In: Haralick, R.M.; Shapiro, L.G. *Computer and Robot Vision*, Chap. 16, Vol. II, pp. 289-379. Addison-Wesley, 1993.

CHAPTER 16

IMAGE MATCHING

16.1 Introduction

16.1.1 Image Matching and Object Reconstruction

Many computer vision tasks require the analysis of two or more images. Time-varying sequences for recognizing parts on a conveyor belt based on their three-dimensional shape or for the visual inspection of the geometry of manufactured parts; for the medical diagnosis of beating hearts; for monitoring land use; for deriving topographic maps from satellite or aerial imagery; can only be accomplished if, at least, pairs of related images are available. Other examples include the analysis of slices of computer tomography images. In principle, the inherent goal of these tasks is object reconstruction, that is, the determination of an object's pose or shape.

The situation we predominantly want to discuss here is symbolically sketched in Fig. 16.1(a). A concrete example is given in Fig. 16.1(b). The object is mapped into the images I^1 and I^2, and possibly further images I^k, via the transformations T_0^1 and T_0^2, and possibly additional transformations T_0^k. These transformations describe all aspects of the imaging process, namely, illumination, reflectance, sensing, and the like. They essentially depend on the geometry of the object (form, pose), of the illumination sources (point, line, or area source), of the sensing devices (interior orientation), and so on.

The mapping can be summarized as

$$I^k = T_0^k(p_G, p_R, p_P, p_A, \ldots)$$

This chapter was written by Wolfgang Förstner.

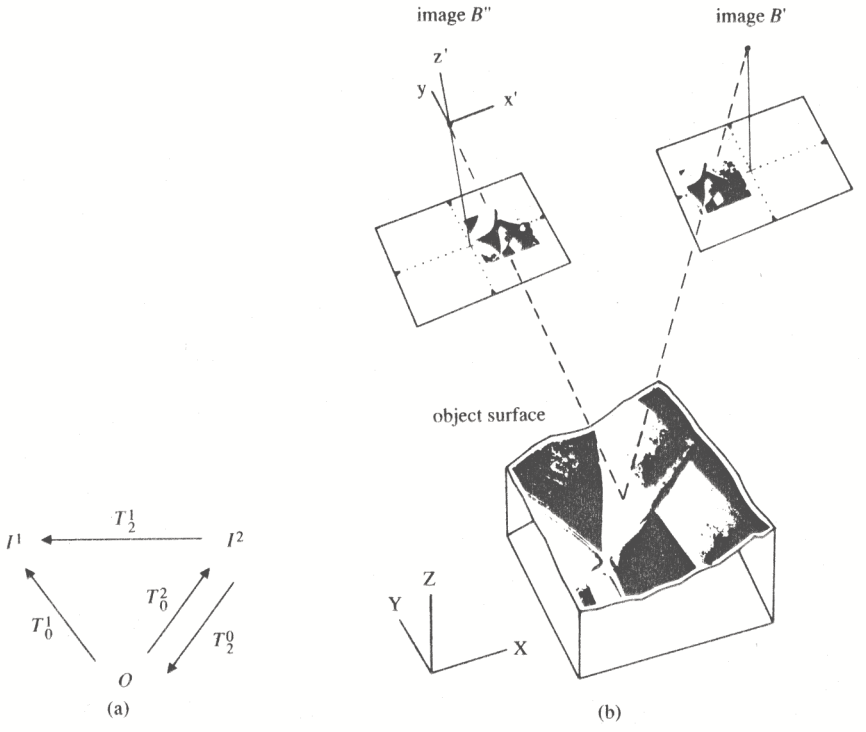


Figure 16.1 General setup of object reconstruction and image matching. (a) The object is mapped to I^1 and I^2 via the transformations T_0^1 and T_0^2. If enough images I^k are available, the reconstruction of the imaging process is possible in principle. If the reflectance properties of the surface O can be derived from one image, the compound transformation $T_2^1 = T_0^1 \cdot (T_0^2)^{-1}$ contains only geometric parameters that may be derived by using image-matching techniques. (b) The setup for two perspective images (from Wrobel, 1987). The reconstruction of the surface is possible either by a direct solution of the parameters describing the surface's geometry and reflectances or by forward intersection after having established correspondences between homologue image features and interpolation.

where $p_G, p_R, p_P, p_A, \ldots$ represent generally unknown parameters specifying the geometry and the reflectance properties of the object, the pose of the camera k, and the atmospheric response. Nearly all object reconstruction tasks, such as those mentioned above, can be modeled this way, possibly after including dynamic models for describing the change of the object over time. If enough images are taken, one in principle can recover the parameters from the observed intensities. Otherwise further regularization has to be applied by putting additional constraints onto the parameters. For analyzing perspective images, the most general case that has been realized assumes that the atmospheric parameters are known, the orientation parameters are restricted to six per image, and the reflectance and surface functions are

described in a finite-element fashion (Helava, 1988; Ebner et al., 1987; Wrobel, 1987).

The general model assumes that the relation between object points or features and image points or features is established. This correspondence between object and image can be found automatically if constraints hidden in the transformation can be made explicit—for example, if the object's geometry is restricted to a polyhedron and if the surface reflectance obeys some regularities. In many cases these constraints and regularities are tight enough to enable one to find homologous features in the images—that is, to establish correspondences between image features—without explicitly referring to the object's properties. In the extreme case, only two images are sufficient to recover the object's form. It is this special case of image matching that we want to discuss in this chapter.

Formally, we want to assume that the transformations T_o^k are one to one, except perhaps up to a few parameters. Thus we assume that we can derive the reflectance parameters p_R from one image, given all other parameters. This implicitly excludes transparent surfaces or occlusions. The image I^2 can be predicted from I^1 by using the compound transformation

$$T_2^1 = T_0^1 \cdot T_2^0 = T_0^1 \cdot (T_0^2)^{-1}(p_G, p_R, p_{P1}, p_{P2}, p_A, \ldots)$$

The geometric part now contains the inverse perspective of image I^2, thus including the surface undulations, and also the perspective of image I^1. Only part of the pose parameters, p_{P1} and p_{P2}, can be derived, namely, those that do not depend on the external coordinate system of the object. The radiometric part is eliminated or at least restricted to a few parameters p_R if the reflectance does not depend strongly on the shape or the pose of the object, as in, for example, diffuse reflectance. In this case the difference between I^1 and I^2 is governed by the form of the object.

Under these restrictions the task of image analysis in all these applications is to establish correspondence between the images or to bring the images into registration. For perspective images the techniques of analytic photogrammetry (see Chapter 14) may be used to determine pose parameters or three-dimensional coordinates of object points.

16.1.2 The Principle of Image Matching

Image matching starts from two digital or digitized images or image patches I' and I''. Their size may vary from 5×5 pixels for tracking points in image sequences to $10,000 \times 10,000$ pixels for registering satellite images. Let the points P' and P'' of images I' and I'' have coordinates (r',c') and (r'',c'') and intensities g' and g'', respectively. Assume that if (r',c') and (r'',c'') are corresponding points, their coordinates can be related by

$$(r',c') = T_G(r'',c'';p_G) \tag{16.1}$$

where T_G is a specified geometric spatial-mapping function reflecting the knowledge about the geometric relation between the images, and p_G is a set or a vector of unknown parameters.

The intensities on one image can be related to those on the other by

$$g' = T_I(g''; p_I) \tag{16.2}$$

where the intensity-mapping function T_I contains the knowledge about the intensity relation between the images, again with the vector p_I being unknown. This leads to the complete model of *image matching*

$$g'(r',c') = T_I\{g''[T_G(r'',c''; p_G)]; p_I\} \tag{16.3}$$

The mapping functions T_G and T_I may be deterministic, stochastic, and/or piecewise continuous.

For optical or range images, the correspondence of points P' and P'' means that P' and P'' relate to the same point on the object. An *arbitrary* pair of points (P', P'') may have two states: P' and P'' are corresponding or they are not corresponding. As shown in Fig. 16.2, the problem of image matching or correspondence now consists of two parts:

1. Finding all corresponding points;
2. Determining the parameters p_G and p_I of the mapping functions T_G and T_I.

The solutions of parts 1 and 2 are not fully equivalent: The solution of part 1 usually does not imply a solution of part 2, as no explicit or implicit determination of the intensity mapping is available. But the solution of part 1 can be derived from the solution of part 2 if the spatial mapping is applied to all positions. In practice, however, this is neither feasible nor necessary, as the mapping function for physical reasons can generally be assumed at least piecewise smooth; thus only a limited number of corresponding points are sufficient to reconstruct the mapping function.

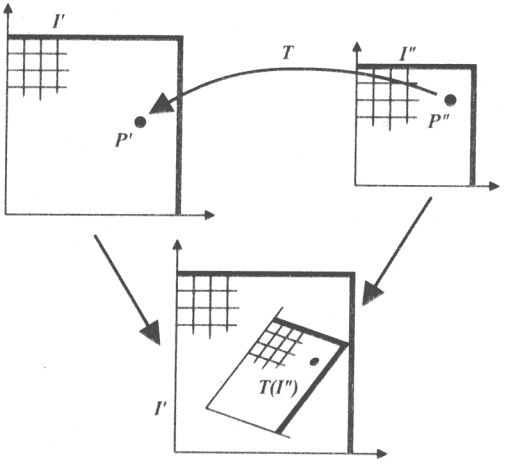


Figure 16.2 Principle of image matching and correspondence: (1) Find corresponding points P' and P'', and (2) determine mapping function T.

This implicitly assumes that the sampling theorem holds for the representation of the geometric spatial mapping.

Most approaches use only a limited number of points and explicitly or implicitly use mapping functions of the type represented in Eqs. (16.1) and (16.2). The solution is always based on the intensity functions g' and g'' or on the attributes a' and a'' of the points

$$P' = P'(r', c'; a') \quad \text{and} \quad P'' = P''(r'', c''; a'')$$

derived from the intensity functions in a neighborhood around (r', c') and (r'', c'').

The solution then is achieved in a three-step procedure:

1. Select appropriate *image features* in one or in both images, possibly by using an interest operator to derive a list of interesting points or edges in one image and another list for the second image. The selection of image features in only one image may be done using the intensity function as the describing attribute. The list of *all* (feasible) positions in the second image must then be searched by an area- or intensity-based image-matching procedure.
2. Find corresponding feature or point pairs (P'_i, P''_j), with P'_i from the first list and P''_j from the second list, that fulfill the criteria of *similarity* and *consistency*. Similarity is based on the attributes a'_i and a''_j of the features or points and thus on the properties of the intensity function, whereas consistency is based on the degree to which the spatial-mapping function is fullfilled. Both similarity and consistency usually are used in an *algorithm* to find an optimal solution, which obviously requires a relative weighting between similarity and consistency.
3. For stereo images, one may have to interpolate the parallaxes $(r'' - r', c'' - c')$ between the selected feature or point pairs in order to obtain a dense parallax field. This interpolation may be based on the *spatial-mapping function*. However, quite different interpolation schemes may also be used, especially if T_G is only given implicitly, for example, by the algorithm to optimize the consistency.

Practically all methods for image matching follow these steps but use significantly different image features, similarity or consistency measures, algorithms, and mapping functions. Some methods are based on the structure of the image content, and thus on a relational description of the images (Price, 1985; Boyer and Kak, 1986, 1988). These methods will be discussed in the following two chapters in detail.

16.1.3 Image Matching Procedures

Table 16.1, which is by no means exhaustive, summarizes the main properties of some matching algorithms. The applications lie in image sequence analysis (TV scenes, robotics, tracking, visual navigation) and in surface reconstruction from stereo images. The approach of Stockman, Kopstein, and Bennett (1982) and Stockman (1987) has also been applied to object location in industrial environments.

Most *interest operators* either try to use no knowledge about the scene and select edges, blobs, or statistically defined points or they rely on special properties

Table 16.1 Properties of some correspondence algorithms for image matching.

Author	Year	Features and Attributes of Interest Operator	Similarity Measure	Mapping Function and Interpolation	Algorithm	Application
Hannah	1974 1989	Variance Moravec	Correlation	NP /-	Local search H	ST
Barnard and Thompson	1980	Moravec	Intensity difference	NP/-	DR	ST
Dreschler	1981	Corners Blobs	Intensity difference class Interest value	NP /-	DR	IS
Baker and Binford	1982	Edges	Sign Strength	NP/ Linear	DP	ST
Grimson	1981 1985	Edges	Direction sign	NP/ PW, smooth	NN	ST
Stockman Kopstein, and Bennett Stockman	1982 1987	Abstract edges	Class	Similarity transfor-mation	Clustering	Registration object location
Benard	1983	edges blobs	grad. + intens. diff., direction	NP linear	DP	ST

Source: Adapted from Förstner, 1986a

Note: NP=non-parametric, PW=piecewise, DR=discrete relaxation, R=relaxation, IS=image sequence, DP=dynamic program, NN=nearest neighbor, H=hierarchical, ST=stereo, and ACF=auto-correlation function.

of the scene, assuming the ability to find corners or junctions of edges. In all cases the attributes of the features are derived from local neighborhoods of the points.

The applied *similarity* measures reflect the assumed intensity-mapping function. In the most simple case of identity mapping $g' = T_I(g'') = g''$, intensity differences serve to measure similarity. If the product moment correlation coefficient is used to measure similarity, T_I implicitly is assumed to be linear: $g' = ag'' + b$. If, however, the intensities are assumed to vary regionwise, properties of edges may be used to measure similarity.

A similar link between *consistency measure* or *interpolation* method and surface type can be established in the case of stereo images in normal orientation

Table 16.1 —*Continued*

Author	Year	Features and Attributes of Interest Operator	Similarity Measure	Mapping Function and Interpolation	Algorithm	Application
Nagel and Enkelmann	1983	Corners	Intensity difference	PW smooth	R	IS
	1986		Interest value	dto.		
Förstner	1984	Roundness	SNR	Affine	R	ST
	1986	and curvature of ACF corners Blobs	Interest value Distinctness	dto.		
Zimmermann and Kories	1984	Blobs	Class	NP/–	NN	IS
Faugeras	1985/	Edges	Orientation	NP /–	Hypothesis	ST
Ayache and Faverjon	1987		sign		Verify technique	IS
Ohta and Kanade	1985	Edges	Intensity difference	NP	DP	ST
Kanade	1987		Interest value	Linear		
Barnard	1986	Intensity values	Intensity constraint	PW/ dto.	Stochastic R H	ST
Witkin, Terzopoulos, and Kass	1987	—	Correlation	PW smooth	H multigrid	general signal

(Fig. 16.3). Constant or piecewise constant T_G (a) corresponds to surfaces or surface patches parallel to the image planes; linear T_G (b) corresponds to locally tilted planar surface patches; and smooth or piecewise smooth T_G or interpolation function (c) corresponds to smooth or piecewise smooth surfaces. Observe that, owing to occlusions (d), smooth surfaces may lead to discontinuities in the mapping.

The main difference between the methods is how an *algorithmic* solution for the consistency of the match is achieved. Most methods are iterative. *Relaxation schemes* seem to be the most flexible for a solution of the highly complex optimization problem, as no parametric form of the mapping function is required, though a parametric form may be used to advantage. In order to speed up convergence, hierarchical or scale space techniques, possibly together with multigrid methods, are

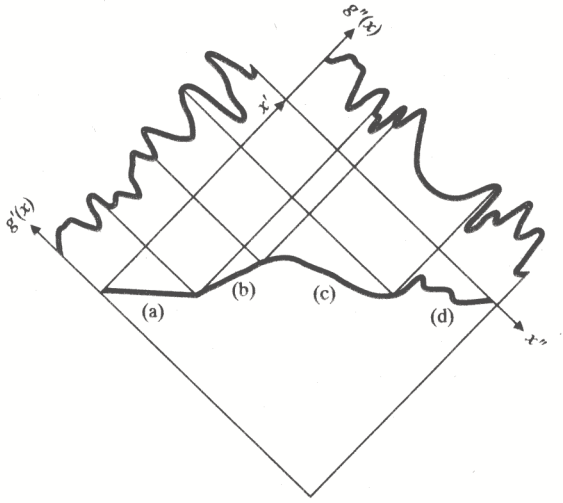


Figure 16.3 One-dimensional surface profile observed from two line cameras by using orthogonal projection: (a) horizontal, (b) sloped, (c) smooth, and (d) piecewise smooth with occlusions.

applied. These approaches guarantee a solution nearly in linear time. *Dynamic programming* also needs no explicit geometric model but optimizes a cost function that takes the similarity of the features into account. In image sequence analysis, consistency can be achieved by selecting the *nearest neighbor* to a point, taking care that the mapping is one to one. These methods can handle occlusions, specifically if the mapping function is assumed to be piecewise smooth. An interesting noniterative approach is the search for a *cluster in parameter space* of the transformation.

The scale space technique of Witkin, Terzopoulos, and Kass (1987) conceptually can handle more than two images, a feature they demonstrated with one-dimensional signals. Table 16.1 does not include the recent approaches by Ebner et al. (1987), Helava (1988), and Wrobel (1987), who independently proposed surface reconstruction schemes with an explicit, finite element approach. They assume piecewise smooth object surfaces. Conceptually they can handle multiple images and then recover possibly locally varying parameters of a reflectance model.

Though quite a variety of approaches exist, certain trends can be observed:

- Most algorithms contain a finite element description that includes discontinuities, which seems to be sufficient for matching two images. If more than two images are available, image matching will loose its power, as truly three-dimensional representations of the object to be recovered have to be used.

- Both intensity-based and feature-based methods are in practical use and may be combined. The result of the feature-based method usually provides the basis for the intensity-based procedure.

Table 16.2 Precision of image matching and edge detection (all figures in pixels).

Author	Year	Empirical Findings	Computer Simulations	Theoretical Values	Applications/Remarks
Sharp, Christensen, and Gilman	1965	1	—	—	Digital terrain models
Bernstein	1973 (cf. 1983)	0.1	—	—	Registration
Klaasman	1975	—	—	0.05	Edge detection
Cafforio and Rocca	1976	0.1	—	—	TV image sequences
McGillem and Svedlov	1976	—	—	0.5/SNR	Registration
Hill	1980	—	0.02–0.1	—	Binary images Target location
Huang/Hsu	1981	—	0.02–0.1	—	Parallax estimation
Förstner	1982	—	—	0.01–0.1	Target location
Thurgood and Mikhail	1982	—	0.02–0.1	—	Target location
Ackermann and Pertl	1983	0.1–0.2	—	—	Parallax estimation
Ho	1983	—	—	0.02–0.2	Binary images Target location
Grün/Baltsavias	1985	0.05–0.1	—	—	Parallax estimation
Vosselman	1986	0.02–0.03	—	—	Target location

Source: Adapted from Förstner, 1984.

- Hierarchical or scale space methods are commonly used to solve the problem of initial approximate values, which nearly all methods require.

Unfortunately, most approaches lack self-diagnosis, and little information on the achievable accuracy is provided by the matching technique itself. Self-diagnosis, however, is indispensable for real applications in order to automatically detect failures of the system and to be certain that the system produces meaningful answers.

Table 16.2 summarizes reported results on the precision of intensity-based matching algorithms and edge detectors in terms of standard deviations. The results are given separately for empirical findings or estimations, for computer simulations,

and for values derived theoretically. There is a tendency toward subpixel accuracy, namely, 0.1 pixel or better. Such results were reported by Bernstein (1973) and Caforio and Rocca (1976). They seem to be realistically obtainable under production conditions and come close to the corresponding results from computer simulations and theoretical studies.

This chapter shows how and under which conditions such high accuracies can be achieved with both intensity-based and feature-based methods. Our development of the algorithms will provide tools for self-diagnosis that can be used first to design matching procedures properly and then to check the performance at run time. The underlying concepts utilize classical least-squares estimation techniques and error propagation, allowing one to derive the precision of the final results in terms of the intensity signal, the noise, and the particular algorithm.

Section 16.2 discusses intensity-based matching of one-dimensional signals, Section 16.3 generalizes the results to two-dimensions. The selection of distinct points in Section 16.4 is discussed. Our approach is based on the expected matching accuracy. It also provides an interpretation of the selected points being independent of intensity-based matching, namely as selection of corners and centers of circular symmetric features. Section 16.5 presents a feature-based matching algorithm using a maximum likelihood type estimation for the geometric transformation. Section 16.6 provides the necessary tools for deriving three-dimensional coordinates from points matched in stereo pairs. Empirical results from an implemented system close the chapter.

16.2 Intensity-Based Matching of One-Dimensional Signals

Intensity-based matching techniques directly refer to the model equation (16.3) and aim at estimating and evaluating the parameters p_G and p_I. To give insight into the principles, we first develop methods for matching one-dimensional signals. As even this task is demanding when taking the statistical properties of the data into account, we restrict this discussion to the case in which one of the signals is assumed to be perfectly known. This situation is relevant to an object location procedure. The model then can be written as

$$g(x) = T_I\{f[T_G(y; p_g)]; p_I\} + n(x)$$

where g', g'', r', and r'' have been replaced by g, f, x, and y, respectively, and the observational noise component $n(x)$ is stated explicitly. We first assume that $f(y)$ is given by sampled data and a fitting or interpolation scheme that allows one to use a derived or estimated continuous $f(y)$ from which the first derivatives can be obtained analytically. This is just like the facet model (Chapter 8). Because of the highly nonlinear character of the estimation problem, we always assume that approximate values $p_I^{(0)}$ and $p_G^{(0)}$ are known from some prior information or prediction scheme. This initial approximation permits us to replace the nonlinear problem by a linear substitute problem whose solution then gives rise to better approximations. In this way we naturally arrive at an iterative solution to the nonlinear problem. We therefore always assume that second-order effects are sufficiently small specifically

(a) when interpolating f or its derivative, (b) when neglecting the random nature of f if it is derived from real data, or (c) when deriving variances for the estimated parameters. A variety of controlled tests have shown that the approximations related to these assumptions are fully acceptable in most applications.

16.2.1 The Principle of Differential Matching

The simplest kind of matching is that of two one-dimensional signals derived from two views of the same scene. For now, one could think of these signals as being image rows.

To simplify the explanation of the estimation procedure involved in the matching, we assume that there are no intensity changes due to viewing direction and that

$$y(x_i) = x_i - \hat{u}(x_i)$$

where $\hat{u}(x_i)$ is the unknown deformation at x_i to produce the corresponding point in the object. The *nonlinear model* can thus be expressed as

$$g(x_i) = f[x_i - \hat{u}(x_i)] + n(x_i), \qquad i = 1, \ldots, m$$

which is valid for all m observed values $g(x_i)$, $i = 1, \ldots, m$, and where $n(x_i)$ is the additive noise at position x_i, which we assume to be independent and identically distributed with mean zero and variance σ_n^2. The nonlinear model makes the assumed observation process explicit. An example for the functions f and g is given in Fig. 16.4. We now assume that we already have a function approximation u_0 of the

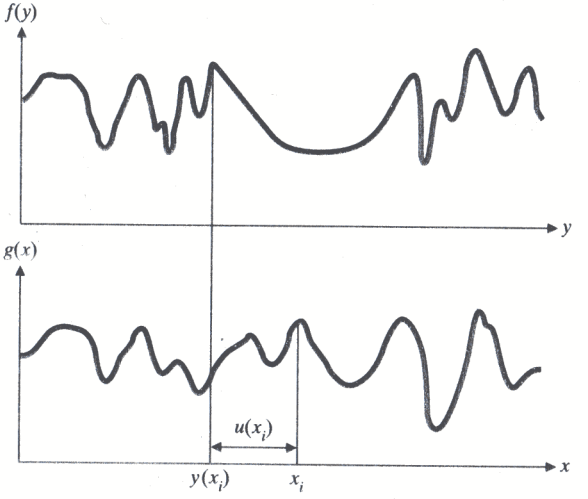


Figure 16.4 Assumed observation process. The observed function $g(x)$ is a noisy version of the deformed function $f(y)$. The deformation $u(x) = x - y(x)$ refers to the coordinates of the observed function $g(x)$ and is positive in the example point x_i, as the point x_i corresponding to $y(x_i)$ is to the right of $y(x_i)$.

function $\hat{u}$. For example, this approximation can be considered given because of the highly constrained geometry inherent in many factory-floor vision tasks. With approximate values $u_0(x_i)$ given, we can write

$$\hat{u}(x_i) = u_0(x_i) + \widehat{\Delta u}(x_i)$$

where the $\widehat{\Delta u}(x_i)$ is the unknown correction for the unknown value $\hat{u}(x_i)$. Thus the nonlinear model can be written as

$$\begin{aligned} g_i &= f(x_i - \hat{u}_i) + n_i \\ &= f(x_i - u_{0i} - \widehat{\Delta u}_i) + n_i, \qquad i = 1, \ldots, m \end{aligned}$$

with the abbreviations $g_i = g(x_i)$, $\hat{u}_i = \hat{u}(x_i)$, $n_i = n(x_i)$ $u_{0i} = u_0(x_i)$, and $\widehat{\Delta u}_i = \widehat{\Delta u}(x_i)$.

Now we linearize f around the point $x_i - u_{0i}$ to obtain

$$g_i = f(x_i - u_{0i}) - f'(x_i - u_{0i})\widehat{\Delta u}_i + \frac{1}{2}f''(x_i - u_{0i} - \xi\widehat{\Delta u}_i)(\widehat{\Delta u}_i)^2 + n_i$$

for some $\xi \in [0, 1]$, and where $f'(y) = df/dy$ and $f''(y) = d^2f/dy^2$.

In the following we assume that f' does not vanish and that the second-order term is negligible. With the differences

$$\Delta g_i = \Delta g(x_i) = g(x_i) - f(x_i - u_{0i})$$

and the derivative

$$f_i' = f'(x_i - u_{0i})$$

we then obtain the *linearized model* in which Δg_i and f_i' are known and $\widehat{\Delta u}_i$ is unknown:

$$\Delta g_i = -f_i' \widehat{\Delta u}_i + n_i, \qquad i = 1, \ldots, m \tag{16.4}$$

or explicitly

$$g(x_i) - f[x_i - u_0(x_i)] = -\frac{df(y)}{dy}\Big|_{y=x_i-u_0(x_i)} \widehat{\Delta u}(x_i) + n(x_i), \qquad i = 1, \ldots, m$$

We can easily determine the random variable $\widehat{\Delta u}_i$ assuming $f_i' \neq 0$:

$$\widehat{\Delta u}_i = -\frac{\Delta g_i}{f_i'} \tag{16.5}$$

or explicitly

$$\widehat{\Delta u}_i = -\frac{g(x_i) - f[x_i - u_0(x_i)]}{\frac{df(y)}{dy}\Big|_{y=x_i-u_0(x_i)}}$$

Thus

$$\hat{u}_i = u_{0i} + \widehat{\Delta u}_i$$

The estimated local deformation—that is, the shift $\hat{u}(x_i)$—therefore simply uses the difference of the functions at the corresponding position $x_i - u_0(x_i)$ and x_i of f and g and the slope f' (Fig. 16.5). This procedure is also termed the *differential approach* when estimating the optical flow (cf. Huang, 1981). As the variance $\sigma^2_{\Delta g_i}$

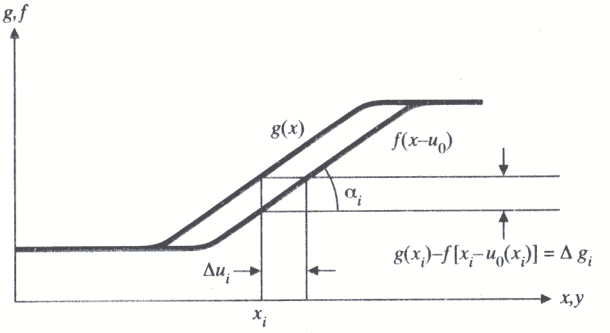


Figure 16.5 Principle of the applied differential approach to estimate the local deformation $\Delta u(x_i)$. The functions g and f are assumed to be locally approximable by a linear function. Then one is able to derive the shift $\Delta u_i = -\Delta g_i / f_i'$ from the difference Δg_i of f and g at the corresponding positions x_i and $x_i - u_0(x_i)$, which depends on the approximate deformation $u_0(x_i)$ and on the known slope $f_i' = \tan\alpha$.

of Δg_i is equal to the noise variance σ_n^2, we obtain for the standard deviation $\sigma_{\hat{u}_i}$ of the estimate $\hat{u}_i = u_0(x_i) + \widehat{\Delta u_i}$

$$\sigma_{\hat{u}_i} = \frac{\sigma_n}{f_i'} \tag{16.6}$$

The precision of the estimate obviously depends on the slope f_i' of the given function and is high at edges with high slope, which is to be expected. No estimate for u is available if $f_i' = 0$, and for small f_i' the weight, $w_i = f_i'^2/\sigma_n^2$, of this estimate is low.

As there is no redundancy in this estimate, it is unreliable in the sense that errors in the approximate values $u_0(x_i)$ and the model—especially wrong assumptions relative to the invariance of lighting—fully influence the estimate. Therefore additional constraints are necessary to improve the result. Specifically, knowledge about the structure of $u(x)$ can allow the use of larger windows. In the following section we discuss deformation functions $u(x)$ with increasing complexity and add parameters for brightness and contrast differences.

16.2.2 Estimating an Unknown Shift

We assume here that $\hat{u}(x)$ is only a uniform shift, $\hat{u}(x) = \hat{u}$, so that $y(x_i) = x_i - \hat{u}$. Further, we assume that u_0 is an initial approximation for u, so that we can write $\hat{u} = u_0 + \widehat{\Delta u}$ for all x. Then by linearizing f around $x_i - u_0$ for each i, there results

$$f[x_i - (u_0 + \widehat{\Delta u})] = f(x_i - u_0) - f'(x_i - u_0)\widehat{\Delta u}$$

The linearized model can then be expressed as

$$g(x_i) = f(x_i - u_0) - f'(x_i - u_0)\widehat{\Delta u} + n_i, \quad i = 1, \ldots, m$$

or in short

$$\Delta g_i = -f_i' \widehat{\Delta u} + n_i, \qquad i = 1, \ldots, m \tag{16.7}$$

holding for all m observed values g_i. Minimizing $\Omega = \sum_{i=1}^{m} n_i^2$ by choosing the appropriate value of $\widehat{\Delta u}$ yields

$$\frac{\partial \Omega}{\partial \widehat{\Delta u}} = \frac{\partial}{\partial \widehat{\Delta u}} \sum_{i=1}^{m} (f_i' \Delta g_i + f_i' \widehat{\Delta u})^2$$

$$= 2 \sum_{i=1}^{m} f_i' \Delta g_i + 2 \left(\sum f_i'^2 \right) \widehat{\Delta u} = 0$$

from which follows the estimate

$$\widehat{\Delta u} = -\frac{\sum_{i=1}^{m} f_i' \Delta g_i}{\sum_{i=1}^{m} f_i'^2} = -\frac{\sum_{i=1}^{m} [g(x_i) - f(x_i - u_0)] f'(x_i)}{\sum_{i=1}^{m} [f'(x_i)]^2} \tag{16.8}$$

To get information about the precision of the estimate $\hat{u}$, we write the estimate as

$$\hat{u} = u_o + \sum_{i=1}^{m} a_i \Delta g_i = u_o + a^T \Delta g$$

where

$$a_i = -f_i' / \sum f_i'^2$$

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

$$\Delta g = \begin{pmatrix} \Delta g_1 \\ \Delta g_2 \\ \vdots \\ \Delta g_m \end{pmatrix}$$

Letting $\sigma_{\hat{u}}^2 = E[(\hat{u} - E(\hat{u})]^2$, we obtain

$$\sigma_{\hat{u}}^2 = E\{[u_0 + a^T \Delta g - E(u_0 + a^T \Delta g)][u_0 + \Delta g^T a - E(u_0 + \Delta g^T a)]\}$$

$$= E\{a^T [\Delta g - E(\Delta g)][\Delta g^T - E(\Delta g^T)] a\}$$

$$= a^T E\{[\Delta g - E(\Delta g)][\Delta g^T - E(\Delta g^T)]\} a$$

Since the covariance matrices $\Sigma_{\Delta g \Delta g}$ and Σ_{nn} are equal to $Diag(\sigma_{n_i}^2) = \sigma_n^2 \cdot I$,

$$\sigma_{\hat{u}}^2 = a^T \Sigma_{nn} a = a^T Diag(\sigma_{n_i}^2) a$$

$$= \left(\sum_{i=1}^{m} a_i^2 \right) \sigma_n^2$$

As

$$\sum_{i=1}^{m} a_i^2 = \frac{\sum_{i=1}^{m} f_i'^2}{(\sum_{i=1}^{m} f_i'^2)^2} = \frac{1}{\sum_{i=1}^{m} f_i'^2}$$

we finally obtain for the variance of $\hat{u}$

$$\sigma_{\hat{u}}^2 = \frac{\sigma_n^2}{\sum_{i=1}^{m} f_i'^2} \tag{16.9}$$

If we define the *mean squared gradient* of f_i' by

$$\sigma_{f'}^2 \doteq \frac{\sum_{i=1}^{m} f_i'^2}{m} \tag{16.10}$$

we can represent the standard deviation for the estimated shift $\hat{u}$ by

$$\sigma_{\hat{u}} = \frac{\sigma_n}{\sqrt{m}\,\sigma_{f'}} \tag{16.11}$$

In an intuitive manner the standard direction depends on the number m of the observed values, the noise variance σ_n^2, and the texture or edge busyness measured by the mean squared gradient of the object.

We would obtain the same result for the estimate $\hat{u}$ and its variance by taking the weighted average of the individual $\widehat{\Delta u}_i$ from Eq. (16.5). For weights we use

$$w_i = \frac{f_i'^2}{\sum_{j=1}^{m} f_j'^2} \tag{16.12}$$

which sum to 1 and are inversely proportional to the variances of the $\widehat{\Delta u}$ that are given by Eq. (16.6). That is,

$$\begin{aligned} \widehat{\Delta u} &= \frac{\sum_{i=1}^{m} \widehat{\Delta u}_i w_i}{\sum w_i} \\ &= -\frac{\sum_{i=1}^{m} \frac{\Delta g_i}{f_i'} f_i'^2}{\sum_{i=1}^{m} f_i'^2} \end{aligned} \tag{16.13}$$

The estimate obviously is independent of the assumed variance σ_n^2 of the noise.

As the common denominator $\sum f_j'^2$ in the weights of Eq. (16.12) cancels in the estimates equation (16.13), one can simply use $w_i = f_i'^2$ as weights. Thus only those parts where the slopes f_i' are nonzero contribute to the estimate, which accords with intuition. The selection of edges is often based on the relative maxima of f'^2; this can be interpreted as a selection of places that locally contribute the most to the estimate of the signal's geometry. The interest operator presented in Section 16.4 is a two-dimensional version of this idea.

If m is large enough, the noise variance can be estimated from the residuals of the least-squares fit:

$$\hat{\sigma}_n^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left[g(x_i) - f(x_i - \hat{u}) \right]^2 \tag{16.14}$$

One can show (Koch, 1987) that the precision of the estimated noise standard deviation $\hat{\sigma}_n$ is

$$\hat{\sigma}_{\hat{\sigma}_n} = \frac{1}{\sqrt{2(m - m_p)}} \hat{\sigma}_n \tag{16.15}$$

with the number $m_p = 1$ of the unknown parameters, if the simple model of Eq. (16.7) really holds. Thus at least 72 samples are necessary to yield an accuracy of better than 10% in $\hat{\sigma}_n$.

EXAMPLE 16.1

Let an edge model $f(y_i)$ and the observed edge $g(x_i)$ be given:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f_i	10	10	10	10	20	30	40	40	40	40	40	40	40	40
g_i	—	—	—	14	13	14	26	37	42	41	42	—	—	—
Δg_i	—	—	—	4	-7	-16	-14	-3	2	1	2	—	—	—
f_i'	0	0	0	5	10	10	5	0	0	0	0	0	0	0
n_i	—	—	—	4	3	-4.8	-2.8	-1.8	2	1	2	—	—	—

The pixel spacing is Δx, the gradients $f_i' = (f_{i+1} - f_{i-1})/2[gr/\Delta x]$, and the approximate value $u_0 = 0[\Delta x]$. Omitting the first and last three observations, we obtain the estimate

$$\hat{u} \doteq -\frac{\sum f_i' g_i}{\sum f_i'^2} = -\frac{5 \cdot 4 + 10 \cdot (-7) + 10 \cdot (-16) + 5 \cdot (-14)[gr^2/\Delta x]}{25 + 100 + 100 + 25[gr^2/\Delta x^2]}$$

$$= -\frac{-280[gr^2/\Delta x]}{250[gr^2/\Delta x^2]} = 1.118[\Delta x]$$

The theoretical precision of this estimate, when assuming $\sigma_n = 1[gr]$, is

$$\sigma_{\hat{u}} = \frac{\sigma_n}{\sqrt{\sum f_i'^2}} = \frac{1[gr]}{\sqrt{250[gr^2/\Delta x^2]}} \approx 0.06[\Delta x]$$

and thus less than one-tenth of a pixel. Only observations 4, 5, 6, and 7 have nonzero weights and hence contribute $\hat{u}$. The residuals $n(x_i)$ are comparatively large. The estimated noise standard deviation according to Eq. (16.14) is $\sigma_n \approx \sqrt{1/(8-1) \cdot 45.32[gr^2]} \approx 3.1[gr]$. This is significantly larger than $1[gr]$. There are obvious differences in brightness, contrast, and possibly also in scale.

16.2.3 Estimating Unknown Shift and Scale

We now augment our model by assuming that the transformation contains a parameter compensating for unknown scale:

$$y_i = s\,(x_i - x_0) - \hat{u} = s\,\bar{x}_i - \hat{u}$$

Note our use of a fixed reference point x_0, which we will later choose for our convenience. From here on we work in the reduced coordinates

$$\bar{x}_i = x_i - x_0$$

The linear model reads as

$$g(x_i) = f(s\,\bar{x}_i - \hat{u}) + n(x_i), \quad i = 1, \ldots, m$$

Using approximate values s_0 and u_0 for the unknown parameters s and $\hat{u}$ and linearizing around $s_0\bar{x}_i - u_0$ yields

$$g(x_i) = f(s_0\bar{x}_i - u_0) - f'(s_0\bar{x}_i - u_0)\,\widehat{\Delta u} + f'(s_0\bar{x}_i - u_0)\,\bar{x}_i\,\widehat{\Delta s} + n(x_i)$$

This result leads to the linearized model

$$\Delta g_i = -f_i'\widehat{\Delta u} + f_i'\bar{x}_i\widehat{\Delta s} + n(x_i), \quad i = 1, \ldots, m$$

again holding for all m observed values, and where $\widehat{\Delta s}$ satisfies $s = s_0 + \widehat{\Delta s}$.

Minimizing $\Omega = \sum n^2(x_i)$ with respect to the two unknown parameters, we set $\partial\Omega/\partial\widehat{\Delta u} = 0$ and $\partial\Omega/\partial\widehat{\Delta s} = 0$ to yield

$$\frac{\partial\Omega}{\partial\widehat{\Delta u}} = \frac{\partial\Omega}{\partial\widehat{\Delta u}} \sum_{i=1}^{m} (\,\Delta g_i + f_i'\,\widehat{\Delta u} - f_i'\bar{x}_i\,\widehat{\Delta s})^2$$

$$= 2\sum_{i=1}^{m} f_i'(\,\Delta g_i + f_i'\,\widehat{\Delta u} - f_i'\bar{x}_i\,\widehat{\Delta s}) = 0$$

and analogously

$$\frac{\partial\Omega}{\partial\widehat{\Delta s}} = -2\sum_{i=1}^{m} f_i'x_i(\,\Delta g_i + f_i'\,\widehat{\Delta u} - f_i'\bar{x}_i\,\widehat{\Delta s}) = 0$$

After some rearranging we can obtain the matrix equation

$$\begin{pmatrix} \sum f_i'^2 & -\sum f_i'^2\bar{x}_i \\ -\sum f_i'^2\bar{x}_i & \sum f_i'^2\bar{x}_i^2 \end{pmatrix} \begin{pmatrix} \widehat{\Delta u} \\ \widehat{\Delta s} \end{pmatrix} = \begin{pmatrix} -\sum f_i'\,\Delta g_i \\ \sum f_i'\bar{x}_i\,\Delta g_i \end{pmatrix}$$

where the sums are to be taken over all i, $i = 1, \ldots, m$. The solution of this 2×2 normal equation system yields the least-squares estimates for the unknown corrections $\widehat{\Delta u}$ to u_0 and $\widehat{\Delta s}$ to s_0:

$$\hat{u} = u_0 + \widehat{\Delta u}, \quad \hat{s} = s_0 + \widehat{\Delta s}$$

The variance of the noise can again be derived from the residuals

$$\hat{\sigma}_n^2 = \frac{1}{m-2} \sum (\Delta g_i + f_i' \widehat{\Delta u} - f_i' \bar{x}_i \widehat{\Delta s})^2$$

In order to analyze the estimates in detail, we shift the coordinate system such that the off-diagonal term in the normal equation becomes zero. The shift x_0 then has to be chosen so that $\sum f_i'^2 \bar{x}_i = 0$ or

$$x_0 = \frac{\sum_{i=1}^{m} f_i'^2 x_i}{\sum f_i'^2} \quad \text{with } f_i' = \left. \frac{df(y)}{dy} \right|_{y=s_0(x_i - x_0) - u_0} \tag{16.16}$$

This is the *weighted center of gravity of the object,* where the weights are again the squares of the gradient of f evaluated at the corresponding positions $s_0(x_i - x_0) - u_0$.

With this choice of x_0, the estimates $\widehat{\Delta u}$ and $\widehat{\Delta s}$ are statistically independent. The estimate here for $\widehat{\Delta u}$ is identical to the estimate without scale parameter. Thus if one is free to choose a point in the image to be transformed into the object by using $y_i' = \hat{s}(x_i - x_0) - \hat{u}$, one may use the weighted center of gravity and solve for the shift only. Also, the precision of this shift is independent of the scale.

In general, a point x_i in the image corresponds to the *predicted point* $\hat{y}_i = \hat{s}(x_i - x_0) - \hat{u}$ in the object. The variance of the predicted point $\hat{y}_i$ is given by

$$\sigma_{\hat{y}_i}^2 = \sigma_n^2 \left(\frac{1}{\sum f_i'^2} + \frac{(x_i - x_0)^2}{\sum f_i'^2 \bar{x}_i^2} \right) \tag{16.17}$$

Thus the variance of points, transformed from the image into the object, increases with increasing distance from the weighted center of gravity. The weighted center of gravity x_0 from Eq. (16.16) is the point that, when transformed into the object, has the smallest variance.

We will generalize this result in the two-dimensional case and use it as a basis for the derivation of an interest operator in Section 16.4.

EXAMPLE 16.2

With the same data as in Example 16.1, we want to determine the shift u and a geometric scale s. We assume $u_0 = 1$, approximating the first estimate, and $s_0 = 1$. The 2×2 normal equation is

$$\begin{pmatrix} 250[gr^2/\Delta x^2] & -1625[gr^2/\Delta x] \\ -1625[gr^2/\Delta x] & 10725[gr^2] \end{pmatrix} \begin{pmatrix} \widehat{\Delta u} \\ \widehat{\Delta s} \end{pmatrix} = \begin{pmatrix} 100[gr^2/\Delta x] \\ -685[gr^2] \end{pmatrix}$$

which yields the estimates

$$\hat{u} = u_0 + \widehat{\Delta u} = 1 - 1.002[\Delta x] = -0.002[\Delta x]$$

and

$$\hat{s} = s_0 + \widehat{\Delta s} = 1 - 0.216[1] = 0.784$$

With the assumed noise standard deviation $\sigma_n = 1[gr]$ and the inverse $Q = N^{-1}$,

$$\begin{pmatrix} 0.2640[\Delta x^2/gr^2] & 0.0400[\Delta x/gr^2] \\ 0.0400[\Delta x/gr^2] & 0.006156[1/gr^2] \end{pmatrix}$$

we obtain the theoretical standard deviations according to

$$\sigma_{\hat{u}} = \sigma_n \sqrt{Q_{11}} = 1[gr] \cdot \sqrt{0.2640[\Delta x/gr^2]} \approx 0.52[\Delta x]$$
$$\sigma_{\hat{s}} = \sigma_n \sqrt{Q_{22}} = 1[gr] \cdot \sqrt{0.006156[1/gr^2]} \approx 0.08$$

The shift is significantly less accurate than in Example 16.1 without the scale parameter. The geometric scale can be determined to approximately 8%. The reason for the increased standard deviation of $\hat{u}$ is that $\hat{u}$ now refers to the origin of the coordinate system alone, which (cf. the table in Example 16.1) lies to the left of the used interval.

The weighted center of gravity of f is

$$x_{0f} = \frac{\sum f_i'^2 x_i}{\sum f_i'^2} = \frac{5 \cdot 4 + 10 \cdot 5 + 10 \cdot 6 + 5 \cdot 7[gr^2/\Delta x]}{250[gr^2/\Delta x^2]} = 5.5[\Delta x]$$

In a similar manner we obtain

$$x_{0g} = 7.12[\Delta x]$$

If we now transform x_{0g}, we get the predicted point $y_{0g} = \hat{s} \cdot x_{0g} - \hat{u} = 0.784 \cdot 7.12[\Delta x] - (-0.002[\Delta x]) = 5.59[\Delta x]$, which is very close to the weighted center of gravity $5.5[\Delta x]$ of f. Its precision is identical to the precision of the shift in Example 16.1, namely, $\sigma_{y_{0g}} = 0.06[\Delta x]$, as it is independent on the scale! This also can be proved by error propagation using $y_{0g} = (-1\ x_{0g})(\hat{u}\ \hat{s})^T$. Thus

$$\sigma_{y_{0g}}^2 = \sigma_n^2(-1\ x_{0g})Q(-1\ x_{0g})^T$$

16.2.4 Compensation for Brightness and Contrast

If object and image have different brightness and contrast, we have to compensate for this difference. Without any geometric scale parameter, the nonlinear model then reads

$$g(x_i) = af(x_i - u) + b + n(x) \tag{16.18}$$

where a represents change in contrast and b change in brightness.

If we start from an approximate value a_0 for a and u_0 for u, we obtain the linearized model in a manner similar to that just described. We let $\Delta g_i = \Delta g(x_i) = g(x_i) - a_0 f(x_i - u_o)$ and obtain

$$\Delta g_i = -a_0 f_i' \widehat{\Delta u} + f_i \widehat{\Delta a} + \hat{b} + n_i, \qquad i = 1, \ldots, m$$

(We do not need an approximate value for b, as the model equation 16.18 is linear in b.)

The normal equation system for the least-squares solution is then given by

$$\underbrace{\begin{pmatrix} a_0^2\sum f_i'^2 & -a_0\sum f_i' f_i & -a_0\sum f_i' \\ -a_0\sum f_i' f_i & \sum f_i^2 & \sum f_i \\ -a_0\sum f_i' & \sum f_i & \sum 1 \end{pmatrix}}_{Q^{-1}} \begin{pmatrix} \widehat{\Delta u} \\ \widehat{\Delta a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} -a_0\sum f_i' \Delta g_i \\ \sum f_i \Delta g \\ \sum \Delta g_i \end{pmatrix} \tag{16.19}$$

where

$$f_i' = \left.\frac{df(y)}{dy}\right|_{y=x_i-u_0}$$

and $f_i = f(x_i - u_0)$, and again the sums are to be taken over all i.

The estimates thus are

$$\hat{u} = u_o + \widehat{\Delta u},\ \hat{a} = a_0 + \widehat{\Delta a},\ \text{and}\ \hat{b}$$

The precision of the estimates can be obtained by using the inverse $Q = (q_{ij})$ of Q^{-1} of the 3×3 normal system in Eq. (16.19), that is

$$\begin{aligned} \sigma_{\hat{u}} &= \sigma_n \sqrt{q_{11}} \\ \sigma_{\hat{a}} &= \sigma_n \sqrt{q_{22}} \\ \sigma_{\hat{b}} &= \sigma_n \sqrt{q_{33}} \end{aligned}$$

The noise variance may be estimated from

$$\hat{\sigma}_n^2 = \frac{1}{m-3} \sum_{i=1}^{m} \left[g(x_i) - \hat{a} f(x_i - \hat{u}) - \hat{b} \right]^2 \tag{16.20}$$

EXAMPLE 16.3

With the same data as in Example 16.1 we want to determine the shift and two radiometric parameters for brightness and contrast. We obtain the 3×3 normal equation system (16.19) (cited without dimensions):

$$\begin{pmatrix} 250 & -750 & -30 \\ -750 & 7900 & 230 \\ -30 & 230 & 8 \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} 100 \\ -90 \\ -1 \end{pmatrix}$$

The estimates are therefore

$$\hat{u} = 1.796[\Delta x], \quad \hat{a} = 0.883, \quad \hat{b} = 6.226[gr]$$

With the inverse normal matrix $Q = N^{-1}$

$$\begin{pmatrix} 0.007877 & -0.000686 & 0.04909 \\ -0.000686 & 0.000836 & -0.02662 \\ 0.04909 & -0.02662 & 1.074 \end{pmatrix}$$

and the estimated noise standard deviation from Eq. (16.20)

$$\sigma_n = \sqrt{20.48[gr^2]/(8-3)} \approx 2.0[gr]$$

we obtain the empirical standard deviations for the estimates

$$\hat{\sigma}_{\hat{u}} \approx 2.0\sqrt{0.007877} = 0.18[\Delta x]$$

and similarly

$$\hat{\sigma}_{\hat{a}} \approx 0.059, \text{ and } \hat{\sigma}_{\hat{b}} = 2.1[gr]$$

The fit between model and data obviously is significantly better than in the two previous examples as the estimated noise. The standard deviation is reduced to only $2[gr]$, which is due to the better modeling of the radiometry.

16.2.5 Estimating Smooth Deformations

If the transformation $y = T_G(x) = y(x)$ (cf. Section 16.2.1) cannot be approximated by a linear function, or if one wants to use larger windows but the transformation is still smooth, it may be represented by

$$y_i = y(x_i) = x_i - u(x_i)$$

with some smoothness constraints on $u(x)$. These could refer to the first, second, or higher derivatives of y or any function $D_u = D(u', u'', \ldots)$ of the derivatives of $u(x_i)$. Thus we could require

$$E\left[D_{u_j}(u)\right] = 0, \qquad V\left[D_{u_j}(u)\right] = \sigma_{D_j}^2, \quad j = 1, \ldots, k \tag{16.21}$$

holding for all j from 1 through k. D_{u_j}, for example, stands for a linear combination

$$D_{u_j} = a_1\, u_j'[u(x)] + a_2\, u_j''[u(x)] + \cdots$$

of the derivatives of u evaluated at x_j.

We now want to estimate the $u_i = u(x_i)$ by using the observed values

$$g(x_i) = f\left[x_i - u(x_i)\right] + n(x_i) \tag{16.22}$$

and the a priori knowledge about the smoothness of $u(x)$ in a Bayesian manner. We can maximize the a posteriori probability

$$p(u \mid g, D_u) = \frac{p(g \mid u)\, p(D_u \mid u)}{\int\int p(g \mid u)\, p(D_u \mid u)\, dg dD_u}$$

Assuming Gaussian distribution for both $g_i = g(x_i)$ and $D_{u_j}(u)$ with

$$p(g_i \mid u) = \frac{1}{\sqrt{2\pi\sigma_{n_i}^2}} \exp\left[-\frac{1}{2}\left(\frac{g_i - f(y_i)}{\sigma_{n_i}}\right)^2\right]$$

$$p(D_{u_j} \mid u) = \frac{1}{\sqrt{2\pi\sigma_{D_j}^2}} \exp\left[-\frac{1}{2}\left(\frac{D_{u_j}}{\sigma_{D_j}}\right)^2\right]$$

and the independence of g_i and D_{u_j}, we obtain the expression

$$\Omega_1(u) = -\log\ p(u \mid g, D_u)$$
$$= \frac{1}{2}\sum_i \left(\frac{g_i - f(x_i - u_i)}{\sigma_{n_i}}\right)^2 + \frac{1}{2}\sum_j \left(\frac{D_{u_j}(u(x))}{\sigma_{D_j}}\right)^2 + \text{const.}$$

to be minimized with respect to u_i.

Now we treat the case in which the smoothness is measured by the second derivative $u''(x_j)$ of $u(x)$. The variances of n_i and of $u''(x_j)$ are assumed to be constant; thus $\sigma_{n_i}^2 = \sigma_n^2$ and $\sigma_{u_j''}^2 = \sigma_c^2$ (c stands for curvature). Then we have to minimize

$$\Omega(u) = 2[\Omega_1(u) - \text{const.}] = \frac{1}{\sigma_n^2}\sum_i [g_i - f(x_i - u_i)]^2 + \frac{1}{\sigma_c^2}\sum_j \{u_j''[u(x)]\}^2 \tag{16.23}$$

The first term in Eq. (16.23) measures the similarity between the given and the observed functions; the second term measures the smoothness. This is equivalent to using, in addition to Eq. (16.21), the fictitious second-derivative curvature-related observations

$$c(x_j) = u''(x_j) + v(x_j), \qquad j = 1, \ldots, k \tag{16.24}$$

with $c(x_j) = 0$ and $\sigma_{v_j}^2 = \sigma_c^2$, and to determine the $u(x_i)$ by a weighted LS technique with the weights being the inverse of the variances. The random variations in the second derivative are represented by $v(x_j)$. If we now represent the second derivative $u''(x_j)$ by

$$u''(x_j) = u(x_{j-1}) - 2u(x_j) + u(x_{j+1})$$

with approximate values $u_0(x_j) = u_{0j}$, and thus

$$u(x_j) = u(x_j) + \Delta u_0(x_j)$$
$$= u_{0j-1} - 2u_{0j} + u_{0j+1} + (\Delta u_{j-1} - 2\Delta u_j + \Delta u_{j+1})$$

then we arrive at the linearized model

$$\Delta g_i = -f_i'\Delta u_i + n_i, i = 1, \ldots, m \tag{16.25}$$
$$-c_{0j} = \Delta u_{j-1} - 2\Delta u_j + \Delta u_{j+1} + v_j, j = 2, \ldots, m-1 \tag{16.26}$$

with

$$\Delta g_i = g(x_i) - f[x_i - u_0(x_i)]$$
$$f_i' = \left.\frac{df(y)}{dy}\right|_{y = x_i - u_0(x_i)}$$
$$c_{0j} = u_{0j-1} - 2u_{0j} + u_{0j+1}$$
$$v_j = v(x_j)$$

Equation (16.25) represents the observation process and Eq. (16.26) represents the smoothness constraint.

The model can be conveniently written in matrix notation

$$\begin{aligned} \Delta g &= -A_1\, \Delta u + n \\ -c &= A_2\, \Delta u + v \end{aligned} \tag{16.27}$$

with the $m \times m$ matrix A_1 containing the derivatives f_i' on the main diagonal

$$\underset{m\times m}{A_1} = \text{Diag}\,(f_i') = \begin{pmatrix} f_1' & & & & \\ & \ddots & & & \\ & & f_i' & & \\ & & & \ddots & \\ & & & & f_m' \end{pmatrix}$$

and the $(m-2) \times m$ matrix A_2 consisting of $m-2$ rows with $(1\ -2\ 1)$ around the main diagonal

$$\underset{(m-2)\times m}{A_2} = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ & 1 & -2 & 1 & & & & \\ & & 1 & -2 & 1 & & & \\ & & & & \ddots & & & \\ & & & & 1 & -2 & 1 & \\ & & & & & & \ddots & \\ & & & & & & 1 & -2 & 1 \end{pmatrix}$$

Minimizing

$$\Omega(u) = w_n \cdot n^T n + w_c \cdot v^T v \tag{16.28}$$

with respect to u yields the normal equation system

$$[w_n A_1^T A_1 + w_c\ A_2^T A_2]\, \Delta u = [w_n A_1^T \Delta g - w_c A_2^T c] \tag{16.29}$$

with the weights $w_n = 1/\sigma_n^2$ and $w_c = 1/\sigma_c^2$. The normal equation matrix N is a pentadiagonal matrix

$$w_n \cdot \text{Diag}\,(f_i'^2) + w_c \cdot \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & & & \ddots & & & \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & 1 & -4 & 5 & -2 \\ & & & & & 1 & -2 & 1 \end{pmatrix}$$

and the right-hand sides are given by

$$\begin{pmatrix} w_n & \cdot & f'_1 & \cdot & \Delta g_1 & - & w_c & \cdot & & & c_{01} \\ w_n & \cdot & f'_2 & \cdot & \Delta g_2 & - & w_c & \cdot & & (-2c_{01} & +c_{02}) \\ w_n & \cdot & f'_3 & \cdot & \Delta g_3 & - & w_c & \cdot & (c_{01} & -2c_{02} & +c_{03}) \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ w_n & \cdot & f'_{m-2} & \cdot & \Delta g_{m-2} & - & w_c & \cdot & (c_{0m-2} & -2c_{0m-2} & +c_{0m}) \\ w_n & \cdot & f'_{m-1} & \cdot & \Delta g_{m-1} & - & w_c & \cdot & (c_{0m-1} & -2c_{0m}) & \\ w_n & \cdot & f'_m & \cdot & \Delta g_m & - & w_c & \cdot & c_{0m} & & \end{pmatrix}$$

This system can be solved by factorization in $O(m)$ time, leading to estimates $\widehat{\Delta u} = (\widehat{\Delta u_i})$ and thus to $\hat{u}_i = u_{0i} + \widehat{\Delta u_i}$ or $\hat{u} = u_0 + \widehat{\Delta u}$.

Often the noise variances σ_n^2 and σ_c^2 are not known. Without proof (Förstner, 1985) we give an estimate for both variances, which in an iterative manner could be used to obtain optimal estimates for both the u_i and the variances σ_n^2 and σ_c^2. The estimation requires the diagonal terms q_{ii} of the inverse $Q = N^{-1}$ of the normal equation matrix. They can be calculated by using a standard routine from a library for solving sparse equation systems.

Including the case of individual variances $\sigma_{n_i}^2$ and $\sigma_{c_j}^2$, the estimates are given by

$$\hat{\sigma}_{n_i}^2 = \frac{1}{r_n}\left(\sum_{i=1}^{m}\left(g(x_i) - f\left[x_i - \hat{u}(x_i)\right]\right)^2 / \sigma_{n_i}^2\right) \cdot \sigma_{n_i}^2$$

and

$$\hat{\sigma}_{c_j}^2 = \frac{1}{r_c}\left(\sum_{j=2}^{m-1}\left[\hat{u}(x_{j-1}) - 2\hat{u}(x_j) + \hat{u}(x_{j+1})\right]^2 / \sigma_{c_j}^2\right) \cdot \sigma_{c_j}^2$$

with

$$Q = N^{-1}$$

$$q_{ii} = (Q)_{ii}$$

$$u_n = \sum_{i=1}^{m}\{f'[x_i - \hat{u}(x_i)]\}^2 \cdot q_{ii}/\sigma_{n_i}^2$$

$$r_n = m - u_n$$

$$r_c = (m-2) - r_n$$

which simplify for $\sigma_{n_i} = \sigma_n$ and $\sigma_{c_j} = \sigma_c$. These estimates assume that no outliers in the observations $g(x_i)$ and no discontinuities are present. The case including outliers and discontinuities may use robust estimation techniques, iteratively weighting down the individual observations depending on the residuals n_i and v_j.

The relative precision of the standard deviations $\hat{\sigma}_n$ and $\hat{\sigma}_c$ is approximately $1/\sqrt{2r_n}$ and $1/\sqrt{2r_c}$, thus leading to reliable results only for sufficiently long profiles. The technique may easily be generalized to two-dimensional functions.

16.2.6 Iterations and Resampling

As $f(x)$ is highly nonlinear, the estimates in general are only improved approximate values. Especially if the initial approximations are crude, the estimates have to be further refined if one wants to exploit the precision inherent in the data. This leads to an iterative estimation scheme.

We obtain the estimates after the νth iteration by, for example,

$$\begin{aligned} \hat{a}^{(\nu)} &= \hat{a}^{(\nu-1)} + \widehat{\Delta a}^{(\nu)} \\ \hat{u}^{(\nu)} &= \hat{u}^{(\nu-1)} + \widehat{\Delta u}^{(\nu)} \end{aligned} \tag{16.30}$$

The constant shift b of the intensity values can be estimated directly, thus we take $b^{(0)} = 0$ for all iterations.

As the positions $y_i = x_i - u(x_i)$ are noninteger values in general, we have to interpolate. Specifically we need

$$f[x_i - u_0(x_i)]$$

and

$$\left.\frac{df(y)}{dy}\right|_{y=x_i-u_0(x_i)}$$

For choosing a proper interpolation, we have to require that the optimization function $\Omega(u)$ be smooth or, equivalently, that the elements in the normal equations change continuously with changing approximate values $u_0(x_i)$. Otherwise small changes of $u_0(x_i)$ may lead to large changes in the estimates and hence to unstable results.

Therefore $f(y)$ and $f'(y)$ at least have to be interpolated linearly:

$$f(x_i - u) = (1 - u/\Delta x) f_i + \frac{u}{\Delta x} f_{i-1} \tag{16.31}$$

$$f'(x_i - u) = (1 - u/\Delta x) f'_i + \frac{u}{\Delta x} f'_{i-1}, \qquad 0 \le u < \Delta x \tag{16.32}$$

where Δx is the spacing of the x_i, $0 \le u \le \Delta x$, and, for example, with the first derivative $f'(x_i)$ from a second-order facet model

$$f'_i = \frac{1}{2\Delta x}(f_{i+1} - f_{i-1}) \tag{16.33}$$

Here we use the abbreviations f_i for $f(x_i)$ and f'_i for $f'_i(x_i)$. Obviously $f(x_i - u)$ and $f'(x_i - u)$ do *not* refer to the same smooth function $f(x)$, as $f'(x_i - u)$ refers to a second-order interpolation scheme of $f(x)$. If we require that $f(x_i - u)$ and $f'(x_i - u)$ refer to the same smooth function, we have to interpolate $f(x_i - u)$ with a second-order function, for example,

$$f(x_i + u) = f_i + f'_i \cdot u + \frac{1}{2} f''_i \cdot u^2, \quad |u| < \frac{1}{2}\Delta x$$

with $f'(x_i)$ as in Eq. (16.33) and

$$f''_i = \frac{2}{\Delta x^2}[f_{i+1} - 2f_i + f_{i-1}]$$

As the slope at $x_i+\Delta x/2$ is then $2(f_{i+1}-f_i)/\Delta x$, and thus twice the first differences, one could instead use the *smoothed* function

$$\tilde{f}(x_i+u)=\overline{f}_i+f_i'\cdot u+\frac{1}{2}f_i''\cdot u^2,\quad |u|<\frac{1}{2}\Delta x$$

with

$$\overline{f}_i=\frac{1}{8}(f_{i-1}+6f_i+f_{i-1})$$

$$f_i'=\frac{1}{2\Delta x}(f_{i+1}-f_{i-1})$$

$$f_i''=\frac{1}{\Delta x^2}(f_{i+1}-2f_i+f_{i-1})$$

Both functions $f(x_i+u)$ and $\tilde{f}(x_i+u)$ pass through the points $\frac{1}{2}(f_{i+1}+f_i)$ and $\frac{1}{2}(f_{i-1}+f_i)$ and have a common slope with the neighboring interpolation elements.

EXAMPLE 16.4

a. With the same data as in Example 16.1, we now perform a second iteration for the shift estimate only. The main data are summarized in the table:

i	4	5	6	7	8	9	10	11
y	2.88	3.88	4.88	5.88	6.88	7.88	8.88	9.88
g	14	13	14	26	37	42	41	42
f	10	10	18.81	28.80	38.80	40	40	40
Δg	4.00	3.00	-4.81	-2.80	-1.80	2.00	1.00	2.00
f'	0	4.4	9.4	10	5.6	0.6	0	0
n	4.00	3.00	-1.93	0.07	1.08	2.00	1.00	2.00

Linear interpolation has been applied to both f and f' (cf. Eqs. 16.31 and 16.32); for example, $f(6-u^{(1)})=f(6-1.118)=f(5-0.118)=(1-0.118)\cdot f_5+0.118\cdot f_4=0.882\cdot 20+0.118\cdot 10=18.81$. The normal equations are

$$\left(\sum_i f_i'^2\right)\cdot\widehat{\Delta u}=\sum_i f_i'g_i\equiv 239.4\cdot\widehat{\Delta u}=68.725$$

leading to the correction $\Delta u^{(2)}=0.288[\Delta x]$ and to a better estimate $u^{(2)}=u^{(1)}+\Delta u^{(2)}=1.406[\Delta x]$ for the shift. Observe the second iteration leading to a change (0.288) in the estimate that is significantly larger than the internal precision (0.06, cf. Example 16.1) of the estimate.

b. In the case of the model of Example 16.3 with the additional radiometric parameters, a second iteration would yield the shift with its standard deviation

$$\hat{u}=1.61[\Delta x],\quad \sigma_{\hat{u}}=0.15[\Delta x]$$

and the estimated noise standard deviation

$$\sigma_n = 1.43[gr]$$

We will compare this result with the one obtained by cross-correlation in Examples 16.5 and 16.6.

16.2.7 Matching of Two Observed Profiles

We have assumed that $f(x)$ is known. Now we want to extend the matching procedure to the case in which *both* profiles are corrupted by noise. We will show how we can reduce the procedure to those methods developed so far.

We start with two noisy profiles

$$g_1(x_i) = f(x_i) + n_1(x_i)$$

$$g_2(x_i) = f(y_i) + n_2(x_i)$$

and the geometric model

$$y_i = x_i - u(x_i)$$

We could have applied a symmetric model distributing $u(x_i)$ equally on both signals (cf. Horn, 1987, sec. 13.9). This would have led to the same results. We again assume that the noise components are independent and white with standard deviations σ_{n_1} and σ_{n_2}. The first profile $g_1(x_i)$ is just the observed function $f(x)$, and $g_2(x_i)$ is the observed and deformed function $f(x)$. With approximate values $u_0(x_i)$, thus $u(x_i) = u_0(x_i) + \Delta u(x_i)$, and again neglecting higher-order terms, we can write

$$g_1[x_i - u_0(x_i)] = f[x_i - u_0(x_i)] + n_1[x_i - u_0(x_i)]$$

and linearize

$$g_2(x_i) = f[x_i - u_0(x_i)] - \left.\frac{df(y)}{dy}\right|_{y=x_i-u_0(x_i)} \cdot \Delta u(x_i) + n_2(x_i)$$

Using the abbreviations

$$\Delta g_i = g_2(x_i) - g_1[x_i - u_0(x_i)]$$

$$f_i' = \left.\frac{df(y)}{dy}\right|_{y=x_i-u_0(x_i)}$$

and

$$\bar{n}_i = n_2(x_i) - n_1[x_i - u_0(x_i)]$$

we obtain the linearized model

$$\Delta g_i = -f_i' \cdot \Delta u(x_i) + \bar{n}_i$$

in full correspondence to Eq. (16.4).

To use this linearized model, we must take into account that the f_i' have to be estimated from the given data. This could be achieved by restoring $g_1(x)$ and $g_2(x)$, yielding $\hat{g}_1(x)$ and $\hat{g}_2(x)$, which are actually estimates for $f(x)$, and by taking the average of the first derivatives $\hat{g}_1'(x)$ and $\hat{g}_2'(x)$ at the corresponding position. Thus, for example,

$$\hat{f}_i' = \frac{1}{2}\{\hat{g}_1'[x_i - u_0(x_i)] + \hat{g}_2'(x_i)\}$$

The restoration of g_1 and g_2 can be based on any of the noise cleaning techniques discussed in Chapter 7. The interpolation, however, again has to take into account the requirements on the smoothness of the optimization function, as discussed earlier. This noise-cleaning step for obtaining reliable estimates for the first derivatives is necessary to achieve consistent results.

However, the determination of the function differences $\Delta g_i = g_2(x_i) - g_1[x_i - u_0(x_i)]$ must be based on the original data, as only then does the error model of white noise hold. The interpolation, which may be needed to obtain $g_1[x_i - u_0(x_i)]$, introduces only negligible correlations (≤ 0.5) between neighboring Δg_i's. If the degree of this smoothing is kept low, we can assume that $\bar{n}(x_i) = n_2(x_i) - n_1[x_i - u_0(x_i)]$ are still white but with variance $\sigma_{\bar{n}}^2 = \sigma_{n_1}^2 + \sigma_{n_2}^2$. For $\sigma_{n_1} = \sigma_{n_2}$ we thus obtain $\sigma_{\bar{n}}^2 = 2\sigma_n^2$. For a constant-shift model $\hat{u}(x_i) = \hat{u}$,

$$\hat{\sigma}_{\bar{n}}^2 = \frac{1}{m-1}\sum_{i-1}^{m}\left[g_2(x_i) - g_1(x_i - \hat{u})\right]^2$$

is now an estimate for the variance of the noise difference between the two profiles and, in the case of $\sigma_{n_1} = \sigma_{n_2}$, can be used to obtain an estimate

$$\sigma_n = \frac{\sigma_{\bar{n}}}{\sqrt{2}}$$

for the noise standard deviation and thus for the observational precision.

16.2.8 Relations to Cross-Correlation Techniques

The earliest applications of intensity-based image matching in remote sensing used cross-correlation techniques. The model assumed simply a shift between two corresponding image sections. To compensate for different brightness and contrast, a linear transformation in the intensity values was assumed, though not always explicitly stated.

We want to discuss the relations between this classical technique and the differential approach, as they give insight into the similarities and differences of both methods. Though only one-dimensional images are treated, the result can be directly transferred to more dimensions. The model reads as

$$g_1(x_i) = f(x_i) + n_1(x_i)$$
$$g_2(x_i) = f(x_i - u) + n_2(x_i)$$

or

$$g_2(x_i) = g_1(x_i - u) + \bar{n}(x_i)$$

with

$$\overline{n}(x_i) = n_2(x_i) - n_1(x_i)$$

The principle of estimating u is to search for the maximal correlation coefficient ρ_{12} of g_1 and g_2 :

$$\max_u \frac{Cov\left[g_1(x_i - u), g_2(x_i)\right]}{\sqrt{V\left[g_1(x_i - u)\right] \cdot V[g_2(x_i)]}} = \frac{\sigma_{g_1 g_2}(u)}{\sigma_{g_1}(u)\sigma_{g_2}} = \rho_{12}(u) \to \hat{u} \tag{16.34}$$

with

$$\begin{aligned}
\sigma_{g_1 g_2}(u) &= \frac{1}{m-1}\left[\sum_{i=1}^{m} g_1(x_i - u)g_2(x_i) - \frac{1}{m}\sum_{i=1}^{m} g_1(x_i - u)\sum_{i=1}^{m} g_2(x_i)\right] \\
\sigma_{g_1}^2(u) &= \frac{1}{m-1}\left[\sum_{i=1}^{m} g_1^2(x_i - u) - \frac{1}{m}\left(\sum_{i=1}^{m} g_1(x_i - u)\right)^2\right] \\
\sigma_{g_2}^2 &= \frac{1}{m-1}\left[\sum_{i=1}^{m} g_2^2(x_i) - \frac{1}{m}\left(\sum_{i=1}^{m} g_2(x_i)\right)^2\right]
\end{aligned} \tag{16.35}$$

As the empirical mean and variances are taken into account, different brightness and contrast are immediately compensated for. Thus cross-correlation corresponds to least-squares matching, using the model of Section 16.2.4. Observe that the cross-correlation term here is the product moment cross-correlation coefficient from statistics, whereas the cross-correlation term as often used in electrical engineering refers to the mean product of two signals and is not normalized with respect to the mean and the variance.

The search for the estimate $\hat{u}$ in general leads to an integer position. The rounding error is dominant in most applications; therefore an interpolation of the correlation function is useful. Let the integer position of the maximum of $\rho(u)$ be u_0 and the two neighboring positions be u_- and u_+, and let the corresponding values of $\rho(u)$ be ρ_0, ρ_- and ρ_+. Then using the quadratic interpolation of $\rho(u)$ leads to the following estimate $\hat{u}$:

$$\hat{u} = u_0 - \frac{\rho'(u_0)}{\rho''(u_0)} = u_0 - \frac{\frac{1}{2\Delta x}(\rho_+ - \rho_-)}{\frac{1}{\Delta x^2}(\rho_+ - 2\rho_0 + \rho_-)} = u_0 - \frac{\rho_+ - \rho_-}{2(\rho_+ - 2\rho_0 + \rho_-)} \cdot \Delta x \tag{16.36}$$

where Δx is the spacing between the u_i.

Thus subpixel accuracy is achievable. Not only was this initially claimed by Bernstein (1973), but it was consistently proved empirically and theoretically by many researchers. The correlation coefficient was used as an acceptance criterion for the problem of setting a proper threshold to reject bad matches. Also, the weakness of the correlation coefficient leads to the use of additional measures of performance, such as the "drop" or "slope" of the correlation function (Helava, 1976; Panton, 1978). This was used to prevent matches with a very flat correlation function from passing.

EXAMPLE 16.5

With the data from Example 16.1, we want to estimate the shift by using cross-correlation and to compare it with the results of the model, adding further parameters for the differences in brightness and contrast. We therefore take the same set of observed values $g^{(4)}$ to $g^{(11)}$ and correlate them with windows of size 8 for f, thus setting $g_1 = g$ and $g_2 = f$ in Eq. (16.35). With the variances $\sigma_g^2 = \sigma_{g_2}^2 = 179.98[gr^2]$ and $\sigma_f^2(u) = \sigma_{g_1}^2(u)$, the covariances $\sigma_{g_1g_2}^2(u) = \sigma_{fg}(u)$, and the correlation coefficients $\rho_{12}(u)$ in the table

u	$\sigma_{g_1}^2(u)$	$\sigma_{g_1g_2}(u)$	$\rho_{12}(u)$
3	183.9	160.5	0.8823
2	200.0	187.9	0.9901
1	183.9	175.2	0.9628
0	135.7	128.2	0.8294
−1	55.4	64.1	0.6423
−2	12.5	20.9	0.4405

we obtain the optimal integer position $u_0 = 2$. The subpixel estimate is $\hat{u} = 2 - (.9628 - 0.8823)/(0.8823 - 2 \cdot 0.9901 + 0.9628)/2 = 2 - 0.30 = 1.70[\Delta x]$ (Eq. 16.36), which is close to the estimate $1.61[\Delta x]$ in the second iteration of Example 16.4.

We now relate these measures to those developed in the context of the differential approach. First let us assume that the two images are not shifted or that the optimal shift has been applied to g_2. Then assuming

$$g_1(x) = f(x) + n_1(x)$$
$$g_2(x) = a\left[f(x) + n_2(x)\right] + b$$

where n_1 and n_2 are independent white noise with variance $\sigma_{n_1} = \sigma_{n_2} = \sigma_n$ and f is stochastic with variance σ_f^2 and independent from n_1 and n_2, we obtain the variances

$$\sigma_{g_1}^2 = \sigma_f^2 + \sigma_n^2$$
$$\sigma_{g_2}^2 = a^2(\sigma_f^2 + \sigma_n^2)$$

and the covariance

$$\sigma_{g_1g_2} = a \cdot \sigma_f^2$$

and therefore the correlation coefficient ρ_{12} (cf. Ballard and Brown, 1982)

$$\rho_{12} = \frac{\sigma_{g_1g_2}}{\sigma_{g_1}\sigma_{g_2}} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_n^2} \tag{16.37}$$

or, with the signal-to-noise ratio,

$$\mathrm{SNR}^2 = \frac{\sigma_f^2}{\sigma_n^2}$$

$$\rho_{12} = \frac{\mathrm{SNR}^2}{\mathrm{SNR}^2 + 1} \tag{16.38}$$

or

$$\mathrm{SNR}^2 = \frac{\rho_{12}}{1 - \rho_{12}} \tag{16.39}$$

Thus the maximum achievable correlation coefficient is limited by the signal-to-noise ratio. On the other hand, we can derive the signal-to-noise ratio from the empirical correlation coefficient. Moreover, if we know the variance $\sigma_{g_1}^2$ of the observed signal, we can derive the noise variance

$$\sigma_n^2 = \sigma_{g_1}^2(1 - \rho_{12}) = \sigma_f^2 \frac{1 - \rho_{12}}{\rho_{12}} \tag{16.40}$$

As the correlation coefficient is identical to the cosine of the angle γ between the vectors g_1 and g_2, we have three equivalent measures for evaluating the intensity of two signals. Examples are:

$$\begin{array}{lll} \mathrm{SNR} = 10 & \rho = 0.99 & \gamma = 8^\circ \\ \mathrm{SNR} = 3.4 & \rho = \sqrt{2}/2 & \gamma = 45^\circ \\ \mathrm{SNR} = 1 & \rho = 0.5 & \gamma = 60^\circ \end{array} \tag{16.41}$$

Values of ρ smaller than 0.5 thus correspond to a signal-to-noise ratio less than 1. Reasonable critical values for ρ_{12} lie in the range between 0.5 and 0.7.

If we use Eqs. (16.37) and (16.39), we can rewrite the variance of the estimated shift from Eq. (16.11):

$$\sigma_{\hat{u}}^2 = \frac{1}{m} \cdot \frac{1}{\mathrm{SNR}^2} \cdot \frac{\sigma_f^2}{\sigma_{f'}^2} = \frac{1}{m} \cdot \frac{1 - \rho_{12}}{\rho_{12}} \cdot \frac{\sigma_f^2}{\sigma_{f'}^2} \tag{16.42}$$

It now becomes obvious that the correlation coefficient gives only partial information about the precision of a match: the number of points used—thus the window size and the ratio $\sigma_f^2/\sigma_{f'}^2$—is not used.

This ratio $\sigma_{f'}^2/\sigma_f^2$ relating the "edge busyness" $\sigma_{f'}^2$ to the variance of the signal can be shown to be proportional to the effective bandwidth of the signal, and, owing to the moment theorem, equal to the (negative) *curvature* of the autocorrelation function (ACF) of $f(x)$ (Papoulis, 1965; Ryan, Gray, and Hunt, 1980). Thus two profiles with the same length and the same variance may lead to significantly different matching precisions (Fig. 16.6). The standard deviation $\sigma_x(B)$ obtainable with profile B is four times higher than the standard deviation $\sigma_x(A)$ with profile A : $\sigma_x(B) = 4 \cdot \sigma_x(A)$.

Using the second derivative $\rho''(u) = (\rho_+ - 2\rho_0 + \rho_-)/\Delta x^2$ of the empirical correlation function $\rho(u)$ (cf. Eq. 16.34), we obtain an expression for the variance

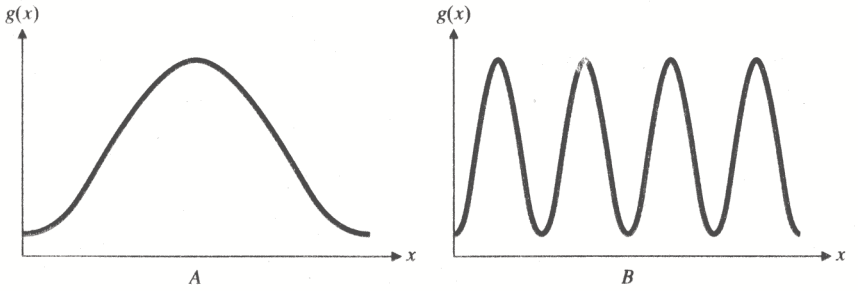


Figure 16.6 Two profiles of common contrast but different sharpness. If one assumes a constant noise variance, matching with profile A leads to four times ~~smaller~~ larger a standard deviation of the estimated shift than does matching with profile B.

of estimated shift

$$\sigma_x^2 = \frac{1}{m} \cdot \frac{1-\rho_0}{\rho_0} \cdot \frac{\Delta x^2}{-\rho_+ + 2\rho_0 - \rho_-} \tag{16.43}$$

with the pixel spacing Δx. This may be used for evaluating the result of correlation techniques and, of course, holds if the geometric model of a pure shift is approximately valid and the position $\hat{u}$ has been estimated by some interpolation of the cross-correlation function (cf. Eq. 16.36).

EXAMPLE 16.6

The result of the correlation in Example 16.5 can now be evaluated. The standard deviation of the estimated shift according to Eq. (16.43) results in $\sigma_{\hat{u}}^2 = 1/8 \cdot (1 - 0.9901)/0.9901/(-0.8823 + 2 \cdot 0.9901 - 0.9628) = 0.009251$, and thus $\sigma_{\hat{u}} = 0.096[\Delta x]$, which seems a bit too optimistic. The estimated noise standard deviation, according to Eq. (16.40), $\sigma_n = 1.41[gr]$, is very close to the one ($1.43[gr]$) obtained in the second iteration of Example 16.4. The difference probably results from the weak estimate of the curvature of the cross-correlation function from only three points, which likely is biased toward too high a curvature (Berman, 1989).

16.3 Intensity-Based Matching of Two-Dimensional Signals

The differential techniques for intensity-based matching can be directly transferred to the case of two-dimensional functions $f(r,c)$ and $g(r,c)$.

16.3.1 The Principle and the Relation to Optical Flow

In the most general case of matching an observed intensity function to a given one, we start from the nonlinear model

$$g(r_i, c_i) = f(p_i, q_i) + n(r_i, c_i) \qquad i = 1, \ldots, m \tag{16.44}$$

with

$$\begin{pmatrix} p \\ q \end{pmatrix}_i = \begin{pmatrix} r \\ c \end{pmatrix}_i - \begin{pmatrix} \hat{u}(r,c) \\ \hat{v}(r,c) \end{pmatrix}_i \tag{16.45}$$

again assuming that the object is given at a grid, together with an appropriate interpolation scheme, and that the noise is white with variance σ_n^2.

If we assume that approximate values $u_0(r_i, c_i)$ of the unknown *deformation* $[\hat{u}(r,c), \hat{v}(r,c)]$ are known, we obtain the linearized model

$$\Delta g_i = -f_{r,i} \cdot \widehat{\Delta u}_i - f_{c,i} \cdot \widehat{\Delta v}_i + n_i \tag{16.46}$$

with

$$\Delta g_i = g(r_i, c_i) - f\left[r_i - u_0(r_i, c_i), c_i - v_0(r_i, c_i)\right]$$

$$f_{r_i} = \left.\frac{\partial f(p,q)}{\partial p}\right|_{p=r_i-u_0(r_i,c_i), q=c_i-v_0(r_i,c_i)}$$

$$f_{c_i} = \left.\frac{\partial f(p,q)}{\partial q}\right|_{p=r_i-u_0(r_i,c_i), q=c_i-v_0(r_i,c_i)}$$

$$\widehat{\Delta u}_i = \hat{u}(r_i, c_i) - u_0(r_i, c_i)$$
$$\widehat{\Delta v}_i = \hat{v}(r_i, c_i) - v_0(r_i, c_i)$$
$$n_i = n(r_i, c_i)$$

Here we want to relate Eq. (16.46) to the optical flow equation used in motion analysis.

If one treats f as a time-varying intensity field $f(r,c,t)$ with the velocity field $[u(r,c), v(r,c)]$, the nonlinear model equation (16.44) can be written as

$$f\left[r + u(r,c) \cdot dt,\ c + v(r,c) \cdot dt,\ t + dt\right] = f(r,c,t)$$

with the the noise term omitted, thus referring to the expected value of g. If we now linearize, for the sake of simplicity at $u = v = 0$, we obtain

$$f(r,c,t) + \frac{\partial f(r,c,t)}{\partial r} \cdot u(r,c) \cdot dt + \frac{\partial f(r,c,t)}{\partial c} \cdot v(r,c) \cdot dt + \frac{\partial f(r,c,t)}{\partial t} dt = f(r,c,t)$$

Setting $\nabla f = (f_r, f_c)^T$ the spatial gradient of f, $V = (u,v)^T$ the velocity field, and $\dot{f} = \partial f / \partial t$ the temporal change of f, we obtain

$$\nabla f \cdot V + \dot{f} = 0$$

the *optical flow equation*.

The equation simply states that the intensity *of corresponding points* is time invariant, and changes in f are due only to V. This relation approximately holds for image sequences but neither at occlusions nor if the time spacing Δt is large. While occlusions cannot explicitly be handled with the model equation (16.44), changes in illumination or reflectance may be taken into account by additional parameters.

Now Eq. (16.46) does not allow us to estimate $\Delta u(r_i,c_i)$ and $\Delta v(r_i,v_i)$ simultaneously, as only one observation—namely, Δg_i—is available. This so-called *aperture problem* can be solved only by adding constraints on the deformation field $V(r,c) = [u(r,c),\ v(r,c)]$. Several stages of constraints are of practical use here:

1. $V(r,c)$ is constant. This is the classical assumption used also for cross-correlation (Barnea and Silverman, 1972).
2. $V(r,c)$ is a linear function in r,c. This would be a first-order approximation of V, which for smooth deformations can be used for local estimates. The first attempts to exceed a shift were made by Schalhoff and McVey (1979) and Huang (1981) (cf. the review given by Förstner, 1984).
3. $V(r,c)$ is smooth. This assumption allows one to model real situations, as in satellite imagery, medical imagery, or reconstruction of smooth surfaces.
4. $V(r,c)$ most generally is best modeled to be piecewise smooth. Then occlusions may also be handled, though no explicit reference is made to the three-dimensional structure in the model.
5. If image matching is used to evaluate stereo images; if the relative orientation of the cameras, including their interior orientation, is known; and if the rows of both images are parallel to the baseline, that is, the line through the two projection centers, then $u(r,c) \equiv 0$ for all r,c. Thus only $v(r,c)$ is to be estimated, which reduces the problem to profile matching, discussed in the previous section. Now, however, the matching can be based on two-dimensional windows, thereby stabilizing the estimates (cf. Section 16.6.3).

We want to treat only the first two cases—those with unknown contrast shift and unknown linear transformation—but will include parameters for differences in brightness or contrast. In addition, we discuss two special cases in order to motivate the interest operator to select distinct points. The more general problem of estimating a complete displacement field was treated in Chapter 15.

16.3.2 Estimating Constant-Shift Parameters

The simplest model for matching is

$$p = x - u, \qquad q = y - v$$

This assumes solely a shift (u,v) between the two windows of concern.

The linearized model then reads as

$$\Delta g_i = -f_{r_i} \cdot \widehat{\Delta u} - f_{c_i} \cdot \widehat{\Delta v} + n_i, \quad i = 1,\ldots,m \tag{16.47}$$

with

$$f_{r_i} = \left.\frac{\partial f(p,q)}{\partial r}\right|_{\substack{r=r_i-u_0 \\ c=c_i-v_0}}$$

$$f_{c_i} = \left.\frac{\partial f(p,q)}{\partial c}\right|_{\substack{r=r_i-u_0 \\ c=c_i-v_0}}$$

$$\begin{aligned}
\widehat{\Delta u} &= \hat{u} - u_0, \\
\widehat{\Delta v} &= \hat{v} - v_0, \\
\Delta g_i &= g(r_i,c_i) - f(r_i - u_0, c_i - v_0) \\
u_i &= u(r_i,c_i)
\end{aligned}$$

for all m pixels (r_i,c_i) in the window. In most cases the points (r_i,c_i) form a grid with $m = m_r m_c$ points, but Eq. (16.47) also can be used if the points do *not* form a grid.

The normal equations are then given by

$$N\hat{y} = h \tag{16.48}$$

with the 2×2 normal equation matrix N

$$N = \begin{pmatrix} \sum_{i=1}^{m} f_{r_i}^2 & \sum_{i=1}^{m} f_{r_i} f_{c_i} \\ \sum_{i=1}^{m} f_{r_i} f_{c_i} & \sum_{i=1}^{m} f_{c_i}^2 \end{pmatrix} = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{21} \end{pmatrix} \tag{16.49}$$

and the right-hand sides

$$h = -\begin{pmatrix} \sum_{i=1}^{m} f_{r_i} \cdot \Delta g_i \\ \sum_{i=1}^{m} f_{c_i} \cdot \Delta g_i \end{pmatrix} \tag{16.50}$$

and $\hat{y} = \begin{pmatrix} \widehat{\Delta u} \\ \widehat{\Delta v} \end{pmatrix}$, which can be resolved for $\hat{y}$, for example, by using the weight coefficient matrix of the unknown parameters $\hat{y}$

$$Q = N^{-1} = \frac{1}{N_{11}N_{22} - N_{12}^2} \begin{pmatrix} N_{22} & -N_{12} \\ -N_{12} & N_{11} \end{pmatrix}$$

leading to the estimates $\hat{u} = u_0 + \widehat{\Delta u}$ and $\hat{v} = v_0 + \widehat{\Delta v}$. We also obtain an estimate for the unknown variance factor, identical here to the noise variance

$$\sigma_n^2 = \frac{1}{m-2} \cdot \sum_{i=1}^{m} [g(r_i,c_i) - f(r_i - \hat{u}, c_i - \hat{v})]^2$$

This can then be used to obtain the covariance matrix $\Sigma_{\hat{y}\hat{y}}$ of the unknown shift $\hat{u}$ and $\hat{v}$

$$\Sigma_{\hat{y}\hat{y}} = \hat{\sigma}_n^2 \cdot Q = \hat{\sigma}_n^2 \cdot \begin{pmatrix} \sum_{i=1}^{m} f_{ri}^2 & \sum_{i=1}^{m} f_{ri} f_{ci} \\ \sum_{i=1}^{m} f_{ri} f_{ci} & \sum_{i=1}^{m} f_{ci}^2 \end{pmatrix}^{-1} \tag{16.51}$$

from which follow the standard deviations

$$\hat{\sigma}_{\hat{u}} = \hat{\sigma}_n \cdot \sqrt{Q_{11}} = \hat{\sigma}_n \cdot \sqrt{\frac{N_{22}}{N_{11}N_{22} - N_{12}^2}}$$

and

$$\hat{\sigma}_{\hat{v}} = \hat{\sigma}_n \cdot \sqrt{Q_{22}} = \hat{\sigma}_n \cdot \sqrt{\frac{N_{11}}{N_{11}N_{22} - N_{12}^2}}$$

The estimates are correlated with correlation coefficient

$$\rho_{\hat{u}\hat{v}} = \frac{-N_{12}}{\sqrt{N_{11}N_{22}}}$$

which could be used to further evaluate the estimates.

If we use estimates

$$\sigma_{fr}^2 \doteq \frac{1}{m} \sum_{i=1}^{m} f_{ri}^2 = \frac{N_{11}}{m}$$

$$\sigma_{fc}^2 \doteq \frac{1}{m} \sum_{i=1}^{m} f_{ci}^2 = \frac{N_{22}}{m}$$

and

$$\sigma_{frfc}^2 \doteq \frac{1}{m} \sum_{i=1}^{m} f_{ri} f_{ci} = \frac{N_{12}}{m}$$

for the *local squared gradient,* or

$$\widehat{\nabla f \nabla f}^t = \begin{pmatrix} \sigma_{fr}^2 & \sigma_{frfc} \\ \sigma_{fcfr} & \sigma_{fc}^2 \end{pmatrix} \tag{16.52}$$

we can rewrite Eq. (16.51) as

$$\Sigma_{\hat{y}\hat{y}} = \frac{\hat{\sigma}_n^2}{m} \cdot \begin{pmatrix} \sigma_{fr}^2 & \sigma_{frfc} \\ \sigma_{fcfr} & \sigma_{fc}^2 \end{pmatrix}^{-1} \tag{16.53}$$

Thus in addition to the estimates $\hat{u}$ and $\hat{v}$, we obtain a measure for the (internal) precision of the estimates. This precision intuitively depends on the following factors:

1. *The noise variance* σ_n^2. If we have good a priori knowledge about σ_n^2, we can introduce it instead of the estimated value $\hat{\sigma}_n^2$. As the model Eq. (16.47) will in many cases be oversimplified, we will not generally use $\hat{\sigma}_n^2$.

2. *The number m of used pixels.* Thus the standard deviation decreases linearly with the width of a square window. This holds only as long as new information is collected when increasing the window, and thus only as long as the average gradient in the window remains constant. Otherwise Eq. (16.51) demonstrates no advantage in using larger windows, as the sums in brackets remain practically constant.
3. *The average squared gradient in the window* ($\widehat{\nabla f \nabla f}^T$). This indicator of the edge busyness is decisive for the precision of the match. We can show that it measures the curvature of the autocovariance (cross-correlation) function of $f(r,c)$, specifically the negative Hessian of the autocorrelation function of $f(r,c)$, if we assume f has zero mean and is thus a direct generalization of the relations to cross-correlation techniques discussed in Section 16.2.8.

The advantage of using Eqs. (16.51) and (16.53) is that, if we assume the noise variance is constant over an entire image, they enable us to determine in advance those places where we can expect high precision—that is, before we actually perform the matching procedure. This is because the equations depend only on the content of *one* window, and not on the actual observations, which would require two windows. Even if both windows are distorted by noise, $f(r,c)$ may be replaced by an adequate estimate from only one of the two windows, also in this case allowing us to check in advance the precision of the match.

Figure 16.7 shows 10 small windows from which the interior 16×16 pixels are used to evaluate the expected precision of a match. The confidence ellipses shown below the template assume a noise standard deviation of $\sigma_n = 5$ gray values. With 99% probability the true shift will lie in the area depicted by the ellipse around the estimated shift. The pixel size is $\Delta x = 20\mu m$. The smallest confidence ellipses have major axes of less than $1\mu m$ and thus less than 1/20 of a pixel. The largest confidence ellipse is obtained at the edge point, indicating that one cannot expect good accuracy for the position along the edge. The other confidence ellipses reflect reasonably well the image content with respect to the expected precision of matching. Thus the covariance matrix, proportional to the inverse of the normal equation matrix, is an ideal measure with which to evaluate the *distinctness* of small windows. As it is a statistical measure, essentially derived by averaging the quadratic gradient, no interpretation of the image content within the window is performed; therefore corners, circles, and random texture are treated simply from the standpoint of distinctness. The interest operator, discussed in Section 16.4, was actually motivated by searching for windows that guarantee good matching accuracy.

16.3.3 Estimating Linear Transformations

In the most important case of matching two small windows, we assume the model

$$g_1(r_i,c_i) = f(r_i,c_i) + n_1(r_i,c_i)$$

$$g_2(r_i,c_i) = h\left[f(p_i,q_i)\right] + n_2(r_i,c_i)$$

$$\begin{pmatrix} p \\ q \end{pmatrix}_i = \begin{pmatrix} a_1 & a_2 \\ a_4 & a_5 \end{pmatrix} \cdot \begin{pmatrix} r \\ c \end{pmatrix}_i + \begin{pmatrix} a_3 \\ a_6 \end{pmatrix} \tag{16.54}$$

$$h(f) = a_7 \cdot f + a_8 \tag{16.55}$$

where the noise n_1 and n_2 in both observed images g_1 and g_2 is white with standard deviations σ_{n_1} and σ_{n_2}, and the eight parameters a_k, $k = 1,\ldots,8$, are unknown. If we assume that the approximate values a_{0k}, $k = 1,\ldots,8$, are known, we obtain the linearized model

$$\begin{aligned} \Delta g_i = \; & f_{r_i} \cdot r_i \cdot \widehat{da}_1 \\ & + f_{r_i} \cdot c_i \cdot \widehat{da}_2 \\ & + f_{r_i} \quad\; \widehat{da}_3 \\ & + f_{c_i} \cdot r_i \cdot \widehat{da}_4 \\ & + f_{c_i} \cdot c_i \cdot \widehat{da}_5 \\ & + f_{c_i} \quad\; \widehat{da}_6 \\ & + f_i \quad\; \widehat{da}_7 \\ & + 1 \quad\; \widehat{da}_8 + \bar{n}_i \end{aligned}$$

with

$$\begin{aligned} \widehat{da}_k &= \hat{a}_k - a_{0k}, \qquad k = 1,\ldots,8 \\ \Delta g_i &= g_2(r_i, c_i) - g_1(p_{0i}, q_{0i}) \\ \bar{n}_i &= n_2(r_i, c_i) - n_1(p_{0i}, q_{0i}) \end{aligned}$$

and estimates for f_{r_i}, f_{c_i}, and f_i, for example,

$$\begin{aligned} f_{r_i} &= \frac{\partial \hat{g}_1(r_i, c_i)}{\partial r} \\ f_{c_i} &= \frac{\partial \hat{g}_1(r_i, c_i)}{\partial c} \\ f_i &= \hat{g}_1(r_i, c_i) \end{aligned}$$

with

$$\begin{pmatrix} p \\ q \end{pmatrix}_{0i} = \begin{pmatrix} a_{01} & a_{02} \\ a_{04} & a_{05} \end{pmatrix} \cdot \begin{pmatrix} r \\ c \end{pmatrix}_i + \begin{pmatrix} a_{03} \\ a_{06} \end{pmatrix}$$

where $\hat{g}_1$—that is, the smoothed version of g_1 (cf. Section 13.2.7)—is restored. Again at least linear interpolation is required for g_1, f_{r_i}, and f_{c_i} in order to avoid unstable solutions.†

†An averaging of g_1 and g_2 and of $\partial g_1/\partial p$ and $\partial g_2/\partial r$, and the like, needs to take the geometric and radiometric transformations into account.

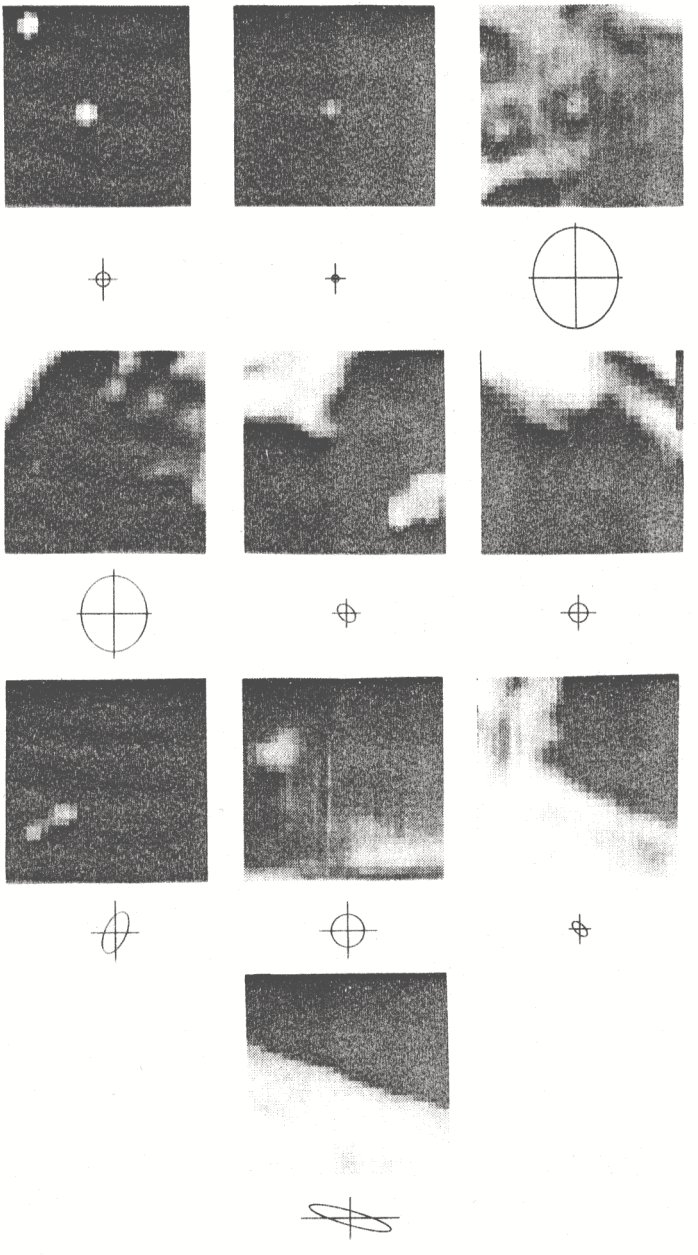


Figure 16.7 Precision of matching to be expected for 10 small windows. Although 32×32 pixels are shown, only the interior 16×16 pixels are assumed to be used. The noise standard deviation is assumed to be $\sigma_n = 5[gr]$. The 99% confidence ellipses given in μm refer to a pixel size of 20μm. Observe the extremely high precision to be expected and the elongated ellipse at the edge.

We then have the following normal equation matrix N:

$$\left(\begin{array}{ccc|ccc|cc}
\Sigma f_r^2 r^2 & \Sigma f_r^2 rc & \Sigma f_r^2 r & \Sigma f_r f_c r^2 & \Sigma f_r f_c rc & \Sigma f_r f_c r & \Sigma f_r fr & \Sigma f_r r \\
 & \Sigma f_r^2 c^2 & \Sigma f_r^2 c & \Sigma f_r f_c cr & \Sigma f_r f_c c^2 & \Sigma f_r f_c c & \Sigma f_r fc & \Sigma f_r c \\
 & & \Sigma f_r^2 & \Sigma f_r f_c r & \Sigma f_r f_c c & \Sigma f_r f_c & \Sigma f_r f & \Sigma f_r \\
\hline
 & & & \Sigma f_c^2 r^2 & \Sigma f_c^2 rc & \Sigma f_c^2 r & \Sigma f_c f_r r & \Sigma f_c r \\
 & & & & \Sigma f_c^2 c^2 & \Sigma f_c^2 c & \Sigma f_c fc & \Sigma f_c c \\
 & & & & & \Sigma f_c^2 & \Sigma f_c f & \Sigma f_c \\
\hline
 & & & & & & \Sigma f^2 & \Sigma f \\
 & & & & & & & \Sigma\, 1
\end{array}\right) \tag{16.56}$$

omitting the index i for convenience and solving the sums over all pixels (r_i, c_i) in the windows. The right-hand side h is given by

$$h = \begin{pmatrix}
\sum_{i=1}^{m} f_{r_i} r_i \Delta g_i \\
\sum_{i=1}^{m} f_{r_i} c_i \Delta g_i \\
\sum_{i=1}^{m} f_{r_i} \Delta g_i \\
- \quad - \quad - \\
\sum_{i=1}^{m} f_{c_i} r_i \Delta g_i \\
\sum_{i=1}^{m} f_{c_i} c_i \Delta g_i \\
\sum_{i=1}^{m} f_{c_i} \Delta g_i \\
- \quad - \quad - \\
\sum_{i=1}^{m} f_i \Delta g_i \\
\sum_{i=1}^{m} 1 \Delta g_i
\end{pmatrix} \tag{16.57}$$

The normal equation system $N\hat{y} = h$ for the eight unknown parameters $\hat{y} = (\widehat{\Delta a_k})$ yields the six corrections Δu_k, $k = 1, \ldots, 6$, to the approximate values of the

geometric transformation and the correction $\widehat{\Delta a_7}$ and $\widehat{\Delta a_8}$ for the radiometric parameter; thus

$$\hat{a}_k = a_{0k} + \widehat{\Delta a_k}, \qquad k = 1, \ldots, 8$$

which may be used as new approximate values in a further iteration.

The main effort consists in building the normal equation matrix. If g_2 is not very noisy, simply a restored version $\hat{g}_1$ or $\hat{g}_2$ could be used for determining the f_i, f_{r_i}, and f_{c_i}, which then need not be updated during an iteration sequence. Only the right-hand sides h of Eq. (16.57) have to be recomputed in each iteration. This speeds up the computation and increases the radius of convergence (Burkhardt and Moll, 1979).

The estimated noise variance

$$\hat{\sigma}_n^2 = \frac{1}{m-8} \cdot \sum_{i=1}^{m} [\hat{a}_7 g_1(\hat{p}_i, \hat{q}_i) + \hat{a}_8 - g_2(\hat{r}_i, \hat{c}_i)]^2 \tag{16.58}$$

with

$$\begin{pmatrix} \hat{p}_i \\ \hat{q}_i \end{pmatrix} = \begin{pmatrix} \hat{a}_1 & \hat{a}_2 \\ \hat{a}_4 & \hat{a}_5 \end{pmatrix} \begin{pmatrix} r_i \\ c_i \end{pmatrix} + \begin{pmatrix} \hat{a}_3 \\ \hat{a}_6 \end{pmatrix}$$

now is a reliable estimate in many cases, especially if windows that are not too large or not too small are used (e.g., between 9×9 and 31×31), and if the distortions between the two images are smooth enough to be modeled by a linear transformation. The two parameters compensating for brightness and contrast are sufficient but also necessary in most practical cases.

The inverse Q of the normal equation matrix may be used to yield the covariance matrix and from there the standard deviation $\sigma_{\hat{a}_k}$ of the estimate

$$\hat{\sigma}_{\hat{a}_k} = \hat{\sigma}_n \cdot \sqrt{Q_{kk}}$$

thus also specifying the precision of the estimated scale, rotation, and shears.

The small windows, however, often do not contain enough detail to enable one to determine all eight parameters. Especially the scales (a_1, a_5, and a_7) and the shears (a_2 and a_4) are frequently not estimable. Therefore a priori knowledge about the transformation may be introduced in a Bayesian manner by using additional observations (possibly fictitious ones)

$$da_k = \widehat{da}_k + n_{a_k}, \qquad w_{a_k} = \frac{\sigma_n^2}{\sigma_{a_k}^2}, \qquad k = 1, \ldots, 8$$

with individual weights depending on the quality, specifically the standard deviations σ_{a_k} of the corrections to the a priori values da_k. This leads to the modified and stabilized normal equation system

$$[N + \text{Diag}\,(w_{a_k})] \cdot \hat{y} = h$$

The right-hand sides h, because of the corrections assumed to be $da_k = 0$, remain unchanged. The following standard deviations can be recommended:

Scales, shears	$\sigma_{a_k} = 0.1 - 1$	$k = 1,2,4,5,7$
Geometric shifts	$\sigma_{a_k} = 1 - 10$ [pixel]	$k = 3,6$
Radiometric shift	$\sigma_{a_k} = 10 - 100$ [gray value]	$k = 8$

The noise standard deviation σ_n has to be estimated or guessed. The result is not too sensitive against errors of a factor 2 in these assumed standard deviations.

16.3.4 Invariant Points

Much as in Section 16.2.3 (Eq. 16.16), we now want to find points within the window that, when transferred into the other window, are invariant with respect to changes or errors in scale or rotation. To simplify the derivation, we restrict the analysis to three parameters.

The first model, including two shifts and scale, yields the symmetric normal equation matrix

$$N = \begin{pmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \\ N_{31} & N_{32} & N_{33} \end{pmatrix} = \begin{pmatrix} \sum_i f_{r_i}^2 & \sum_i f_{r_i} f_{c_i} & \sum_i f_{r_i}(f_{r_i}\bar{r}_i + f_{c_i}\bar{c}_i) \\ & \sum_i f_{c_i}^2 & \sum_i f_{c_i}(f_{r_i}\bar{r} + f_{c_i}\bar{c}_i) \\ & & \sum_i (f_{r_i}\bar{r}_i + f_{c_i}\bar{c}_i)^2 \end{pmatrix} \tag{16.59}$$

using $\bar{r}_i = r_i - r_0$ and $\bar{c}_i = c_i - c_0$ with the unknown reference point (r_0, c_0). If we require the estimates $\hat{u}$ and $\hat{v}$ of the two shifts u and v to be independent on the estimated scale, then we have to determine r_0 and c_0 from

$$N_{13} = 0 = \sum f_{r_i}\left[f_{r_i}(r_i - r_0) + f_{c_i}(c_i - c_0)\right]$$
$$N_{23} = 0 = \sum f_{c_i}\left[f_{r_i}(r_i - r_0) + f_{c_i}(c_i - c_0)\right]$$

leading to the 2×2 system of equations

$$\begin{pmatrix} \sum_i f_{r_i}^2 & \sum_i f_{r_i} f_{c_i} \\ \sum_i f_{r_i} f_{c_i} & \sum_i f_{c_i}^2 \end{pmatrix} \begin{pmatrix} r_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} \sum_i (f_{r_i}^2 r_i + f_{r_i} f_{c_i} c_i) \\ \sum_i (f_{r_i} f_{c_i} r_i + f_{c_i}^2 c_i) \end{pmatrix} \tag{16.60}$$

for (r_0, c_0). This point has several important properties:

- If only the shift parameters $(\hat{r}, \hat{c})$ would have been determined, but the contents of the windows of concern are different in scale, then the shift should be applied to (r_0, c_0) in $g(r, c)$, as the transferred point $(r_0 - \hat{r}, c_0 - \hat{c})$ then would not be biased owing to scale differences.
- The transferred point $(r_0 - \hat{r}, c_0 - \hat{c})$ has the minimum variance (cf. Eq. 16.17), provided the model with three parameters is adequate.
- The invariance of (r_0, c_0) with respect to scale differences especially holds for *corner points* (cf. Fig. 16.8a, b), for example, points where several (≥ 2) edges meet. This is because (in the noiseless case) all gradients $\nabla f_i = (f_{r_i}, f_{c_i})$ are orthogonal to $(r_i - r_0, c_i - c_0)$, leading to the condition $f_{r_i}\bar{r}_i + f_{c_i}\bar{c}_i = 0$ and thus to $N_{13} = N_{23} = 0$. This indicates that the estimates for the shifts and the scale are uncorrelated. But then also $N_{33} = 0$ (cf. Eq. 16.59), from which it is clear that the scale cannot be determined. The condition mentioned above was also used by Negahdaripour and Horn (1989)

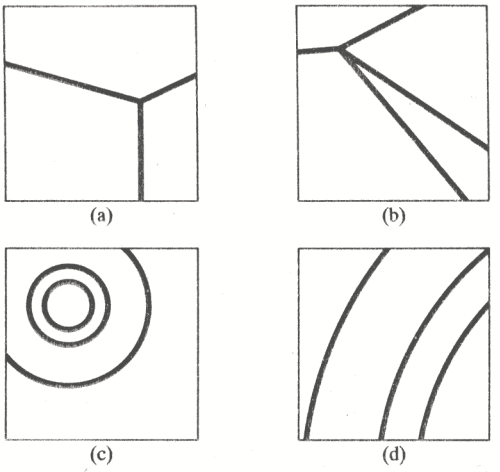


Figure 16.8 Two pairs of images where a scale difference (a,b) and a rotation (c,d) cannot be determined or can be only weakly determined. The points (r_0, c_0) and (r_0', c_0') can be determined directly from eqs.(16.60) and (16.61). Estimating the shifts (r,c) without additional geometric parameters for scale or rotation (e.g., by two-dimensional cross-correlation) and applying the shifts to these points leads to unbiased transferred points.

to estimate the focus of expansion, leading to the same normal equation system, Eq. (16.60).

The same reasoning can be followed to determine the point $(r_o^\bullet, c_0^\bullet)$ in the window that, when transferred to the other window, is invariant with respect to differences in *rotation*. If we want to estimate an unknown rotation in addition to unknown shifts $(\hat{r}, \hat{c})$, the nonlinear model

$$g(r_i, c_i) = f(\cos\phi \cdot r_i + \sin\phi \cdot c_i - u, -\sin\phi \cdot r_i + \cos\phi \cdot c_i - v) + n(r_i, c_i)$$

with $\phi \approx 0$ in linearized form, reads as

$$g(r_i, c_i) = f(r_i + \phi c_i - u, -\phi r_i + c_i - v) + n(r_i, c_i)$$

This leads to the 3×3 normal equation matrix

$$N = \begin{pmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \\ N_{31} & N_{32} & N_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_i f_{r_i}^2 & -\sum_i f_{r_i} f_{c_i} & -\sum_i f_{r_i}(f_{r_i}\bar{c}_i - f_{c_i}\bar{r}_i) \\ & \sum_i f_{c_i}^2 & \sum_i f_{c_i}(f_{r_i}\bar{c} - f_{c_i}\bar{r}_i) \\ & & \sum_i (f_{r_i}\bar{c}_i - f_{c_i}\bar{r}_i)^2 \end{pmatrix}$$

For the shifts (r,c) to be stochastically independent from ϕ, the conditions for $(r_0^\bullet, c_0^\bullet)$ now are

$$N_{13} = 0 = \sum f_{r_i}\left[f_{r_i}(c_i - c_0^\bullet) - f_{c_i}(r_i - r_0^\bullet)\right]$$
$$N_{23} = 0 = \sum f_{c_i}\left[f_{r_i}(c_i - c_0^\bullet) - f_{c_i}(r_i - r_0^\bullet)\right]$$

Therefore $(r_0^\bullet, c_0^\bullet)$ in this case can be estimated from the 2×2 equation system:

$$\begin{pmatrix} \sum_i f_{c_i}^2 & -\sum_i f_{r_i} f_{c_i} \\ -\sum_i f_{r_i} f_{c_i} & \sum_i f_{r_i}^2 \end{pmatrix} \begin{pmatrix} r_0^\bullet \\ c_0^\bullet \end{pmatrix} = \begin{pmatrix} \sum_i (f_{c_i}^2 \cdot r_i - f_{r_i} f_{c_i} \cdot c_i) \\ \sum_i (-f_{r_i} f_{c_i} \cdot r_i + f_{r_i}^2 \cdot c_i) \end{pmatrix} \tag{16.61}$$

Also, this point has the following important properties:

- The transferred point $(r_0^\bullet - \hat{r}, c_0^\bullet - \hat{c})$ among all others has the highest precision, provided the three-parameter model holds.
- If one wishes to determine the shift parameters (r,c) from the two windows without taking a rotation into account, the transferred point $(r_0^\bullet - \hat{r}, c_0^\bullet - \hat{c})$ is unbiased, that is, invariant to possible rotations between the two windows around $(r_0^\bullet, c_0^\bullet)$ and $(r_0^\bullet - \hat{r}, c_0^\bullet - \hat{c})$, respectively.
- The invariance of $(r_0^\bullet, c_0^\bullet)$ with respect to rotation differences especially holds for the centers of figures, which are circularly symmetric with respect to $(r_0^\bullet, c_0^\bullet)$ (cf. Fig. 16.8c, d). The gradients are expected to be parallel to (r_i, c_i), leading to the condition $f_{r_i}\bar{c}_i - f_{c_i}\bar{r}_i = 0$. Therefore not only does $N_{13} = N_{23} = 0$, proving the independence of the shift estimates and the rotation, but also $N_{33} = 0$, showing that the rotation is not determinable.

We will use these relations and the similarity of the normal equation matrices (16.49), (16.60), and (16.61) as a basis for the interest operator discussed next.

16.4 An Interest Operator

16.4.1 Introduction

Image matching as well as general image analysis may require one to find interesting points in the image. *Interesting* here has several meanings, depending on the context:

1. *Distinctness.* Points should be distinct, that is, distinguishable from immediate neighbors. This definition especially excludes points sitting on the same edge. Distinct points may be corners, blobs, highly textured places, and so on. One way to measure distinctness is to compare the intensity function within a window at a point with the intensity function of all surrounding windows. One

could, for example, use the correlation coefficient. If the maximum of the correlation coefficients of the point with its neighbors is small, then the point is dissimilar to all neighbors and is thus a distinct point. This condition is identical to requiring the autocorrelation function at a point to be peaked or to show a high curvature in all directions.

2. *Invariance.* The position as well as the selection of the interesting point should be invariant with respect to the expected geometric and radiometric distortions, which may include robustness with regard to gross or unexpected errors. Invariance and distinctness obviously are the main properties that interesting points should have, as they influence all subsequent steps in the analysis.
3. *Stability.* The position as well as the selection should be invariant with respect to viewing excluding "virtual interestingness." This is to ensure that interesting points in the image correspond to interesting points in the object. For example, corner points of polyhedra can be assumed stable, while T-junctions usually are unstable, as they almost always result from occlusions. Stability thus is decisive for image-matching three-dimensional reconstruction as well as for general image analysis tasks.
4. *Uniqueness.* Whereas distinctness aims at local separability, uniqueness aims at global, that is, imagewide, separability. This is to avoid locally distinct but repetitive features or points that confuse or at least slow down many matching procedures. Unique points thus may significantly increase the reliability of the results of matching and analysis procedures. Uniqueness is probably the notion closest to *interestingness* and the reason to use the term *interest operator* for procedures extracting such points.
5. *Interpretability.* While the previous notions specifying interestingness can be used for both matching and image analysis, we may in addition require the extracted points to have a meaning with respect to image interpretation. Such points then may be corners, junctions of lines, centers of circles, rings, and so on.

Chapter 7 discussed several interest operators for detecting corners. Some, like Moravec's operator (Moravec, 1980; Thorpe, 1983) do not locate corners precisely enough; others rely on a geometric description of the intensity surface and thus are not able to handle corners with more than two edges meeting. All techniques for finding corners of polyhedra with three or even more edges meeting require extracting edges before grouping them in order to obtain corner points. Techniques for finding the centers of circles follow the same type of procedure.

Here we discuss an interest operator that practically fulfills all the criteria mentioned above and contains the extraction procedures as a special case (Paderes, Mikhail, and Förstner, 1984; Förstner, 1986a; Förstner and Gülch, 1987). It follows a three-step procedure:

1. *Selection of optimal windows.* The selection is based on the average gradient magnitude within a window of prespecified size. Searching for local maxima, while suppressing windows on edges, guarantees (local) distinctness. The measure used is also invariant with respect to rotation.

2. *Classification of the image function within the selected windows.* The classification distinguishes between types of singular points such as corners, rings, and spirals, on one hand, and isotropic texture, on the other. Excluding spirals here, a classification of corners, rings, and general texture, based on a statistical test, is available.
3. *Estimation of the optimal point within the window as the classification.* The estimation is precise for corners and for the centers of circular symmetric features or spirals.

The interest operator has several salient features:

1. The selection of the windows is *optimal* with respect to the following tasks:
   - Find windows that guarantee optimal precision for matching (cf. Section 16.4.4);
   - Find corners;
   - Find centers of circular symmetric features;
   - Find centers of logarithmic spirals (possibly).
2. The selection of the windows is the same for all these tasks and needs no a priori knowledge of the number of edges meeting at a corner or the number of rings.
3. The corner points are *invariant* with respect to *rotations* of a polyhedron in three-dimensions around that corner. In addition, the operator is scale invariant at corners. This is probably the most important property.
4. The *decision* with respect to corner, circular symmetric features, and texture can be based on an *F-test*.
5. The estimation of the optimal point within the window can be represented as a least-squares fit that allows a *rigorous evaluation* of its precision. Specifically, one can derive a covariance matrix for the located point, which may be used in further steps of a geometric analysis.
6. The procedure for finding the optimal windows as well as for locating the optimal points within the windows allows performance in parallel and thus enables real-time feature extraction—in the extreme, down to a few lines' delay during the scanning of the image.

We will discuss the three steps in detail, omitting the operator's property of locating centers of spirals. As the window selection and the distinction are based on the point location within the windows, we treat these estimation procedures first.

16.4.2 Estimating Corner Points

Let us first assume that an $m_r \times m_c$ window is known to contain a corner point $p_0 = (r_0, c_0)'$. The aim is to obtain an estimate $\hat{p}_0 = (\hat{r}_0, \hat{c}_0)'$ for the corner point. As we do not know how many edges intersect at p_0, we treat the edge

elements individually, each being representative of a straight line passing through $p_i = (r_i, c_i)'$, $i = 1, \ldots, m$ (cf. Fig. 16.9 a). We assume that the estimated point $\hat{p}_i$ is the point closest to all straight lines crossed by the edge elements in the window, taking individual uncertainty into account.

Let $p_i' = (r_i, c_i)$ be a pixel within the window, with

$$e_i' = e(r_i, c_i) \neq \nabla f_i' = (f_{r_i}, f_{c_i}) = \left[f_{r_i}(r_i, c_i), f_{c_i}(r_i, c_i)\right]$$

its gradient, and

$$\frac{\nabla f_i'}{|\nabla f_i|} = (\cos\phi_i, \sin\phi_i) = \left[\cos\phi(r_i, c_i), \sin\phi(r_i, c_i)\right]$$

its unit vector in the direction of the gradient. The straight line passing through $p_i = (r_i, c_i)$ parallel to the edge direction is given by

$$(p - p_i)' e_i = 0$$

with $p = (r, c)'$.

We now make two assumptions that allow a noniterative solution for determining the unknown corner point $p_0 = (r_0, c_0)'$:

- The directions ϕ_i of the edge elements, that is, of the lines orthogonal to ∇f_i passing through (r_i, c_i), are error free.
- All uncertainty is taken care of by the distances $l_i = p_i' e_i$ which we assume to be the original observations.

The weights of the edge elements are assumed to be $w_i = ||\nabla f_i||^2$, as if the edge position across had been determined in an image-matching approach using

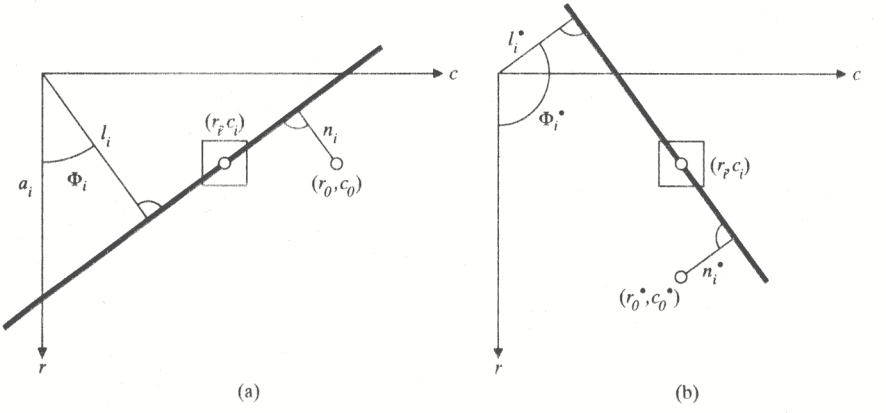


Figure 16.9 Models for estimating corners (a) and centers of circular symmetric features (b). In (a) the edge element through (r_i, c_i) is represented by a straight line (l_i, ϕ_i). The optimal point (r_o, c_o), a corner, is the one that minimizes the weighted sum of the squared distances n_i. In (b) the slope element through (r_i, c_i) is represented by the straight line (l_i', ϕ_i'). The optimal point, the center of the circular feature, minimizes the weighted sum of the distances n_i'.

a model of the edge (cf. Eq. 16.6). Moreover, the individual edge elements are assumed to be independent. Then the linear model reads as

$$l(r_i,c_i) = \cos\phi(r_i,c_i)\cdot\hat{r}_0 + \sin\phi(r_i,c_i)\cdot\hat{c}_0 + n(r_i,c_i)$$

or, with $l_i = l(r_i,c_i), \phi_i = \phi(r_i,c_i)$, and $n_i = n(r_i,c_i)$,

$$l_i = \cos\phi_i\cdot\hat{r}_0 + \sin\phi_i\cdot\hat{c}_0 + n_i, \quad i = 1,\ldots,m \tag{16.62}$$

which is assumed to hold for all $m = m_r \times m_c$ pixels (r_i,c_i). The weight of l_i is assumed to be

$$w_i = \|\nabla f_i\|^2 = f_r^2(r_i,c_i) + f_c^2(r_i,c_i)$$

Minimizing $\Omega(r_0,c_0) = \sum_{i=1}^{m} n_i^2\cdot w_i$ with respect to $\hat{r}_0$ and $\hat{c}_0$ yields

$$\frac{1}{2}\frac{\partial\Omega(\hat{r}_0,\hat{c}_0)}{\partial\hat{r}_0} = \sum_{i=1}^{m}\cos\phi_i\cdot(l_i - \cos\phi_i\cdot\hat{r}_0 - \sin\phi_i\cdot\hat{c}_0)\cdot w_i = 0$$

$$\frac{1}{2}\frac{\partial\Omega(\hat{r}_0,\hat{c}_0)}{\partial\hat{c}_0} = \sum_{i=1}^{m}\sin\phi_i\cdot(l_i - \cos\phi_i\cdot\hat{r}_0 - \sin\phi_i\cdot\hat{c}_0)\cdot w_i = 0$$

which leads to the normal equation system

$$\begin{pmatrix} \sum_{i=1}^{m} w_i cos^2\phi_i & \sum_{i=1}^{m} w_i\cos\phi_i\cdot\sin\phi_i \\ \sum_{i=1}^{m} w_i\cos\phi_i\cdot\sin\phi_i & \sum_{i=1}^{m} w_i\sin^2\phi \end{pmatrix}\begin{pmatrix}\hat{r}_0\\ \hat{c}_0\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^{m} l_i w_i\cos\phi_i \\ \sum_{i=1}^{m} l_i w_i\sin\phi_i\end{pmatrix}$$

Substituting $l_i = r_i\cos\phi_i + c_i\sin\phi_i$ and using $f_{r_i} = f_r(r_i,c_i) = \|\nabla f_i\|\cos\phi_i$ and $f_{c_i} = f_c(r_i,c_i) = \|\nabla f_i\|\sin\phi_i$, we arrive at

$$\begin{pmatrix} \sum_{i=1}^{m} f_{r_i}^2 & \sum_{i=1}^{m} f_{r_i}f_{c_i} \\ \sum_{i=1}^{m} f_{r_i}f_{c_i} & \sum_{i=1}^{m} f_{c_i}^2 \end{pmatrix}\begin{pmatrix}\hat{r}_0\\ \hat{c}_0\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^{m}(f_{r_i}^2 r_i + f_{r_i}f_{c_i}c_i) \\ \sum_{i=1}^{m}(f_{r_i}f_{c_i}r_i + f_{c_i}^2 c_i)\end{pmatrix} \tag{16.63}$$

which is identical to Eq. (16.60). This was to be expected, as the intensity function at an ideal corner is invariant with respect to scale when taking the corner point as center of the origin.

We can also write Eq. (16.63) in the form

$$\left(\sum_{i=1}^{m} W_i\right)\hat{p}_0 = \sum_{i=1}^{m}(W_i\cdot p_i) \tag{16.64}$$

with the singular-weights matrices

$$W_i = \nabla f_i \nabla f_i' = \|\nabla f_i\|^2 \cdot e_i\, e_i' = \|\nabla f_i\|^2 \cdot \begin{pmatrix} \cos^2\phi_i & \cos\phi_i sin\phi_i \\ \cos\phi_i \sin\phi_i & \sin^2\phi_i \end{pmatrix} \quad (16.65)$$

This shows $\hat{p}_0$ to be the *weighted center of gravity* of all points p_i with *the squared gradient* $\nabla f_i \nabla f_i'$ *as weight matrix,* thus using the model

$$p_i = \hat{p}_0 + n_i \quad \text{or} \quad \begin{pmatrix} r_i \\ c_i \end{pmatrix} = I_2 \cdot \begin{pmatrix} \hat{r}_0 \\ \hat{c}_0 \end{pmatrix} + \begin{pmatrix} n_{r_i} \\ n_{c_i} \end{pmatrix} \quad (16.66)$$

We will use this model in the following because of its simplicity.

We can arrive at a third interpretation of the normal equation system (16.63). If we divide the model equation (16.62) by $\cos\phi_i$ and substitute $s_i = \tan\phi_i$ and $\bar{n}_i = n_i/\cos\phi_i$, we obtain for the observation $a_i = l_i/\cos\phi_i$

$$a_i = \hat{r}_0 + s_i \cdot \hat{c}_0 + \bar{n}_i \quad (16.67)$$

which is the intercept of the straight line through p_i with the r-axis (see Fig. 16.9). If we now take (a_i, s_i) as the representation of the edge element in Hough space with intercept and slope as parameters, we can interpret the model equation (16.67) to be a straight-line fit in Hough space (cf. Chapter 7) with unknown parameters $(\hat{r}_0, \hat{c}_0)$, which in the image domain correspond to the intersection point of the edge elements. If we take the proper weights for the observations a_i, namely, $\|\nabla f_i\|^2 \cos\phi_i^2$, we arrive at the same normal equation system as before Eq. (16.63).

We can easily transfer this reasoning to the case in which the window is supposed to contain a circular symmetric feature, such as a circle or a set of rings. The idea is to use the slope elements, that is, the straight lines going through the points $p_i' = (r_i, c_i)$ and having the direction of the gradient ∇f_i. If the window contains a circular symmetric feature, then these lines intersect at the center $p_0^\bullet = (r_0^\bullet, c_0^\bullet)$. With the unit vector $e_i^\bullet$ being orthogonal to $e_i, e_i^\bullet = (-\sin\phi_i, \cos\phi_i)$, the equation for these straight lines reads as

$$(p - p_i)' \cdot \tilde{e}_i = 0$$

With the same reasoning as above, thus fixing the orientations ϕ_i and taking the distance $l_i^\bullet = l_i^\bullet(r_i, c_i) = p_i' e_i^\bullet = -r_i \sin\phi_i + c_0 \cdot \cos\phi_i$ as the original observation, we obtain the linear model

$$l_i^\bullet = -\sin\phi_i \cdot \hat{r}_0^\bullet + \cos\phi_i \cdot \hat{c}_0^\bullet + n_i^\bullet, \quad i = 1, \ldots, m \quad (16.68)$$

which is valid for all pixels in the window and assumes the weight to be $w_i = \|\nabla f_i\|^2$. The resulting normal equation system for the center $(\hat{r}_0^\bullet, \hat{c}_0^\bullet)$ of the circular symmetric feature reads as

$$\begin{pmatrix} \sum_{i=1}^{m} w_i \sin^2\phi_i & -\sum_{i=1}^{m} w_i \cos\phi_i \sin\phi_i \\ -\sum_{i=1}^{m} w_i \cos\phi_i \sin\phi_i & \sum_{i=1}^{m} w_i \cos^2\phi_i \end{pmatrix} \begin{pmatrix} \hat{r}_0^\bullet \\ \hat{c}_0^\bullet \end{pmatrix} = \begin{pmatrix} -\sum_{i=1}^{m} w_i l_i^\bullet \sin\phi_i \\ \sum_{i=1}^{m} w_i l_i^\bullet \cos\phi_i \end{pmatrix} \quad (16.69)$$

which can also be written as

$$\begin{pmatrix} \sum_{i=1}^{m} f_{c_i}^2 & -\sum_{i=1}^{m} f_{r_i} f_{c_i} \\ -\sum_{i=1}^{m} f_{r_i} f_{c_i} & \sum_{i=1}^{m} f_{r_i}^2 \end{pmatrix} \begin{pmatrix} \hat{r}_0' \\ \hat{c}_0' \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} (f_{c_i}^2 r_i - f_{r_i} f_{c_i} c_i) \\ \sum_{i=1}^{m} (-f_{r_i} f_{c_i} r_i + f_{r_i}^2 c_i) \end{pmatrix} \tag{16.70}$$

This equation system is identical to Eq. (16.61) for the point that is invariant with respect to rotation during intensity-based matching.

Equation (16.70) can be written as

$$\left(\sum_{i=1}^{m} W_i^{\bullet} \right) \cdot p_0^{\bullet} = \sum_{i=1}^{m} (W_i^{\bullet} \cdot p_i)$$

Now, however, with the singular-weight matrix

$$W_i^{\bullet} = \| \nabla f_i \|^2 e_i^{\bullet} \cdot e_i^{\bullet} = \| \nabla f_i \|^2 \begin{pmatrix} \sin^2 \phi_i & -\cos \phi_i \sin \phi_i \\ -\cos \phi_i \sin \phi_i & \cos^2 \phi_i \end{pmatrix} \tag{16.71}$$

the center of circular symmetric features can also be interpreted as the weighted center of gravity. Thus the functional models equation (16.66) for estimating corners and centers of circular symmetric features is the same, whereas the weights W_i and $W_i^{\bullet}$ are different.

Here also the third interpretation of the estimation procedure using the Hough space can be applied: The determination of the intersection of all slope elements corresponds to the determination of the parameter $(\hat{r}_0^{\bullet}, \hat{c}_0^{\bullet})$ of a straight line in Hough space. This may be achieved by dividing Eq. (16.68) by $-\sin \phi_i$, as $a_i^{\bullet} = -l_i^{\bullet} / \sin \phi_i$ is the intercept of the line passing through $p_i(r_i, c_i)$ and having the direction of the gradient.

The estimation procedures provided in this discussion may be used independently in selecting the windows and in determining the window content. But with the least-squares models for estimating the distinct points, we obtain statistical means for automatically classifying the window content and searching for those positions of the window within the image that generate the best local estimate of the image features.

16.4.3 Evaluation and Classification of Selected Windows

We first want to evaluate the estimates $\hat{p}_0$ and $\hat{p}_0^{\bullet}$ for corners and centers of circular symmetric features. In both cases we need an estimate $\hat{\sigma}_n^2$ for the noise variance σ_n^2 and the inverse Q of the 2×2 normal equation matrix N.

We immediately obtain the weighted sums Ω and Ω' of the squared residuals:

$$\Omega = \sum_{i=1}^{m} (r_i - \hat{r}_0, c_i - \hat{c}_0) \cdot W_i \cdot (r_i - \hat{r}_0, c_i - \hat{c}_0)' = \sum_{i=1}^{m} n_i^2 w_i$$

$$\Omega^{\bullet} = \sum_{i=1}^{m} (r_i - \hat{r}_0^{\bullet}, c_i - \hat{c}_i^{\bullet})' \cdot W_i^{\bullet} \cdot (r_i - \hat{r}_0^{\bullet}, c_i - \hat{c}_0^{\bullet}) = \sum_{i=1}^{m} n_i^{\bullet 2} w_i^{\bullet}$$

from which follows the estimate for the noise variance

$$\hat{\sigma}_n^2 = \frac{\Omega}{m-2} \text{ and } \hat{\sigma}_{n\bullet}^2 = \frac{\Omega^\bullet}{m-2} \tag{16.72}$$

Thus we obtain the covariance matrices

$$\Sigma_{pp} = D\begin{pmatrix}\hat{r}_0\\ \hat{c}_0\end{pmatrix} = \hat{\sigma}_n^2 \cdot \begin{pmatrix}\sum f_{r_i}^2 & \sum f_{r_i} f_{c_i}\\ \sum f_{r_i} f_{c_i} & \sum f_{c_i}^2\end{pmatrix}^{-1} \tag{16.73}$$

and

$$\Sigma_{p^\bullet p^\bullet} = D\begin{pmatrix}\hat{r}_0^\bullet\\ \hat{c}_0^\bullet\end{pmatrix} = \hat{\sigma}_{n\bullet}^2 \cdot \begin{pmatrix}\sum f_{c_i}^2 & -\sum f_{r_i} f_{c_i}\\ -\sum f_{r_i} f_{c_i} & \sum f_{r_i}^2\end{pmatrix}^{-1}$$

for the two estimated points $\hat{p}$ and $\hat{p}^\bullet$.

EXAMPLE 16.7

Figure 16.10 shows 18 small windows between 5×5 and 9×9 pixels in size. The edge and slope elements sit between the pixels as the Roberts gradient is taken for determining the first derivatives. The hypothesis about the image content is indicated by a small cross for corners and a circle for circular symmetric features. The true points with 99% probability lie within the shown confidence ellipses. The estimation procedure obviously can handle corners with multiple edges coming together. Extrapolation is performed, and

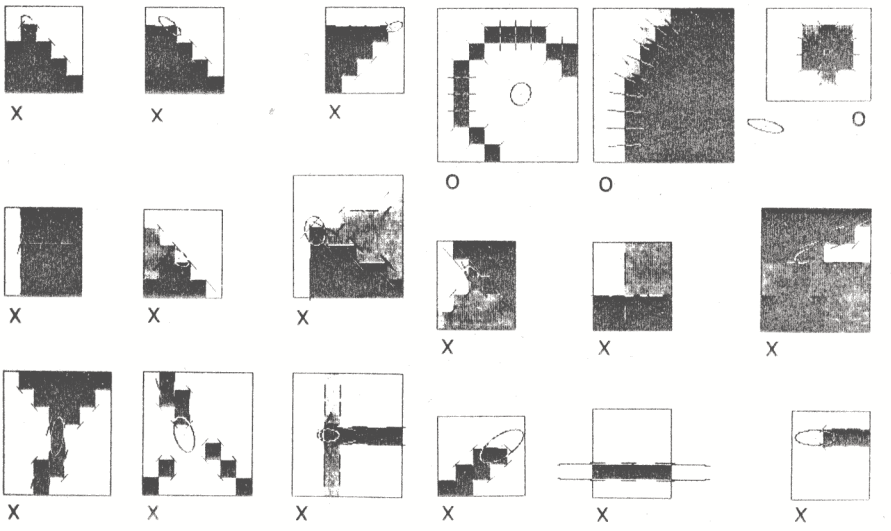


Figure 16.10 Weighted centers of gravity for 18 windows with 99%-confidence ellipses. The assumed model is indicated by the cross (corners) and the circle (circular symmetric figures). Observe the high precision and the extrapolation capacity of the estimation procedure.

the intersection point of the edge lines can be estimated. The example of a simple straight line passing through the window is solved by introducing the center $P_c = (r_1, c_1)$ of the window as a priori information in a Bayesian manner, thus adding the observations

$$\begin{pmatrix} r_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} \hat{r}_0 \\ \hat{c}_0 \end{pmatrix} + \begin{pmatrix} n_{r_1} \\ n_{c_1} \end{pmatrix}$$

with a very low weight $(\sigma_1^2 \cdot I_2)$ to the other observed points p_i.

The covariance matrices Σ_{pp} and $\Sigma_{p^*p^*}$ can be represented by three values, which we will use later for the window selection step:

1. The average precision of the point or the weight w of the point, assuming $\sigma_n^2 = 1$

$$w = \frac{1}{\frac{1}{2}tr\Sigma} = \frac{1}{\frac{1}{2}trN^{-1}} = \frac{det\ N}{\frac{1}{2}tr\ N} \tag{16.74}$$

with N being the 2×2 matrix in Eqs. (16.60) or (16.61). If the eigenvalues of N are equal, then

$$w = \frac{1}{2}tr\ N \tag{16.75}$$

2. The direction of the major axis of the confidence ellipse

$$\phi = \frac{1}{2}\arctan\frac{-2N_{12}}{N_{11} - N_{22}}$$

3. The form of the confidence ellipse, which may be derived from the ratio λ_1/λ_2 of the two eigenvalues of N or Σ or (to avoid the calculation of the eigenvalues) from the *form factor*

$$q = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 = \frac{4\det N}{tr^2 N} = \frac{4\det\Sigma}{tr^2\Sigma} \tag{16.76}$$

which lies in the range between 0 and 1. This can be determined either from Σ directly or from the normal equation matrix N. Here $q = 1$ corresponds to a circular error ellipse, whereas $q = 0$ indicates that the smaller eigenvalue is zero, the normal equation matrix is singular, and thus the window contains a straight edge. We will use the form parameter q to exclude points lying on edges when searching for optimal windows. The term q can be said to measure the circularity of the ellipse or the degree of isotropy of the texture within the window.

As the normal equation matrix for the corner is the same as for matching two windows, assuming a shift only (cf. Eq. 16.47), the confidence ellipses of the matching examples have the same form and orientation as those one would have obtained for the estimated corner point in these windows (cf. Fig. 16.7).

The classification of the window content aims at deciding whether

$$\begin{aligned} H_0 &: \text{ the window contains an isotropic texture;} \\ H_A &: \text{ the window contains a corner;} \\ H_{A\bullet} &: \text{ the window contains a circularly symmetric feature.} \end{aligned} \tag{16.77}$$

We therefore assume q to be large enough to exclude windows containing straight edges.

Owing to the orthogonality of the models, one can use the test statistic

$$T = \frac{\Omega}{\Omega^{\bullet}} \sim F_{m-2,m-2} \tag{16.78}$$

which is F-distributed with $m-2$ and $m-2$ degrees of freedom, where m is the number of edge or slope elements. With two critical values k_1 and $k_2 = 1/k_1$, we can classify the window content:

$$\begin{aligned} T > k_1 &\rightarrow \text{ circular symmetric feature} \\ T < k_2 &\rightarrow \text{ corner} \\ \text{Else} &\rightarrow \text{ isotropic texture.} \end{aligned}$$

Thus the test checks which of the models in (16.77) hold or not.

The parameterization of a spiral would permit estimation of the parameter, including the corner and the circular symmetric windows as special cases (cf. Bigün, 1990).

EXAMPLE 16.8

For the binary window

$$f = \begin{matrix} \\ \downarrow r \\ \\ \\ \\ \end{matrix} \begin{matrix} \rightarrow c \\ \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{matrix} \end{matrix}$$

(cf. Fig. 16.10 #1) we obtain the derivatives f_r, f_c, and the product $f_r \cdot f_c$.

$f_r =$

	1	2	3	4
1	2	2	2	1
2	0	0	-1	-1
3	0	-1	-1	0
4	-1	-1	0	0

$f_c =$

	1	2	3	4
1	0	0	0	-1
2	0	0	-1	-1
3	0	-1	-1	0
4	-1	-1	0	0

$$f_r \cdot f_c = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & 0 & 0 & -1 \\ 2 & 0 & 0 & 1 & 1 \\ 3 & 0 & 1 & 1 & 0 \\ 4 & 1 & 1 & 0 & 0 \end{array}$$

Similarly, gradients $f_r(r,c) = f(r+\frac{1}{2}, c-\frac{1}{2}) + f(r+\frac{1}{2}, c+\frac{1}{2}) - f(r-\frac{1}{2}, c-\frac{1}{2}) - f(r-\frac{1}{2}, c+\frac{1}{2})$, and $f_c(r,c)$ are used. The coordinate system is chosen such that the gradients refer to integer positions. The normal equation system (16.60) thus reads as

$$\begin{pmatrix} 19 & 5 \\ 5 & 7 \end{pmatrix} \begin{pmatrix} \hat{r}_0 \\ \hat{c}_0 \end{pmatrix} = \begin{pmatrix} (13\cdot 1+2\cdot 2+2\cdot 3+2\cdot 4)+(1\cdot 1+2\cdot 2+2\cdot 3+0\cdot 4) \\ (-1\cdot 1+2\cdot 2+2\cdot 3+2\cdot 4)+(1\cdot 1+2\cdot 2+2\cdot 3+2\cdot 4) \end{pmatrix}$$
$$= \begin{pmatrix} 42 \\ 36 \end{pmatrix}$$

This leads to estimates $\hat{r}_0 = 114/108 = 1.056, \hat{c}_0 = 474/108 = 4.389$. The weight sum of the squared residuals is

$$\Omega = 5/12 = 0.4166$$

In case one would assume that the window represents a circular symmetric feature,

$$\Omega^\bullet = 77/12 = 6.4166$$

Therefore the test statistic Eq. (16.78) is obtained as

$$T = \Omega/\Omega^\bullet = 5/77 \approx 0.065$$

which is significantly smaller than 1, indicating that the window represents a corner.

16.4.4 Selection of Optimal Windows

We can now describe the goal of the window selection scheme precisely: The interest operator should find distinct points—points that are discernible from immediate neighbors. We therefore search for local optima of the expected precision for estimating the location of points within a window or for matching the window with one or another image.

Recall that the weight w of the estimated points depends on the noise variance σ_n^2 and the signal content of the window, specifically, the average squared gradient. If one assumes constant σ_n^2, the search need be based only on the average squared gradient N or its inverse. The decisive values—namely, the form factor q and the traces trN and trN^{-1}—are identical for the three tasks:

- Area based matching,
- Corner estimation,
- Estimation of the center of a circular symmetric feature.

As for large q, and thus nearly isotropic window texture $trN \approx 4trN^{-1}$, we base the selection of the optimal window on $w_i = \frac{1}{2}trN_i$ and q_i and require the following:

1. The confidence ellipse should be round. This is to ensure that selected windows do not contain a straight edge or strongly oriented texture, but rather that the point determination is equally precise in all directions. This leads to the condition
$$q_i > q_{min}$$
2. The confidence ellipse should be smaller than those obtained from neighboring windows and should not exceed a certain size. This ensures maximum local separability or distinctness and good accuracy of the point determination and leads to the conditions
$$w_i > w_{min}$$
and
$$w_i \geq w_l, \text{ for all } l \in \{ \text{ neighborhood of } i \}$$
The procedure for finding the centers of optimal windows is therefore the following:
Input parameters:
   - A noise-cleaned image $f(r,c)$
   - A gradient operator
   - The window size for the operator (m_r, m_c)
   - A neighborhood size for finding local extrema
   - A threshold q_{min} for q
   - A threshold w_{min} for w

1. Determine two derivative images $f_r(r,c)$ and $f_c(r,c)$ by using the gradient operator.
2. Determine the three images containing $f_r^2(r,c), f_r(r,c) \cdot f_c(r,c)$ and $f_c^2(r,c)$.
3. Convolve the three images with a box filter of size $m_r \times m_c$ yielding the three images $N_{11}(r,c), N_{12}(r,c)$, and $N_{22}(r,c)$, which represent the elements of the normal equation matrix for all positions in the image.
4. Determine the images $q(r,c)$ and $w(r,c)$ from Eqs. (16.76) and (16.75).
5. Threshold $w(r,c)$ leading to $w^*(r,c)$:
$$w^*(r,c) = \begin{cases} w(r,c) & \text{if } w(r,c) > w_{min} \text{ and } q(r,c) > q_{min} \\ 0 & \text{else} \end{cases}$$
6. Suppress the nonmaxima of $w^*(r,c)$, setting all w^* to zero where there is no relative maximum of w^* within the prespecified window.

Remarks:

a. Any noise-cleaning procedure that preserves edges and corners may be used. If noise is moderate, a linear filter, preferably Gaussian, is sufficient. In this case the noise cleaning and the gradient determination may be combined by using a larger window for the gradient operator. If the images are already cleaned, even Roberts gradient works sufficiently well.

b. The window size has to be chosen according to the task and the image content. If the selected points are used for matching highly textured images, window sizes of 5×5 already give reliable results. If corners of polyhedra or other interpretable features are to be detected, the window size should be chosen as large as possible so that two neighboring points on an average do not fall into the same window.

c. The neighborhood for the nonmaximum suppression also has to be chosen according to the task and the image content. A large neighborhood suppresses more points but also may eliminate good points sitting near stronger ones. Taking the window and neighborhood sizes to be equal has been proved a reasonable choice in many applications, such as stereo, image sequence analysis, and general image analysis.

d. The threshold q_{min} for the form parameter q due to Eq. (16.76) can be based on a critical value for the ratio $\lambda_1/\lambda_2 (\lambda_1 > \lambda_2)$ of the eigenvalues of N or Σ. If the window contains a straight edge, the larger eigenvalue equals $m \cdot \sigma_{g'}^2$ that is, is proportional to the variance of the gradient across the edge, whereas the smaller eigenvalue equals $m \cdot \sigma_{n'}^2$, that is, is proportional to the variance of the noise gradient. A requirement that the window significantly contain no edge of $\lambda_1/\lambda_2 = 4$, say, using Eq. (16.76), leads to a threshold of $q_{min} = 0.64$. It corresponds to requiring the ratio of the two semiaxes of the error ellipse to be smaller than 2, or the two sides of a corner to meet at an acute angle larger than approximately $53° \approx 2 \arctan(1/2)$.

e. A similar reasoning leads to a choice for the threshold w_{min}. The weight should be significant with respect to noise in the input image (possibly already restored). The expectation of w_i is $m \cdot \sigma_{n'}^2$ where m is the number of edge elements used for determining w. The noise gradient variance $\sigma_{n'}^2$ may be estimated from

$$\sigma_{n'}^2 = \frac{t_{f'} \cdot M}{\#(f'^2 < t_{f'})}$$

where $t_{f'}$ is a threshold for the gradients $||\nabla f||^2 = f'^2$; where $\#(f'^2 < t_{f'})$ is the number of gradients of the image that are smaller than the chosen threshold; and where M is the total number of the edge elements of the image used for this estimate. The threshold $t_{f'}$ should be approximately three to five times an initial estimate of the noise gradient variance $\sigma_{n'}^2$, which could be obtained from a histogram of the f'^2. The estimate is based on the assumption that the noise in the image is white and Gaussian, and therefore the gradient squares $f_i'^2$ at flat areas are χ_2^2-distributed $[\exp(-x), x > 0]$

and that edges influence only the right-hand side of the (cumulative) histogram (theoretically: 1-exp(-x)), which — in contrast to Vorhees' and Poggio's approach (1987) — does not necessarily have to be built up explicitly, except for an initial guess. The threshold w_{min} then can be chosen to be a critical value for the noise gradient, that is, $km\sigma_n^2$, with $k = 10$, say, guaranteeing that only significant windows are selected. Though because of the discreteness of the intensity values and possibly the additional smoothing, the initial guess σ_n^2, may be 0, $t_{f'}$ should be chosen > 0.

f. It may happen that different selected windows refer to the same distinct point. One representative may be selected *after* the class and the *accurate position* of the points have been determined and their equivalence checked; the one with the highest weight may then be chosen. The reference of multiple windows to the same point partially results from using a box filter for determining the elements of the normal equation system in step 3 and may be reduced, not eliminated, by using a Gaussian, a triangular, or an equivalent filter with a clear maximum at the center of the filter mask.

16.4.5 Uniqueness of Selected Points

The uniqueness of the selected points can be based on their similarity derived from the describing attributes. In our case this would be the intensity function in a window *around the optimal position* of the selected points, as the selected windows in certain cases may sit arbitrarily. Those points that have no features or attributes in common with other points should obtain the highest uniqueness measure.

We want to present a uniqueness measure that is based on the total correlation of a point with all others. It is based on the assumption that the attributes, here the intensity values, are Gaussian distributed (Förstner, 1988). We need the correlation matrix

$$R = (\rho_{ij})$$

between all selected points, with $\rho_{ii} = 1$ and ρ_{ij}, for example, from Eq. (16.34), replacing x_i by (r_i, c_i) and setting $u = 0$, that is, assuming no shift. Large correlation coefficients indicate high similarity. Then the total correlation ρ_i of point P_i and all others P_l can be derived from

$$\rho_i^2 = r_i'(R_{ii})^{-1} r_i = 1 - \frac{1}{(R^{-1})_{ii}} \tag{16.79}$$

where r_i is the vector containing all correlation coefficients between P_i and the P_l, R_{ii} is the submatrix of R after canceling the ith column and the ith row, and $(R^{-1})_{ii}$ is the ith diagonal element of the inverse of R. The total correlation obviously is a weighted mean over all correlations contained in r_i. For only two points the total correlation reduces to the normal correlation.

We now may define the uniqueness of point P_i by

$$u_i = \frac{1 - \rho_i}{\rho_i} \tag{16.80}$$

This is analogous to the signal-to-noise ratio in Eq. (16.40) but takes into account the inverse reasoning here: High correlations lead to low total uniqueness, whereas low total correlations lead to high uniqueness.†

The practical calculation has to take into account that R may be singular, especially if the number of points is larger than the number of attributes. Then one can calculate the uniqueness by using the correlation coefficient from

$$\rho_i^2 = 1 - \frac{1}{[(R + \delta \cdot I)^{-1}]_{ii}} + k \tag{16.81}$$

$$k = \frac{\delta(n + \delta)}{n - 1 + \delta}$$

with δ a small number, for example, 0.001, and n the number of points involved. The correction term k is motivated by assuming R to consist only of 1's, thus assuming the worst case, namely, that all points are completely similar. Then, owing to the $\delta \cdot I$-term modifying R, the correlation shows its largest value, namely, $1 - k$, not 1 as it should be.

Remarks:

1. The uniqueness measure based on the correlation coefficient of the surrounding intensity function derived from Eq. (16.34) is not invariant with respect to rotation, scale, or other distortions. Using the correlation coefficient from Eq. (16.37) derived from the least-squares fit between the intensity functions may provide this invariance, however, since it is practical only for moderate distortions.
2. This definition of uniqueness can be applied to all types of attributes as long as a correlation matrix can be derived, for example, based on a metric $d^2(P_i, P_j)$ between the points and using a correlation function, for example, $1/[1 + (d/d_0)^2]$, guaranteeing R to be at least positive semidefinite. This opens the door to measuring uniqueness based on attributes that are invariant with respect to a wide class of transformations, especially also symbolic attributes.
3. Both distinctness and uniqueness are based on the same similarity measure, the correlation coefficient. If we follow the reasoning that leads to Eq. (16.39), high distinctness corresponds to high curvature of the autocorrelation function, thus low correlation between a point and its immediate neighbors, whereas high uniqueness goes along with low correlation with the other selected points.
4. For large numbers n of points, the calculation of the inverse, being of order $O(n^3)$, may be prohibitive. This can be circumvented by using the maximum

†A different definition could apply an information-theoretic view and use the mutual information $H_i = -1/2 \log(1 - \rho_i^2)$ (Papoulis, 1984, Eq. 75–98) of the P_i and all P_1, leading to the definition $u_i' = 1/H_i$. (Förstner, 1988). u_i is a monotonic function of u_i'. We prefer u_i here, as we want to use it as a factor for modifying the weight $1/\sigma_x^2$ of correspondences derived from the expected precision, where $\text{SNR}^2 = \rho/(1 - \rho)$ is used (cf. below).

square correlation $\rho_i^2 = \max_j(\rho_{ij}^2)$ of P_i with all other P_j to determine the uniqueness (cf. Eqs. 16.80 and 16.81).

EXAMPLE 16.9

Figure 16.11 shows two images with the selected windows classified according to Section 16.4.3. The estimated points obviously sit at the right positions, when one takes the discreteness of the presentation into account. Observe that in the upper right of Fig. 16.11(b) even an extrapolation of the edges was

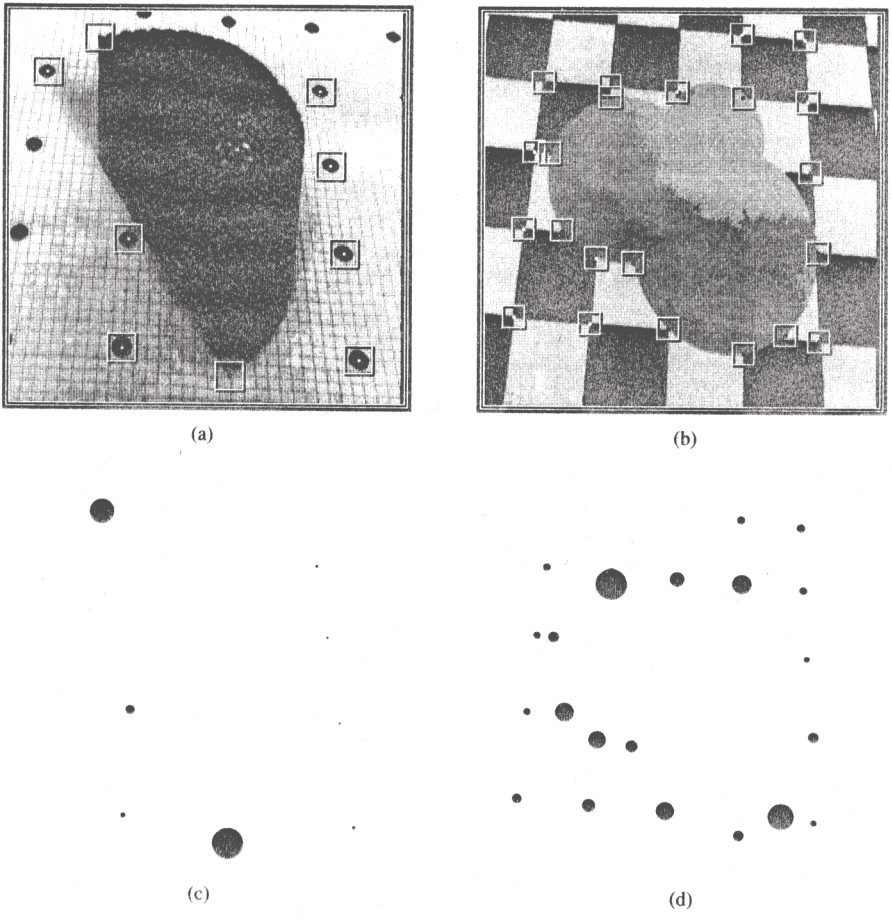


Figure 16.11 Two images (a) and (b) with selected windows and optimally estimated points after classification. Parts (c) and (d) show the uniqueness measure $(\rho/1-\rho)$ represented by the area of the circles. Observe that the repetitive features show low uniqueness.

achieved. Also the uniqueness measures, represented by the area of the circles in (c) and (d), correspond to intuition. The images were prefiltered with a 7×7 binomial filter. The window sizes were 17×17 and 21×21, respectively; Roberts gradient and the threshold $q_{min} = 0.75$ were used.

16.5 Robust Estimation for Feature-Based Matching

16.5.1 The Principle of Feature-Based Matching

Recall the three steps of feature-based matching (cf. Section 16.1.2):

1. Selecting features by using some interest operator or some other feature extraction scheme;
2. Finding correspondences between the features by using some similarity and consistency measure;
3. Interpolating between the parallaxes by using a spatial-mapping function.

After having discussed criteria for selecting appropriate features and one possibility for extracting distinct points, we now develop criteria for finding correspondences and for interpolating between the matched features in order to obtain a dense parallax field. Steps 2 and 3 are often designed separately. Specifically, Step 2, finding corresponding features, often does not utilize the same information that is later used for Step 3, interpolating the parallax field. Consequently the strength of the model, namely, the spatial-mapping function, is not exploited, thereby leading to a suboptimal solution of the correspondence step, unless a refinement goes along with the interpolation. Therefore integrating steps 2 and 3 may lead to optimal procedures. Intensity-based matching is an example, the features there being just the pixels themselves. Optical flow using differential techniques—the scale space filtering technique of Witkin, Terzopoulos, and Kass, (1987) or the maximum likelihood technique proposed by Cernusch-Frias, et al. (1989)—falls into the same category. Also, techniques using dynamic programming (Baker and Binford, 1982; Ohta and Kanade, 1985) to obtain global optima while matching the features (line segments) integrate correspondence and interpolation, though this is not made explicit.

For finding an optimal solution, the similarity of the features and the consistency of the correspondence have to be weighted properly. In the following we will discuss a measure for similarity that can be used directly in an estimation process for the parameters of the mapping function serving as a model for consistency. The structure of the setup is closely related to the one by Barnard and Thompson (1980). The individual steps, however, are replaced in order to arrive at an algorithm that tracks

the uncertainty from the basic observations, the intensities, to the final result, the spatial-mapping function, and then allows one to evaluate the individual steps and their effect on the result:

1. The similarity between the extracted features leads to a preliminary list of correspondences, including their weights. We restrict the discusssion to feature points, as described in the previous section.
2. A hypothesis for the mapping function is found by using a robust estimation procedure, similar to relaxation techniques, and by enforcing the one-to-one correspondence between the image features.
3. The final parameters of the mapping function are achieved by using a maximum likelihood estimate, allowing a rigorous evaluation of the match.

Though the spatial-mapping function may be quite general, for example, a piecewise-smooth deformation field, we treat only the case of an affine transformation between the images for simplicity. The generalization is straightforward.

16.5.2 The Similarity Measure

We want to determine the weight

$$w_{12} = \frac{1}{\hat{\sigma}^2_{p_{12}}}$$

of the parallax p_{12} in row and column directions by using the average variance

$$\hat{\sigma}^2_{p_{12}} = \frac{1}{2}\left(\hat{\sigma}^2_{r_{12}} + \hat{\sigma}^2_{c_{12}}\right)$$

of the points 1 and 2 in the left and the right image of concern. In order to arrive at a simple expression, we assume that the selected points have isotropic texture ($q = 1$, cf. Eq. 16.76), assuming derivatives f_r and f_c to be uncorrelated (covariance = 0), and that the points have equal variance. Then, using Eq. (16.73), we obtain

$$\hat{\sigma}^2_{p_{12}} = \frac{1}{2}\hat{\sigma}^2_n\left(\frac{1}{\sum f_r^2} + \frac{1}{\sum f_c^2}\right)$$

and with $\sum f_r^2 = \sum f_c^2$ (because of $q = 1$), we have

$$\hat{\sigma}^2_{p_{12}} = \hat{\sigma}^2_n \frac{1}{\frac{1}{2}\sum f_r^2 + f_c^2}$$

With the weights $w_i = \frac{1}{2} trN_i = \frac{1}{2}(\sum f_r^2 + \sum f_c^2)_i, i = 1,2$ of the two extracted points, and with $w = \sqrt{w_1 \cdot w_2}$ as an average, this yields

$$\hat{\sigma}^2_{p_{12}} = \frac{\hat{\sigma}^2_n}{\sqrt{w_1 \cdot w_2}}$$

Thus with some arbitrary variance factor σ_0^2, the weight w_{12} of the parallax can be determined from

$$w_{12} = \frac{\sqrt{w_1 \cdot w_2}}{\hat{\sigma}_n^2} \cdot \sigma_0^2 \tag{16.82}$$

If we now use Eq. (16.40) and replace σ_f^2 by $\sigma_{f_1} \cdot \sigma_{f_2}$, we finally obtain

$$w_{12} = m \cdot \frac{\rho_{12}}{1 - \rho_{12}} \cdot \frac{\sigma_o^2}{\sigma_{f_1} \cdot \sigma_{f_2}} \cdot \sqrt{w_1 \cdot w_2} = m \cdot \frac{\sigma_o^2}{\sigma_n^2} \cdot \sqrt{w_1 \cdot w_2} \tag{16.83}$$

This weight of the parallax can be determined by using:

- The correlation coefficient ρ_{12} of the two windows around the points based on Eq. (16.34), with the shift being zero;
- The empirical standard deviation σ_f of the intensities $f_1(r,c)$ and $f_2(r,c)$ in the two windows;
- The weights of the two windows, essentially being the variances of the gradients;
- The variance factor σ_0^2, usually taken to be 1.

Remarks

1. As the aim is to measure the similarity of the selected points that have been determined to subpixel values, the windows used for measuring the similarity have to refer to these estimated points. Thus they are usually placed on noninteger positions, which requires resampling. Rounding to full integer positions introduces errors of $1/\sqrt{12} \approx 0.3$ pixel spacing.
2. The weights w_i and the variances $\sigma_{f_i}^2$ can be computed separately and possibly in parallel for both images, the complexity being proportional to the total number of points selected and the window size. The main effort is to compute the correlation coefficients ρ_{12}, which in principle have to be determined over all pairings of points for the left and the right images. Heuristics, discussed later, may reduce this effort.
3. As the weight mainly depends on the correlation coefficient, it is only invariant with respect to geometric shifts and to linear radiometric transformations. As shown by Svedlow, McGillem, and Anuta (1976), geometric deviations below 20° or 30% do not have much influence on the correlation coefficient, thereby making this value useful for a wide class of applications. Strong rotations or occlusions, of course, cannot be handled with this similarity measure.
4. The three parts of the weight can be interpreted as similarity $[\rho_{12}/(1 - \rho_{12})]$, strength $(\sigma_{f_1} \cdot \sigma_{f_2})$, and distinctness $(\sqrt{w_1 \cdot w_2})$ and therefore can be replaced by any other measure having these properties. This especially holds for the correlation coefficient ρ_{12} or equivalently—including the strength—the empirical noise variance σ_n^2. The noise variance could also be derived from an intensity-based match of the two windows, then possibly leading to invariance with respect to more complex geometric differences. Using a different measure ρ with the properties of a correlation coefficient—for example, one derived from a distance d

and an appropriate correlation function, such as $\rho(d) = 1/(1+d^2)$—allows one to arrive at similarity measures that are invariant with respect to a much wider class of transformations than simply a geometric shift and a linear radiometric tranformation; measures of symbolic attributes of the points might also be included (cf. Exercise 16.19).

16.5.3 Heuristics for Selecting Candidate Pairs

The first step after selecting image features is to find *candidate pairs* of features. This is to reduce the algorithmic complexity in the final match. Here all types of a priori knowledge may be included and all types of strategies may be applied. The only requirement is that these heuristics and strategies be conservative, that is, they should eliminate only truly wrong correspondences. Some of these heuristics and strategies follow (cf. Ballard and Brown, 1982):

1. The *expected parallax* may be used to exclude feature pairs that are unlikely. This expectation may result from a *scale space* approach using an image *pyramid,* where the result of one level serves as approximation for the level below. The model of the object or, the movement of the object or the camera, may lead to weak or even strong constraints on the parallax field. When one knows the focus of expansion in an image sequence with constant velocity and no rotation, or when one knows the relative orientation of the two cameras, then the corresponding points have to sit on straight lines, the epipolar lines (cf. below).
2. These heuristics may be coupled with the requirement that the similarity of the features be above a certain threshold. One possibility is to require a minimum correlation coefficient, such as $\rho_{12} \geq 0.5$, which, according to Eq. (16.41), corresponds to requiring SNR ≥ 1, a reasonable threshold. Or one may require that $\sigma_p = \sigma_o/\sqrt{w_{12}}$ be better than some standard deviation, such as 2 pixels.
3. Finally, the uniqueness (cf. Section 16.4.5) of the selected points may also be used, for example, to reduce the number of candidate pairs. This can be done separately for both images by canceling all points that are too similar to one or several points within the same image. A criterion could be that the total correlation coefficient (e.g. Eq. 16.79) must be larger than a threshold to keep the point an interesting one. Or the weight w_{12} of the correspondence could be modified by including the uniqueness leading to

$$w_{12} = \frac{\sigma_0^2}{\sigma_n^2}\sqrt{w_1 \cdot w_2} \cdot \sqrt{u_1 \cdot u_2} \tag{16.84}$$

(cf. Eq. 16.82) thus increasing the weight with the average uniqueness of the two points in concern. (This is why we used the definition from Eq. 16.80 and not the one based on the mutual information.)

Other heuristics may be used too. The problem with all of them is that up to now no immediate evaluation of their effect on the result has been possible, specifically with respect to the savings in the algorithmic complexity of the total procedure.

The result of this selection step is a list $\{(r',c'),(r'',c''),w\}$ of candidate pairs with their coordinates in the right and left images and the weight of the correspondence as additional attribute.

16.5.4 Robust Estimation for Determining the Spatial-Mapping Function

The preliminary correspondences now have to be compared to see whether they are consistent with some model. The most general model applied here could be a local smoothness constraint that should hold almost everywhere (Barnard and Thompson, 1980; Terzopoulos, 1986b). The application of a finite-element description of the spatial-mapping function would be particulary appropriate. In order to demonstrate the principle, we restrict our discussion to the same model as the one we used with intensity-based matching, namely, a linear geometric transformation between the images. The model can be expressed as the parallaxes $p = r'' - r'$ and $q = c'' - c'$ by Eq. 16.54.

The model is assumed to hold for the correct correspondences. Assuming that the random errors are small, we can treat r' and c' to be fixed, which is an acceptable approximation and sufficient in most cases. The parallaxes p_k and q_k of the corresponding point a are supposed to have the same weight.

The normal equations for the least-squares estimates have a special structure that can be exploited to reduce the numerical effort, namely,

$$\begin{bmatrix} \sum r_k'^2 w_k & \sum r_k' c_k' w_k & \sum r_k' w_k \\ \sum r_k' c_k' w_k & \sum c_k'^2 w_k & \sum c_k' w_k \\ \sum r_k' w_k & \sum c_k' w_k & \sum w_k \end{bmatrix} \begin{bmatrix} \hat{a}_1 & \hat{a}_4 \\ \hat{a}_2 & \hat{a}_5 \\ \hat{a}_3 & \hat{a}_6 \end{bmatrix} = \begin{bmatrix} \sum p_k r_k' w_k & \sum q_k r_k' w_k \\ \sum p_k c_k' w_k & \sum q_k c_k' w_k \\ \sum p_k w_k & \sum q_k w_k \end{bmatrix} Q^{-1} \tag{16.85}$$

with the residuals

$$n_k = \begin{pmatrix} n_p \\ n_q \end{pmatrix}_k = \begin{pmatrix} \hat{a}_1 & \hat{a}_2 \\ \hat{a}_4 & \hat{a}_5 \end{pmatrix} \begin{pmatrix} r' \\ c' \end{pmatrix}_k + \begin{pmatrix} \hat{a}_3 \\ \hat{a}_6 \end{pmatrix} - \begin{pmatrix} p \\ q \end{pmatrix}_k \qquad k = 1,\ldots,l \tag{16.86}$$

We may obtain an estimate for the variance factor

$$\hat{\sigma}_o^2 = \frac{1}{2l-6} \sum_k n_k' n_k w_k \tag{16.87}$$

which, together with the inverse $Q = N^{-1}$ of the normal equation matrix N, can be used to determine the estimated standard deviation of the unknowns:

$$\begin{aligned} \hat{\sigma}_{\hat{a}_1} &= \hat{\sigma}_{\hat{a}_4} = \hat{\sigma}_o \sqrt{Q_{11}} \\ \hat{\sigma}_{\hat{a}_2} &= \hat{\sigma}_{\hat{a}_5} = \hat{\sigma}_o \sqrt{Q_{22}} \\ \hat{\sigma}_{\hat{a}_3} &= \hat{\sigma}_{\hat{a}_6} = \hat{\sigma}_o \sqrt{Q_{33}} \end{aligned} \tag{16.88}$$

With the contribution h_k of each observation p_k or q_k to the determination of the six unknowns

$$h_k = (r_k' \; c_k' \; 1) \cdot Q^{-1} \cdot (r_k' \; c_k' \; 1)' \cdot w_k \tag{16.89}$$

we obtain (cf. Huber, 1981; Förstner, 1987) the standard deviation of the residuals

$$\hat{\sigma}_{\hat{n}_p} = \hat{\sigma}_{\hat{n}_q} = \hat{\sigma}_o \sqrt{(1-h_k)/w_k}$$

which are used for the evaluation of the correspondences.

This estimation procedure leads to correct results only if the true correspondences are known. Actually we have no information on which of the elements in the list of preliminary correspondences is correct. But we can now define precisely what we mean by consistency:

1. The true correspondences should fullfill the geometric model. The wrong correspondences therefore can be treated as *outliers* with respect to the geometric model.
2. The true correspondences should form a one-to-one mapping.

We will exploit the requirement on the geometric consistency first. One obvious possibility is to assume either

$$A: \quad n_k \sim N(o, \sigma_k^2 I_2) \quad \text{with probability } 1 - \epsilon$$

or

$$B: \quad n_k \sim U[-L, +L] * U[-L, +L] \quad \text{with probability } \epsilon$$

Case A states that the parallaxes follow a normal distribution with zero mean and a small standard deviation; thus the two points concerned are corresponding ones. Case B states that the parallaxes come from a broad distribution, namely, a uniform distribution with a large L; thus the two points are not corresponding. The probability density function of the standardized true residuals $t = n_k/\sigma_{n_p}$ or $t = n_q/\sigma_{n_q}$ can then be stated as

$$f(x) = (1 - \epsilon) \cdot \varphi(x) + \epsilon \cdot H(x)$$

which is a narrow density function with heavy tails. This gives rise to a robust estimation. The maximum-likelihood estimate can be approximated by the sum of

$$\sum_k \rho\left(\frac{n_k}{\sigma_k}\right) \rightarrow \min$$

with ρ being a nonconvex function. (Here ρ should not be confused with the correlation coefficient used above.) With the method of modified residuals (cf. Huber, 1981), the solution can be found by iteratively weighting down the residuals, with the weight function $w(x) = \rho'(x)/x$ being of the form

$$w(x) = e^{-x^2/2}$$

The idea is that large residuals, indicating wrong correspondences, are weighted down in the course of iterations.

Only for convex minimum functions $\rho(x)$ is convergence guaranteed if the problem is linear (for a more precise statement of these conditions, cf. Huber, 1981).

Therefore the first few iterations should be performed with a convex function $\rho(x)$, which leads to a weight function

$$w(x) = \frac{1}{\sqrt{1+x^2}}$$

Including the iterative adaptation of the variance factor, the total procedure is as follows:

1. Choose initial values for the parameters $a_j^{(o)}, j = 1,\ldots,6$, say (1, 0, 0, 1, 0, 0), and the weights $w_k^{(o)}$ (cf. the discussion above). Set the iteration number $\nu = 0$.
2. Solve the weighted least-squares problem according to Eqs. (16.85) to (16.89), which specifically leads to new estimates $\hat{a}_j^{(\nu+1)}, n_k^{(\nu)}, \hat{\sigma}_o^{2(\nu)}$, and $h_k^{(\nu)}$.
3. Update the weights according to

$$w_k^{(\nu+1)} = w_k^{(o)} \cdot w\left(\frac{|n_k^{(\nu)}| \cdot \sqrt{w_k^{(o)}}}{c \cdot \hat{\sigma}_o^{(\nu)} \cdot \sqrt{1-h_k^{(\nu)}}}\right) \tag{16.90}$$

   Use the weight function $w(x) = 1/\sqrt{1+x^2}$ during the first few (3) iterations; use the weight function $w(x) = e^{-x^2/2}$ for the last two or three iterations; c should be in the range between 1.5 and 2.5.
4. If the change of $\hat{a}_j$ is less than a certain percentage, say 0.1, of their standard deviation, then stop; otherwise, increase ν by 1 and and go to 2.

The results of this robust estimation are parameters $\hat{a}_j$, which in general are close to the true ones and thus are not influenced by the wrong correspondences. In addition, we get weights that—owing to the fast falloff of the exponential function—are practically zero for clearly wrong matches and are significantly less than the original weights. Therefore all correspondences with weights $w_k^{(\nu)}$ that are significantly less than the original weight $w_k^{(o)}$—for example, $< 0.1 w_k^{(o)}$—are rejected and excluded from the following steps.

Though the remaining correspondences now are consistent with the geometric model they need not be unique, as nearby points or image features that are similar to the same points or image features in the other image might still remain. The deviations from the geometric model now intuitively can serve as a criterion for finding a one-to-one mapping: The list of correspondences keeps those which have the smaller—possibly weighted—residual. With these a final least-squares fit for determining the parameters of the geometric model is performed. In this estimation process one must use the original weights, not taking into account the weighting function or the possibly used uniqueness factors. This is necessary if we want to evaluate the final result by using statistical techniques. The *robust estimation* thus is used only *for finding a good hypothesis* for the match. An example is given in Fig. 16.12.

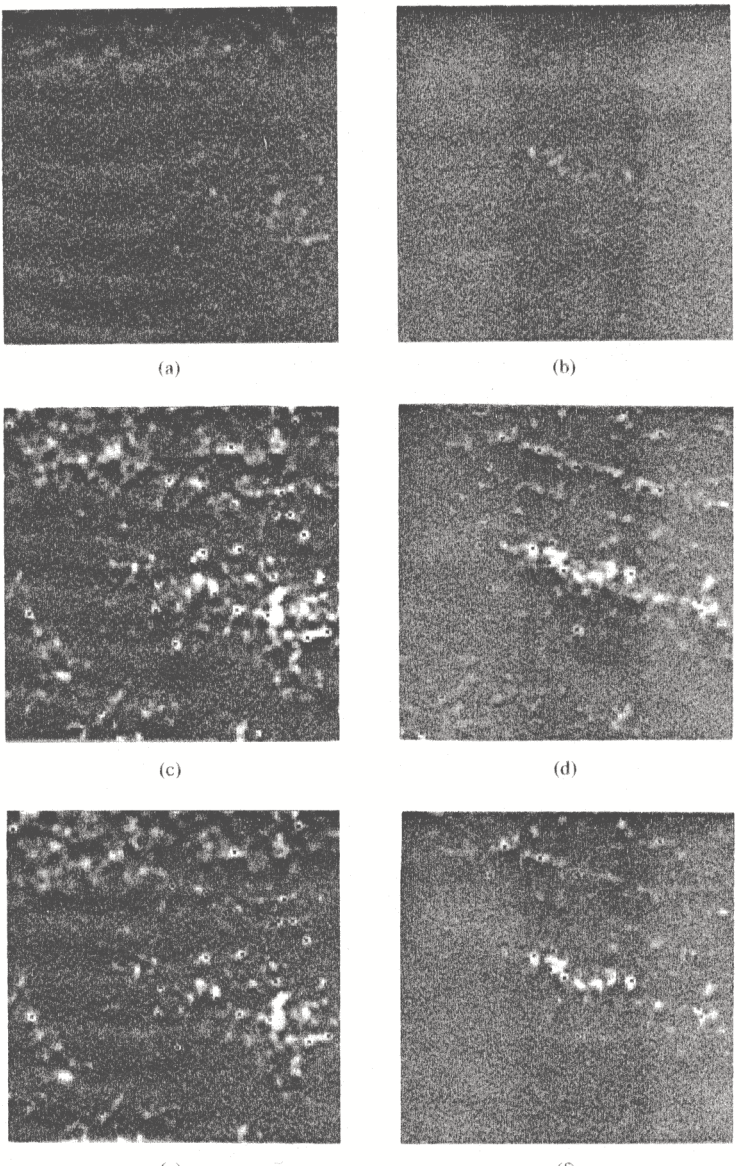


Figure 16.12 Result of the feature-based matching procedure. The points automatically selected in the original images (a) and (b) are shown in (c) and (d). The final correspondences are indicated by circles in (e) and (f). Correspondences with residuals > 2 pixels are indicated by crosses.

16.5.5 Evaluating the Final Result

The result of the previous step, namely, finding the most likely parameters of the mapping function or a one-to-one mapping of the extracted features, finally has to be evaluated. This evaluation is necessary

- to be sure the solution is correct;
- to have quantitative measures for the quality of the parameters and the parallaxes to be used in the following steps.

These requirements motivated the strict separation between hypothesis generation (robust estimation) and hypothesis testing after final parameter estimation using least squares: As the least-squares principle is a special case of the maximum-likelihood principle for Gaussian-distributed random errors, we can use the derived density functions of the estimation processes and apply classical hypothesis tests to evaluate the result:

1. A global check whether data and model are consistent uses Ω, the sum of the squared residuals where the sum is taken over all accepted correspondences.

   The ratio Ω/σ_o^2 distributed with χ_{2m-6} degrees of freedom. The denominator is the variance factor used for determining the weights. In case one were to use *unit weights* in the last estimation step, the ratio

   $$\hat{\sigma}_0^2 = \frac{\Omega}{(2m-6)}$$

   would be an estimate for the average parallax if a linear transformation really held. It could be compared with an expected value σ_0^2 of 1/2 pixel, say, using a Fisher test:

   $$\frac{\hat{\sigma}_0^2}{\sigma_0^2} \approx F_{2m-6,\infty}$$

   If the global test is rejected, one may conclude that the model is oversimplified or the data are much worse than expected. Without further testing, no hint is given on the cause of the rejection. If one is not certain about the a priori value σ_n, one may assume σ_n to be the result of a previous estimation step and assume a limited number $m' < \infty$ of degrees of freedom in the Fisher distribution. This leads to larger critical values $F_{2m-6,m',\alpha} > F_{2m-6,\infty,\alpha}$, and thus to a more conservative test (cf. Spiegelhalter, 1985).
2. The precision of the estimated parameters can be determined in the usual way. Also the precision of the parallaxes may be determined as $\sigma_{p_k} = \sigma_{q_k} = \sigma_0\sqrt{h_k}$.
3. The result can be termed reliable only if enough points are used to determine the spatial-mapping function and if they are distributed well enough that the result is not sensitive to errors in the correspondences. When leaving one correspondence out of the estimation, the effect on the result should be small. This maximum influence on one of the parameters a_i is bounded by

   $$\Delta_k a_i \leq \frac{1}{1-h_k}\sqrt{n_k' n_k \cdot h_k \cdot q_{ii}}$$

with h_k from Eq. (16.89), n_k from Eq. (16.86), and q_{ii} again being the diagonal element of the inverse $Q = N^{-1}$ (for the derivation, see Förstner, 1987), with $1 - h_k$ being identical to the redundancy number r_k.

4. The previous measures refer only to the extracted features, not to the original image. A final check could measure the correlation coefficient (similar to Eq. 16.35) after one has rectified the left image. This projection of one image into the other is the decisive check on the correctness of the matching, as all available information is exploited. Of course a wrong match due to repetitive patterns cannot be detected this way, a limitation common to all tests relying only on the image content. Such errors can be detected only if new information that has not been used in the image-matching step becomes available.

We finally have arrived at describing the match of the two images both by a set of corresponding image features, namely, distinct points, and by a mapping function, enabling one to determine the deformation field at each position.

16.6 Structure from Stereo by Using Correspondence

Now we deal with a special application of image matching. We discuss procedures for recovering the three-dimensional structure of objects from image pairs. This forms a link to Chapter 14 on analytic photogrammetry, where the basic relations were worked out. We assume that the interior orientation of the cameras is determined by some calibration procedure. It could be based on a set of targeted control points given in object space. These targets in the image could be located by the intensity-based matching procedure discussed in Section 16.3. We also assume that at least the relative orientation of the two cameras is known. It could be determined from a set of (≥ 5) matched image points. For reasons of stability at least *five groups of points* should be used (Förstner, 1987), which in general requires no great additional effort. If only the relative orientation is known, we can derive three-dimensional coordinates of the object in a local coordinate system that has an arbitrary origin, orientation, and scale. If in addition some (≥ 3) given points have been identified in the images, we can determine the exterior orientation of the cameras.

16.6.1 Epipolar Geometry

Image matching can be tremendously simplified if the relative orientation is known, as the two-dimensional search space is reduced to a one-dimensional one by the so-called epipolar geometry inherent in the oriented image pair. Figure 16.13 shows the general setup of two cameras. The projection centers O' and O'' form the baseline of length b; the principal points H' and H'' are assumed to be the origin of the two image coordinate systems (u',v') and (u'',v''), derived from the pixel coordinates (r',c') and (r'',c'') by using the interior orientation. The object point $P(x,y,z)$ then is mapped into $P'(u',v')$ and $P''(u'',v'')$ in the image planes μ'

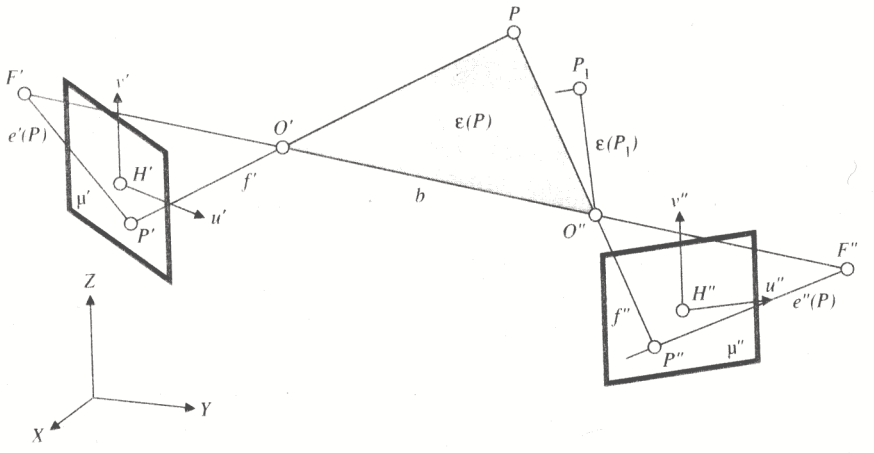


Figure 16.13 Epipolar geometry of a general image pair: image planes f' and f'', projection centers O' and O'', principal points u' and u'', image coordinate systems (u',v') and (u'',v''), baseline b, epipolar plane $\epsilon(P)$, epipolar lines $e'(P)$ and $e''(P)$, and epipoles F' and F''.

and μ''. Because of the geometric model of the perspective projection—specifically the collinearity condition—the five points P, O', O'', P', and P'' lie in one plane, the so-called *epipolar plane* $\epsilon(P)$ associated with P. The intersection lines of $\epsilon(P)$ with μ' and μ'' result in the two *epipolar lines* $e'(P)$ and $e''(P)$ associated with P. For points P_1 not sitting in the same epipolar plane, we obtain different pairs of epipolar lines. All epipolar planes form a pencil of planes passing through the baseline $b = (O'O'')$. The epipolar lines intersect in the *epipoles* F' and F'', which are the intersection of the baseline b intersects with the image planes μ' and μ'', respectively. Thus in general epipolar lines are not parallel.

The main advantage of these geometric relationships is that the epipolar plane $\epsilon(P)$ is defined by P', O', and O''. Thus when only one image point is given, the epipolar line $e''(P)$ is fixed and P'' must sit on this line. Therefore search is necessary in only one dimension. *The epipolar line constraint is the strongest constraint in image matching and should be used as soon as available.* Specifically it is independent of the shape of the object.

Remark:

The epipolar line constraint holds as long as the lens distortions are not too large to be negligible or, more generally, as long as the perspective projection can be modeled by a projective projection, where straight lines in object space map into straight lines in image space. This includes the affine distortion caused by different pixel distances in row and column directions. For high-precision applications the lens distortion (part of the interior orientation) has to be determined and taken into account to correct the observed image (pixel) coordinates in order to exploit the epipolar constraint as a crisp geometric condition. If one does not perform a

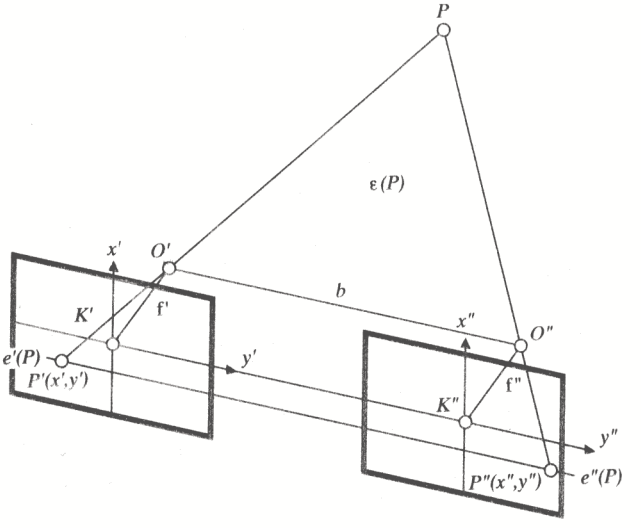


Figure 16.14 Epipolar geometry of a normal image pair: image plane; projection centers O' and O''; principal points K' and K''; image coordinate system (x', y') and (x'', y''); baseline b; x' is parallel to x'', which is parallel to b; focal lengths $f', f''(< 0)$; epipolar plane $\epsilon(P)$; epipolar lines $e'(P)$ and $e''(P)$ identical.

reduction of the observed pixel coordinates for lens distortion, the point P'' corresponding to P' still has to sit on a line that, however, is not straight anymore. In the following we want to neglect possible deviations from the ideal geometry.

If the images are in "normal" position, the determination of epipolar lines reduces to triviality (cf. Fig. 16.14): If the image planes μ' and μ'' are identical and parallel to b, the epipolar lines e' and e'' are parallel to b. Thus the search space for P'', given $P'(x', y')$, is the line $y = y'$.

16.6.2 Generation of Normal Images

It may be useful to rectify evaluation image pairs such that the geometry of a normal image pair can be exploited both for correspondence and for determining the three-dimensional coordinates of object points. Though this advantage cannot be exploited when three or more images are used, it has great computational advantages in the special case of two images, as this rectification has to be performed only once. The method for rectifying the images given below is independent of the representation used for the given pose of the two cameras. It uses the fact that any perspective projection is a projective projection. The principal idea is to replace the images in planes μ' and μ'' by images in plane ν, while keeping the geometry of the bundle of rays spanned by the points in μ' and μ'' and the projection centers O' and O''. Figure 16.15 shows the planes μ' and μ'', the new plane ν, and the images $P', P'', \underline{P'}$,

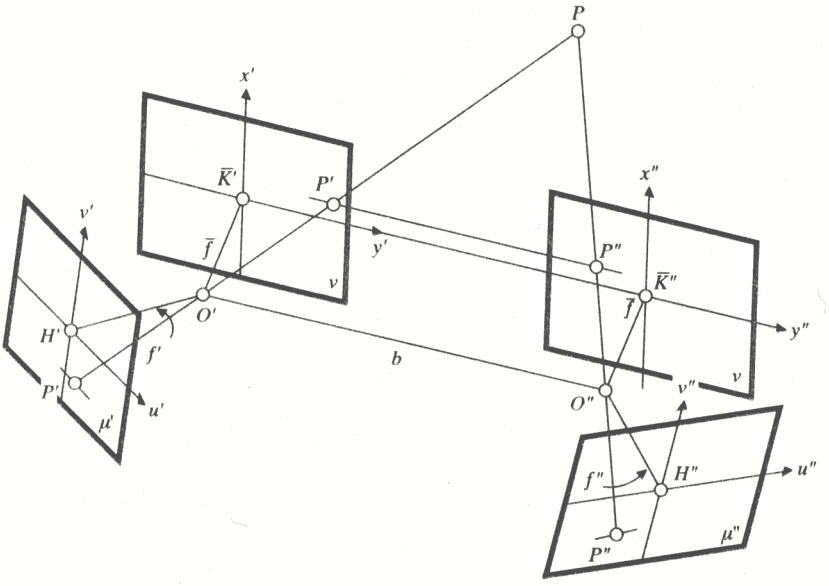


Figure 16.15 Relation between general image pair and normal image pair: common projection centers O' and O'', common baseline; common focal length f; and collinear, normal images in positive position (focal length >0).

and P'' of P. The procedure has to guarantee that P' maps into P. Four points in each image are sufficient. The procedure follows:

1. Choose a plane ν parallel to b. This fixes the focal length f of the normal images. Choose the new image coordinate systems (x', y') and (x'', y'') with origins K' and K'', where x' is parallel to x'', which is parallel to b, and where K' and K'' are the principal points of the new images and $K'O'$ is parallel to $K''O''$, which is parallel to ν. Choose a common pixel spacing in the normal images.
2. For each of the images:
   a. Choose four points well distributed over the new image, for example, forming a square. The coordinates are (x_i, y_i), $i = 1, \ldots, 4$, measured in pixels of the normal image.
   b. Project the four points into the original image, using the known pose of the cameras. This yields four coordinates (u_i, v_i), $i = 1, \ldots, 4$.
   c. Solve the equation system

$$(g\,x_i + h\,y_i + 1) \cdot u_i = a\,x_i + b\,y_i + c$$
$$(g\,x_i + h\,y_i + 1) \cdot v_i = d\,x_i + e\,y_i + f$$

which is linear in the parameters a to h of the projective transformation

$$(u, v) = T(x, y; a, \ldots, h)$$

   d. Rectify the original image. Each pixel (x_j, y_j) in the normal image deter-

mines its position (u_j, v_j) in the original image and derives its intensity value by resampling (linear or higher order).

For this discussion we now assume such normal images to be available.

16.6.3 Specializing the Image-Matching Procedures

In the preceding sections we assumed that no geometric relation between the two images was known. Moreover, we approximated the nonlinear mapping between the images by a linear transformation with six parameters (cf. Eqs. 16.54 and 16.55). We now know that the y-coordinates y' and y'' of corresponding points are equal. Thus the complete model reads as

$$\begin{aligned} x' &= a_1 x'' + a_2 y'' + a_3 \\ y' &= y'' \\ g' &= h_1 g'' + h_2 \end{aligned}$$

We need only five parameters. In an intuitive manner they correspond to depth (a_3), to slope along the epipolar line (a_1), to slope along the epipolar line (a_1), and to brightness (h_2) and contrast (h_1) (cf. Fig. 16.15).

When using the intensity-based matching technique in Section 16.3, only the rows and columns 4 to 6 in Eqs. (16.56) and (16.57) have to be canceled. This is because the parameters a_4, a_5, and a_7 do not have to be determined, owing to the epipolar constraint, and are thus eliminated from the estimation process. The normal equations (16.56) and (16.57) therefore are reduced to a 5×5 linear equation system. The number of degrees of freedom reduces to $m - 5$, changing the ratio $1/(m-8)$ in Eq. (16.58) to $1/(m-5)$.

The feature-based matching algorithm also uses the geometric model from above, but the selection of distinct points has to be modified: We are interested only in the x-parallaxes, as the y-parallaxes are zero. Instead of searching for local maxima of $\sum f_r^2 + f_c^2$, we now only need to search for local maxima of $\sum_r \sum_c f_c^2$.

These local maxima yield the centers of optimal windows whose size should be a minimum of three rows and five columns, which proves to be sufficient in good imagery. If accuracy greater than 1/2 pixel is necessary, larger windows should be used. The optimal point within the window is the weighted center of gravity from Eq. (16.16), but now the sum is evaluated over the entire window, which leads to more stable results.

Remark:

Maximizing f_x^2 is equivalent to searching for the zeros of $2 \cdot f_x \cdot f_{xx}$, and if $f \neq 0$, it is equivalent to determining the zero crossings of the second derivative. But since $f_x \neq 0$, the second derivative $2 \cdot f_{xx}^2 + 2 \cdot f_x \cdot f_{xxx}$ of f_x^2 reduces to $2 \cdot f_x \cdot f_{xxx}$, which is negative because the maxima of f_x^2 are searched for. Thus spurious zero crossings of the second derivative are automatically excluded. Therefore the interest operator, reduced to one dimension, is equivalent to searching edges across the epipolar lines.

16.6.4 Precision of Three-Dimensional Points from Image Points

If we know the coordinates (x', y') and (x'', y'') of corresponding points in the normal images, the three-dimensional coordinates can easily be determined, as shown in Chapter 14. With the object-space coordinate system defined there, we have

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{b}{x'' - x'} \cdot \begin{pmatrix} x' \\ y' \\ f \end{pmatrix} + \begin{pmatrix} -b/2 \\ 0 \\ 0 \end{pmatrix}$$

with the base length b and the focal length f.

Following the line of thought of this chapter, we finally want to determine the precision of the three-dimensional coordinates. Instead of going to the general case with arbitrary orientation of the cameras, we want to restrict the discussion to the coordinates of this "normal model" of the object. If the right base b is chosen, the model is in the same scale as the object.

We first want to give the precision of the z-coordinate. With the parallax $p = x'' - x'$, we have

$$z = \frac{b \cdot f}{p}$$

or $p \cdot z = b \cdot f$, which results in $z\,dp + p\,dz = f\,db + b\,df$, or if p, f, and b are given, we have

$$\frac{dz}{z} = \frac{db}{b} + \frac{df}{f} - \frac{dp}{p}$$

Therefore we obtain the simple relation for the relative precision σ_z/z :

$$\left(\frac{\sigma_z}{z}\right)^2 = \left(\frac{\sigma_b}{b}\right)^2 + \left(\frac{\sigma_f}{f}\right)^2 + \left(\frac{\sigma_p}{p}\right)^2$$

EXAMPLE 16.10

Given

$f = 50\text{mm}$, $\sigma_f = 0.1\text{mm}$
$b = 200\text{mm}$, $\sigma_b = 0.1\text{mm}$
$z = 2\text{ m} = 2000\text{mm}$
$p = fb/z = 5\text{mm}$ $\sigma_p = 0.005\text{mm}$

we have

$$\left(\frac{\sigma_z}{z}\right)^2 = \left(\frac{1}{500}\right)^2 + \left(\frac{1}{2000}\right)^2 + \left(\frac{1}{1000}\right)^2$$
$$= \frac{16 + 1 + 4}{2000^2} = \frac{21}{2000^2}$$

thus a relative precision of $\sqrt{21}/2000 \approx 1/400$, and so $\sigma_z = \sqrt{21}\text{mm} = 4.5$ mm. Obviously the worst relative precision counts, which in this case is the focal length with 1/500 relative precision.

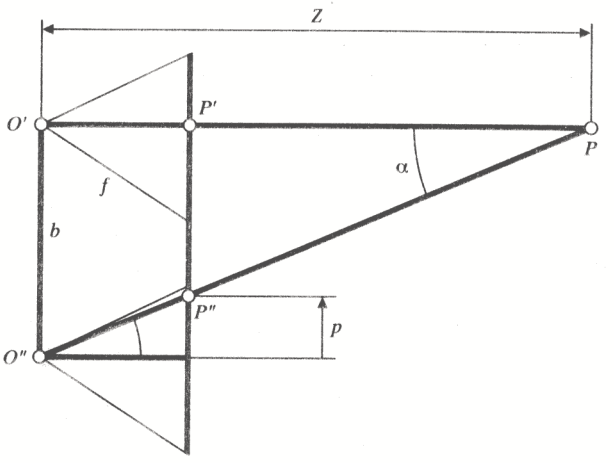


Figure 16.16 Assumed relation between parallax p and parallactic angle α.

In the case of perfect calibration and orientation, the relative precision of z reduces to

$$\frac{\sigma_z}{z} = \frac{\sigma_p}{p}$$

an extremely simple relation to remember. Using $z = fb/p$, we also obtain

$$\frac{\sigma_z}{z} = \frac{z}{f} \cdot \frac{\sigma_p}{b}$$

which relates the relative precision σ/z of z to the relative distance z/f, and the relative precision σ_p/b of the parallax, measured in units of the baseline. The relative precision thus decreases linearly with distance of p.

Finally, we can use the *parallactic angle* $\alpha \approx \arctan(p/f)$ at point P (cf. Fig. 16.16) to relate the relative distance precision to the angular error $\sigma_\alpha = \sigma_p/f$:

$$\frac{\sigma_z}{z} = \frac{z}{b} \cdot \frac{\sigma_p}{f} = \cot\alpha \cdot \sigma_\alpha$$

or for small angles α,

$$\frac{\sigma_z}{z} \approx \frac{\sigma_\alpha}{\alpha}$$

Thus for angles not much larger than σ_α, the relative precision of z also is very low. This situation may easily occur in navigation applications, when the time interval between successive frames is short.

EXAMPLE 16.11

The precision of z under the conditions mentioned above, but with f and b error free, is

$$\sigma_z = z \cdot \frac{\sigma_p}{p} = 2\text{mm}$$

which obviously would be more than a factor 2 too optimistic if the orientation parameters (f and b) were uncertain. We can also derive the precision of depth differences $h = z_2 - z_1$ of two points of an object, which is given by

$$h = z_2 - z_1 = f\ b \cdot \left(\frac{1}{p_2} - \frac{1}{p_1}\right)$$

We obtain for $z_1 \approx z_2 \approx z$, thus $h/z << 1$:

$$\left(\frac{\sigma_h}{h}\right)^2 = \left(\frac{\sigma_f}{f}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 + 2\left(\frac{z}{h}\right)^2 \cdot \left(\frac{\sigma_p}{p}\right)^2$$

EXAMPLE 16.12

With the same assumptions as above, namely,

$$\sigma_f/f = 1/500,$$

$$\sigma_b/f = 1/2000,$$

$$\sigma_p/p = 1/1000$$

$$h = z/5 = 400\text{mm}$$

we obtain

$$\left(\frac{\sigma_h}{h}\right)^2 = \left(\frac{1}{500}\right)^2 + \left(\frac{1}{2000}\right)^2 + 2 \cdot 25 \cdot \left(\frac{1}{1000}\right)^2$$

$$= \frac{16 + 1 + 2 \cdot 25 \cdot 4}{2000^2} = \frac{217}{2000^2}$$

and thus

$$\sigma_h = \sqrt{217}/2000 \cdot 400\text{mm} \approx 3\text{mm}$$

This is *less* than $\sigma_z = 4.5$mm from above.

Observe that the effect of the errors in f and b are negligible here, $\sqrt{17}/2000^2$ against $\sqrt{200}/2000^2$. The reason why the depth differences are much more precise is that possible errors in f and b cancel when taking the difference $z_2 - z_1$.

This result can be generalized for the case when all parameters of the pose are uncertain. The precision of the absolute depth depends strongly on the precision of the relative depth, or depth differences depend mainly on the parallax precision. Thus

$$\frac{\sigma_h}{h} = \sqrt{2} \cdot \frac{\sigma_p}{p}$$

Nearly all other effects cancel. These results can easily be applied:

1. If the form of an object has to be determined, only relative depth is required. Thus most orientation errors cancel, and the precision of the form is determined by the accuracy of the parallaxes.
2. The same holds if one reference point in object space is given and the pose of the object has to be determined with respect to this (close) reference point.
3. However, if the pose of the camera has to be determined relative to some object, as in navigation, *all* orientation errors influence the overall accuracy. Thus all orientation parameters have to be determined with great care.

These are rules of thumb; a proper planning should apply a rigorous analysis.

Finally, we give the accuracy of the x- and y-coordinates of the point in object space:

$$x = z\frac{x'}{f} - \frac{b}{2}$$

$$y = \frac{z \cdot y'}{f}$$

Under the assumption $\sigma_f = \sigma_b = 0$,

$$\left(\frac{\sigma_x}{x}\right)^2 = \left(\frac{\sigma_z}{z}\right)^2 + \left(\frac{\sigma_{x'}}{x'}\right)^2$$

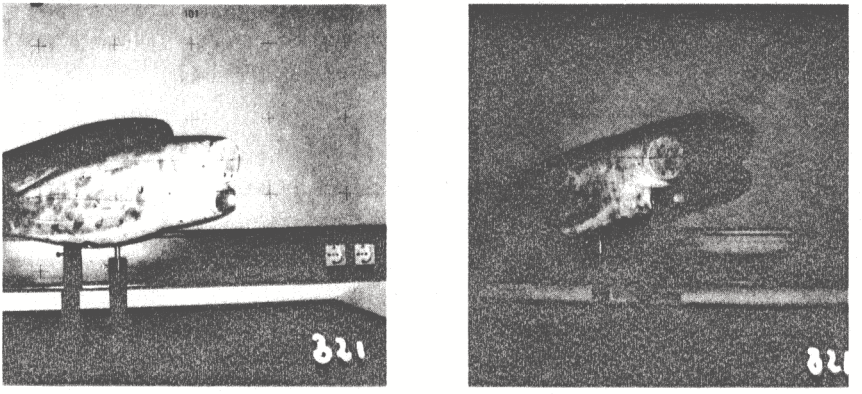
and

$$\left(\frac{\sigma_y}{y}\right)^2 = \left(\frac{\sigma_z}{z}\right)^2 + \left(\frac{\sigma_{y'}}{y'}\right)^2$$

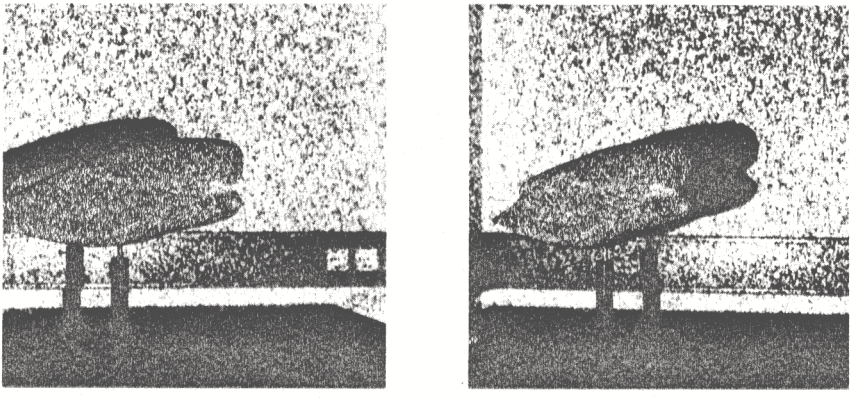
EXAMPLE 16.13

Figure 16.17(a) shows a stereo image pair taken with a photogrammetric stereo camera. The object is a model of an engine crank case. The task is to determine the object's surface along profiles in prespecified planes in order to provide the input for a CAD system, where the model is then processed further. The accuracy requirements are 0.4mm tolerance. In order to cover the complete surface, six stereo image pairs were taken. The orientation of the camera is determined on the basis of precisely measured targeted control points, partly visible in Fig. 16.17.

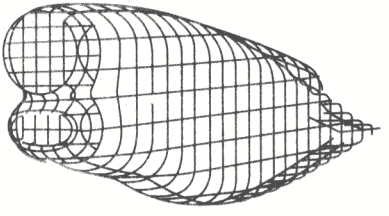
The baseline of the stereo camera is $b = 0.8$m; the distance to the object is approximately 1.5m; the focal length f is 0.1m. Figure 16.17(b) shows a stereo image pair of the same object now illuminated with a texture projector to permit measurements over the complete surface.



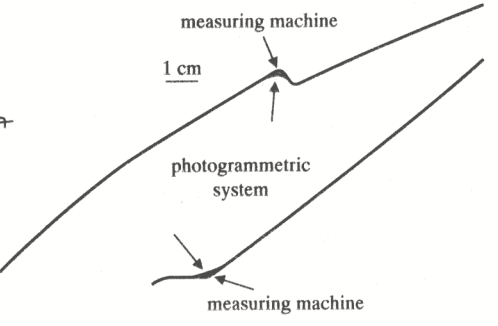
(a)



(b)



(c)



(d)

Figure 16.17 (a) stereo pair; (b) illustrated by texture projects (c) three-dimensional (d) profile compression.

The images are digitized with a pixel size of 20μm. Interactively given an initial match, the profiles are automatically measured by using the differential approach discussed in Section 16.3 with the specific model of Section 16.6.3 (cf. Schewe and Förstner, 1986; Schewe, 1989). The result of the measured profiles of four of the six stereo image pairs is shown in Fig. 16.17(c). Figure 16.17(d) shows two pairs of profiles, one measured with the photogrammetric system including the automatic matching procedure, and the other measured with a classical mechanical three-dimensional measuring machine. The deviations mainly are below 0.2μm. Only at the edges of the surface does the photogrammetric system produce a certain amount of blending, which is due to the limitation of the model.

Exercises

16.1. Image matching, in contrast to object reconstruction, should be used only if a geometric transformation between the images is an appropriate representation of the model or if certain invariant low-level features can be found in the images of concern. Give examples for stereo tasks in which image matching is appropriate: (1) for the complete image content; (2) for 80% of the image content; and (3) for less than 50% of the image content shown in both images.

16.2. You want to determine the distance between two holes of an object by using a digital camera with an accuracy of 1:10000. Is this possible? Describe the mensuration design, the size of the camera used (in pixels), and the technique used for locating the hole (refer to Table 16.2).

16.3. Give practical examples for the different types of object surfaces discussed in Fig. 16.2.

16.4. (Reading study.) Study the papers of Barnard and Thompson (1980), Stockman (1987), Ohta and Kanade (1985), and Witkin, Terzopoulos, and Kass (1987): (1) What type of surfaces do they refer to when applied to image matching? (2) How do they achieve consistent results? (3) How far do the algorithms provide means for evaluating the final result?

16.5. (Computer study.) Use Example 16.1 and apply it to the models "shift and scale," "shift + brightness + contrast," "shift + brightness," and all four parameters. Repeat the experiment with: (a) two ideal, noise-free ramp edges; (b) ramp edges with artificial noise ($\sigma_n = 2$); and (c) a noisy box with a shift that is not an integer. Generate the data using noninteger values for the geometric and radiometric transformation parameters. First use real numbers for the gray values, then use gray values rounded to integers. Compare the results with respect to the actual differences in the transformation parameters, the theoretical standard deviations of the estimates, assuming $\sigma_n = 2$, and the empirical standard deviation of the estimated parameters.

16.6. Name conditions such that the intensities g' and g'' at the corresponding points of two images can be related reasonably well by

$$\text{a.} \quad g'(r',c') = g''(r'',c'')$$

or

$$\text{b.} \quad g'(r',c') = ag''(r'',c'') + b$$

Name typical situations in which these simple models do not hold (refer to Chapter 12 on illumination).

16.7. Name conditions such that extracted edges or edge points refer to the same object point when: (a) the camera is moving; and (b) the lighting conditions change. Name conditions such that extracted edges or edge points do *not* refer to the same object point when: (a) the position of the camera is changing; (b) the position of the light source is changing; and (c) the pose of the object is changing.

16.8. Under which conditions are two images enough for reconstructing the surface visible in both images? Under which conditions may a third image be of advantage? In which geometric relation should the three cameras be? Refer to Sections 16.3–16.6 and explain where the inability of matching possibly shows up. What objective criteria could be used to detect situations in which no image matching is possible with two and three cameras, respectively? (*Hint:* Analyze the matching ability in dependency on the type of texture, on the orientation of edges, and on the complexity of the object's surface.)

16.9. The ratio $\underline{T} = \underline{\Omega}/\sigma^2 = (m-1) \cdot \underline{\hat{\sigma}_n}^2/\sigma_n^2$ (cf. Eq. 16.14) is χ^2-distributed with $m-1$ degrees of freedom. The expectation and variance of $\underline{T}$ is given by

$$E(\underline{T}) = m - 1, \quad V(\underline{T}) = \sigma_{T^2} = 2(m-1).$$

Prove Eq. (16.15). (*Hint:* If two stochastic variables are related by $\underline{s} = \sqrt{\underline{t}}$, then their standard deviations are approximately related by $\sigma \approx \frac{1}{2}\sigma_t \sqrt{E(\underline{t})}$. Apply this approximation to $\underline{t} = \underline{\hat{\sigma}_n}^2$ after having derived the variance of $\underline{\hat{\sigma}_n}^2 = i\underline{T} \cdot \sigma_n^2/(m-1)$. Stochastic variables are underscored.)

16.10. Estimate the weighted center of gravity of: (a) a smoothed step edge

$$g_a(x) \begin{cases} a/c \cdot x & 0 \le x \le c \\ a & x > c \\ 0 & x < 0 \end{cases}$$

$$(c \neq 0)$$

(b) a smoothed box

$$g_b(x) = g_a(x+d) - g_a(x-e)$$

(c) a skew triangle

$$g_c(x) = \begin{cases} \frac{c}{a}x + c & -a \le x \le 0 \\ a & x > c \\ 0 & \text{otherwise} \end{cases}$$

$$(a \neq 0, b \neq 0)$$

with the window w not sitting symmetric to the signal and using integrals instead of sums $s : x_o = \int_w x \cdot f^{12}(x)dx / \int_w f^{12}(x)dx$ (see Fig. 16.18). Compare the center of gravity to the center of the window. Vary the window size and its position,

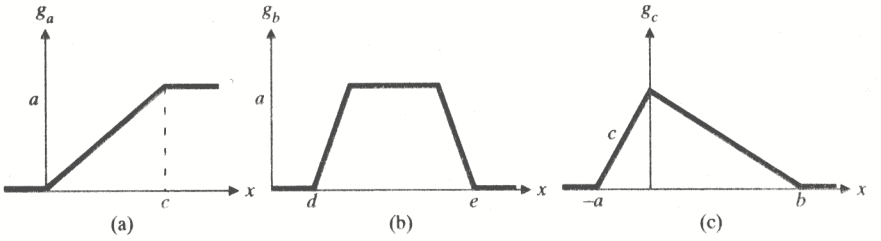


Figure 16.18 Images for Exercise 16.10.

that is, the lower and upper bounds of the integrals. Repeat the comparison with sampled data.

16.11. Prove the precision of the weighted center of gravity to be identical to Eq. (16.9).

16.12. (a) Apply a second iteration to Example 16.1 (cf. Exercise 16.2). Evaluate the improvement in position with respect to the standard deviation of the shifts. Repeat the experiment for two profiles that differ by more than two pixels in position. Does the second (or a third) iteration significantly improve the result? (b) Assume two ramp edges of width d and arbitrary height being shifted by an amount $> d$ with respect to each other. Explain why no solution for the shift can be obtained when using the differential approach. Where does this show up in the practical solution (use simulated data)? Compare the difference of the position of the weighted centers of gravity with the generated shift between the ramps. Comment on the results of (a) and (b) with respect to the rate and the range of convergency of the differential approach and its limitations and on the range of convergency of the feature-based matching techniques.

16.13. Use robust estimation for the shift of two profiles, one having one or two outliers. How does the robust procedure react (a) if the outliers are at a flat portion of the signal? (b) if the outliers are at an edge?

16.14. Generalize the subpixel estimate using correlation and its precision to two dimensions. (a) Show that the position can be obtained from

$$(\hat{r},\hat{c})' = (r_0,c_0)' - [H\rho|_{(r_0,c_0)}]^{-1} \cdot \nabla\rho|_{(r_0,c_0)}$$

where $H\rho$ and $\nabla\rho$ are the Hessian and the gradient of $\rho(r,c)$

$$H\rho = \begin{pmatrix} \rho_{rr} & \rho_{rc} \\ \rho_{cr} & \rho_{cc} \end{pmatrix} \qquad \nabla\rho = \begin{pmatrix} \rho_r \\ \rho_c \end{pmatrix}$$

evaluated at the integer position (r_0,c_0), being the shift where the correlation coefficient is maximum. The gradient can be determined by using the Sobel operator, and the elements of the Hessian by using a similar convolution kernel. (b) Show that the covariance matrix of the estimated shift can be determined from

$$D\begin{pmatrix} \hat{r} \\ \hat{c} \end{pmatrix} = \frac{1}{m} \cdot \frac{1-\rho_{12}}{\rho_{12}} \cdot [-H\rho|_{(r_0,c_0)}]^{-1} \cdot \Delta x^2$$

16.15. How can the optical flow equation be generalized: (a) in the case of different brightness? (b) in the case of different brightness and contrast? (*Hint:* Refer to model Eq. 16.16 or use an appropriate function of the image densities.)

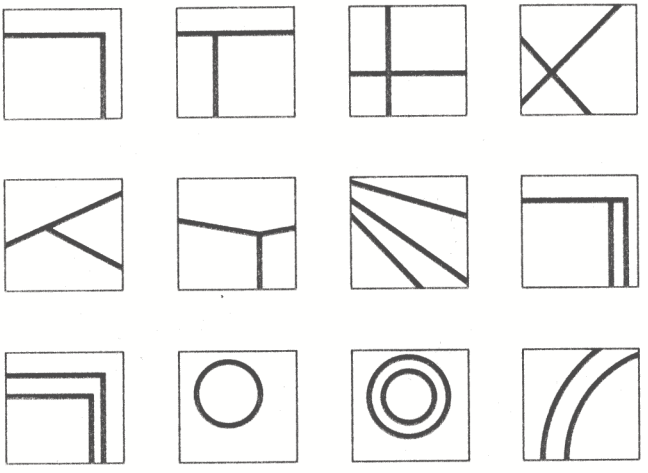


Figure 16.19 Windows for Exercise 16.17.

16.16. Specialize the estimate for a constant shift in two dimensions (Eq. 16.48) for binary images. Use logical operations. (*Hint:* $f_r(r,c)$ and $f_c(r,c)$ can only have the values $-1, 0$, and $+1$. The positive part f_{r+} of f_r can be determined from $f_{r+} = \{[f(r,c)$ xor $f(r+1,c)]$ and $f(r+1,c)\}$, while the negative part is $f_{r-} = \{[f(r,c)$ xor $f(r+1,c)]$ and $f(r,c)\}$, assuming 0=false and 1=true. Thus, for example, the sum $S = \sum_{rc} f_r(r,c)$ can be split into $S = S_+ - S_-$, with $S_+ = \sum_{rc} f_{r+}$ and $S_- = \sum_{rc} f_{r-}$.)

16.17. Apply the interest operator to the images in Fig. 16.19. (a) Compare the center of the optimal window with the estimated point. (b) Vary the window size used and describe the effect on the optimal position of the window and on the estimated point within the window.

16.18. Show that the correlation matrix of the three windows in Fig. 16.20 is

$$R = \begin{pmatrix} 1.000 & 0.033 & 0.199 \\ 0.033 & 1.000 & 0.260 \\ 0.199 & 0.260 & 1.000 \end{pmatrix}$$

and the uniqueness is

$$u_1 = 4.077,\ u_2 = 2.868,\ u_3 = 2.122$$

16.19. Apply the uniqueness measure to the symbol sequences "south," "north," and "august." The Levenshtein distances are

$$D = \begin{pmatrix} 0 & 2 & 5 \\ 2 & 0 & 6 \\ 5 & 6 & 0 \end{pmatrix}$$

Use the correlation function $\rho_{ij} = 1/[1+(d_{ij}/l_{\max})^2]$, $l_{\max}$ being the maximal length of either symbol sequence, to show that the uniqueness of the three sequences is

$$u_1 = 0.136,\ u_2 = 0.161,\ u_3 = 0.697$$

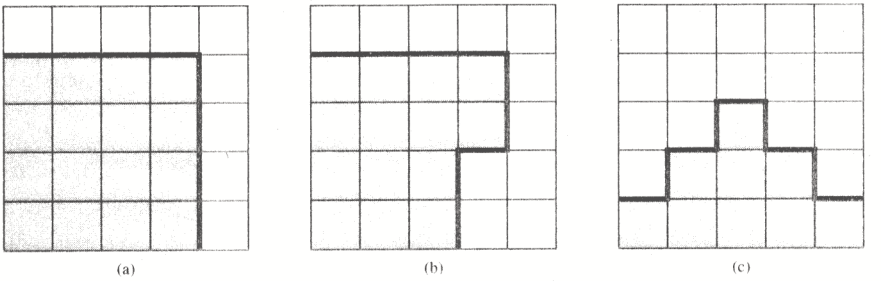


Figure 16.20 Three windows for Exercise 16.18.

16.20. Refer to Chapter 14 (analytical photogrammetry) and assume two cameras in the normal case. The focal length is $f = 35mm$, the base length is $b = 1\text{m}$, the pixel size is $20\mu\text{m}$, and the distance to the object varies between 9 and 10m. The standard deviation of the parallaxes is 0.1 pixels. (a) How accurately can you determine the length l of the object parallel to the basis? (b) Determine the standard deviation of the distance of a point relative to the cameras. (c) How accurately can the depth difference $d_1 - d_2$ of two points be determined? Assume $h = d_1 - d_2 = 1\text{m}$. What relative accuracy σ_h / h can you achieve?

Bibliography

Ackermann, F., and A. Pertl, "Zuordnung Leleiner Bildflächen durch digitale Korrelation zur Verknüpfung verschiedener oder verschniedenartiger Bilder in Anwendingsbereich," *Photogrammetrie und Fernerkindung*, DFG-Abschlupbericht, 1983.

Akey, M. L., and R. Mitchell, "Detection and Subpixel Location of Objects in Digitized Aerial Images," *Proceedings of the Seventh Conference on Pattern Recognition*, Montreal, 1984, pp. 411–414.

Alliney, S., and C. Morandi, "Digital Image Registration Using Projections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, 1986, pp. 222–233.

Altmann, J., and H. J. P. Reitbock, "A Fast Correlation Method for Scale-and Translation-Invariant Pattern Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, 1984, pp. 46–57.

Ayache, N., and B. Faverjon, "A Fast Stereovision Matcher Based on Prediction and Recursive Verification of Hypothesis," *Proceedings of the Third Workshop on Computer Vision*, Bellaire, MI, 1985, pp. 27–37.

———, "Efficient Registration of Stereo Images by Matching Graph Descriptions of Edge Segments," *International Journal on Computer Vision*, Vol. 1, 1987, pp. 107–131.

Baker, H. H., and T. O. Binford, "Depth from Edges and Intensity Based Stereo," *Proceedings of the International Joint Conference on Artificial Intelligence*, Vancouver, 1982, pp. 631–636.

Ballard, D. H., and C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

Barnard, S. T., "A Stochastic Approach to Stereo Vision," *Proceedings of the Fifth National Conference on Artificial Intelligence,* Philadelphia, 1986, pp. 676-690.

———, "Stereo Matching by Hierarchical, Microcanonical Annealing," *Proceedings of the Image Understanding Workshop,* Los Angeles, 1987, pp. 792–797.

Barnard, S. T., and W. B. Thompson, "Disparity Analysis of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-2, 1980, pp. 333–340.

Barnea, D. I., and H. F. Silverman, "A Class of Algorithm for Fast Digital Image Registration," *IEEE Transactions on Computers,* Vol. C-21, 1972, pp. 179–186.

Benard, M., "Automatic Stereophotogrammetry: A Method Based on Feature Detection and Dynamic Programming," *Proceedings of the Specialist Workshop on Pattern Recognition in Photogrammetry,* Graz, Austria, 1983.

Berenstein, C. A., L. N. Kanal, D. Lavine and E. C. Olson, "A Geometric Approach to Subpixel Registration Accuracy," *Computer Vision, Graphics, and Image Processing,* Vol. 40, 1987, pp. 334–360.

Berman, M., "Large Sample Bias in Least Squares Estimators of a Circular Arc and Its Radius," *Computer Vision, Graphics, and Image Processing,* Vol. 45, 1989, pp. 126–128.

Bernstein, R., "Scene Correction (Precision Processing) of ERTS Sensor Data Using Digital Image Processing Techniques," *Proceedings of the Third ERTS Symposium,* Vol. 1-A, NASA SP-351, 1973.

———, "Image Geometry and Rectification," *Manual of Remote Sensing,* 2d ed., American Society Photogrammetrics, Little Falls Church, Va, 1983, chap. 21.

Bigün, J., "A Structure Feature for Some Image Processing Applications Based on Spiral Functions," *Computer Vision, Graphics, and Image Processing,* Vol. 51, 1990, pp. 166–194.

Bigün, J., and G. H. Granlund, "Optimal Orientation Detection of Linear Symmetry," *Proceedings of the First International Conference on Computer Vision,* London, 1987, pp. 433–438.

Blostein, S. T., and T. S. Huang, "Quantization Errors in Stereo Triangulation," *Proceedings of the First International Conference on Computer Vision,* London, 1987, pp. 325–334.

Bolles, R.C., L.H. Quam, M.A. Fischler, and H.C. Wolf, "Automatic Determination of Image-to-Database Correspondences," *Proceedings of the Sixth International Joint Conference on Artificial Intelligence*, Tokyo, 1979, pp. 73–78.

Borgefors, G., "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 10, 1988, pp. 849–865.

Bouthemy, P., and A. Benveniste, "Modeling of Atmospheric Disturbances in Meteorological Pictures," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-6, 1984, pp. 587–600.

Boyer, K. L., and A. C. Kak, "Symbolic Stereo from Structural Descriptions," School of Electrical Engineering, Purdue University, West Lafayette, IN, TR-EE 86-12, 1986.

———, "Structural Stereopsis for 3-D Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-10, 1988, pp. 144–166.

Burkhardt, H., and H. Moll, "A Modified Newton-Raphson Search for the Model-Adaptive Identification of Delays," *Proceedings of the IFAC Symposium on the Identification and System Parameter Estimation,* Darmstadt, W. Germany, 1979, pp. 1279–86.

Burr, D. J., "A Dynamic Model for Image Registration," *Computer Graphics and Image Processing,* Vol. 15, 1981, pp. 102–112.

Cafforio, C., and F. Rocca, "Methods for Measuring Small Displacements of Television Images," *IEEE Transaction on Information Theory,* Vol. 22, 1976, pp. 573–579.

Castro, E. de, and C. Morandi, "Registration and Rotated Images Using Finite Fourier Transforms," *IEEE Transactions on Pattern Matching and Machine Intelligence,* Vol. PAMI-9, 1987, pp. 700–703.

Cernuschi-Frias, B., et al., "Toward a Model-Based Bayesian Theory for Estimating and Recognizing Parameterized 3-D Objects Using Two or More Images Taken from Different Positions," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-11, 1989, pp. 1028–52.

Dreschler, L., "Ermittlung markanter Punkte auf den Bildern bewegter Objekte und Berechnung einer 3D-Beschreibung auf dieser Grundlage," *Dissertation Fachbereich Informatik,* Universität Hamburg, 1981.

Ebner, H., "Berücksichtigung der lokalen Geländeform bei der Höheninter polation mit finiten Elementen," *Bildmessung und Luftbildwesen,* Vol. 51, 1983, pp. 3–9.

Ebner, H., et al., "Integration von Bildzuordnung und Objektrekonstruktion innerhalb der Digitalen Photogrammetrie," *Bildmessung und Lufbildwesen,* Vol. 55, 1987, pp. 194–203.

Förstner, W., "On the Geometric Precision of Digital Correlation," *International Archives of Photogrammetry and Remote Sensing,* Vol. 24-III, Helsinki, 1982, pp. 176–189.

——, "Reliability and Discernability of Extended Gauss-Markov Models," *Deutsche Geodätische Kommision,* München A 98, Munich, 1983, pp. 79–103.

——, "Quality Assessment of Object Location and Point Transfer Using Digital Image Correlation Techniques," *International Archives of Photogrammetry and Remote Sensing,* Vol. 25-A3a, Rio de Janeiro, 1984, pp. 197–219.

——, "Determination of the Additive Noise Variance in Observed Autoregressive Processes Using Variance Component Estimation Techniques," *Statistics and Decisions,* Suppl. Issue No. 2, 1985, pp. 263–274.

——, "A Feature Based Correspondence Algorithm for Image Matching," *International Archives of Photogrammetry and Remote Sensing,* Vol. 26-3/3, Rovaniemi, 1986a, pp. 150–166.

——, "On Automatic Measurement of Digital Surface Models," *Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart 11,* Stuttgart, 1986b, pp. 69–90.

——, "Reliability Analysis of Parameter Estimation in Linear Models with Applications to Mensuration Problems in Computer Vision," *Computer Vision, Graphics, and Image Processing,* Vol. 40, 1987, pp. 273–310.

——, "Statistische Verfahren für die automatische Bildanalyse und ihre Bewertung bei der Objekterkennung und- vermessung," Habilitationsschrift, Stuttgart, 1988.

Förstner, W., and E. Gülch, "A Fast Operator for Detection and Precise Location of Distinct Points, Corners, and Centres of Circular Features," *Proceedings of the Intercommission Conference on Fast Processing of Photogrammetric Data,* Interlaken, Switzerland, 1987, pp. 281-305.

Gale, W.A., *Artificial Intelligence and Statistics,* Addison-Wesley, Reading, MA, 1985.

Giri, N.C., *Multivariate Statistical Inference,* Academic Press, New York, 1977.

Goshtasby, A., "Piecewise Cubic Mapping Functions for Image Registration," *Pattern Recognition,* Vol. 20, 1987, pp. 525–533.

——, "Image Registration by Local Approximation Methods," *Image and Vision Computing,* Vol. 6, 1988, pp. 255–261.

Grimson, W. E. L., *From Images to Surfaces: A Computational Study on the Human Early Visual System,* MIT Press, Cambridge, MA, 1981.

——, "Computational Experiments with a Feature Based Stereo Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-7, 1985, pp. 17–34.

Grün, A., and E. Baltsavias, "Adaptive Least Squares Correlation with Geometrical Constraints," *Proceedings of the SPIE,* Vol. 595, Cannes, 1985.

Haggren, H., "Photogrammetric Prototype System for Real-Time Engineering Applications," *Optics in Engineering Measurement,* SPIE Proceedings, Vol. 599, 1985, pp. 330–335.

Hannah, M. J., "Computer Matching of Areas in Stereo Images," Ph.D. Diss., Stanford University, Stanford, CA, Report STAN-CS-74-483, 1974.

——, "A System for Digital Stereo Matching," *Phototgrammetric Engineering & Remote Sensing,* 1989.

Haralick, R. M., "Statistical and Structural Approach to Texture," *Proceedings of the IEEE,* Vol. 67, 1979, pp. 786–804.

Helava, U. V., "Digital Correlation in Photogrammetric Instruments," *International Archives of Photogrammetry and Remote Sensing,* Vol. 23-2, Helsinki, 1976.

——, "Object-Space Least-Squares Correlation," *Phototgrammetric Engineering & Remote Sensing,* Vol. 54, 1988, pp. 711–714.

Herbin, M., et al., "Automated Registration of Dissimilar Images: Application to Medical Imagery," *Computer Vision, Graphics, and Image Processing,* Vol. 47, 1989, pp. 77–88.

Herman, M., and T. Kanade, "The 3D Mosaic Scene Understanding System," in Pentland, 1986, pp. 322–358.

Hill, J. W., "Dimensional Measurement for Quantized Images," *SRI Project 4391,* Stanford Research Institute, Menlo Park, 1980.

Ho, C. S., "Precision of Digital Vision Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-5, 1983, pp. 593–601.

Horn, B. K. P., *Robot Vision,* McGraw-Hill, New York, 1987.

Horn, B. K. P., and B. L. Bachman, "Using Synthetic Images to Register Real Images with Surface Models," *Communications of the ACM,* Vol. 21, 1978, pp. 914–924.

Huang, T. S. (ed.), *Image Sequence Analysis,* Springer, New York, 1981.

Huber, P., *Robust Statistics,* Wiley, New York, 1981.

Kass, M., and A. Witkin, "Analysing Oriented Patterns," *Computer Vision, Graphics, and Image Processing,* Vol. 37, 1987, pp. 362–385.

Klaasman, H., "Some Aspects on the Accuracy of the Approximated Position of a Straight Line on a Square Grid," *Computer Graphics and Image Processing,* Vol. 4, 1975, pp. 225–235.

Koch, K. R., *Parameterschätzung und Hypothesentests,* in linearen Modellen, Dümmler, 1987.

Kories, R. R., "Bildzuordnungsverfahren für die Auswertung von Bildfolgen," *Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart,* Vol. 11, 1986, pp. 157–168.

Lam, K. P., "Position Determination Using Generalized Multidirectional Gradient Codes," *Computer Vision, Graphics, and Image Processing,* Vol. 28, 1984, pp. 228–239.

Lavine, D., B. A. Lambird, and L. N. Kanae, "Recognition of Spatial Point Patterns," *Pattern Recognition,* Vol. 16, 1983, pp. 289–295.

Li, X., and R. C. Dubes, "The First Stage in Two-Stage Template Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-7, 1985, pp. 700–707.

Lindenberger, J., "Consideration of Observation Errors when Modelling Digital Terrain Profiles," *Proceedings of the Workshop on Progress in Digital Terrain Modelling of WG III/3 of the International Society of Photogrammetry and Remote Sensing,* Lyngby, Denmark, 1987, pp. 227-238.

Longuet-Higgins, M. S., "The Statistical Analysis of a Random Moving Surface," *Philosophical Transactions of the Royal Society of London,* Ser. A 249, 1957, pp. 321–387.

Maître, H., and Y. Wu, "Improving Dynamic Programming to Solve Image Registration," *Pattern Recognition,* Vol. 20, 1987, pp. 443–462.

Markarian, H., et al., "Digital Correction for High-Resolution Images," *Photogrammetric Engineering,* Vol. 39, 1973, pp. 1311–20.

———, *Implementation of Digital Techniques for Correcting High Resolution Images,* American Institute of Aeronautics and Astronautics, Report No. A72–10454, 1972.

McClure, D. E., "Image Models in Pattern Theory," *Image Modeling*, Rosenfeld (ed.), Academic Press, New York, 1981, pp. 259–276.

McGillem, C. D., and M. Svedlov, "Image Restoration Error Variance as a Measure of Overlay Quality," *IEEE Transactions on Geoscience Electronics,* Vol. 14, 1976, pp. 44–49.

———, "Optimum Filter for Minimizing of Image Registration Error Variance," *IEEE Transactions on Geoscience Electronics,* Vol. 15, 1977, pp. 257–259.

Medioni, G., and R. Nevatia, "Matching Images Using Linear Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-6, 1984, pp. 675–685.

Merickel, M., and M. McCarthy, "Registration of Contours for 3-D Reconstruction," in *Proceedings of the Seventh Annual Conference of the IEEE Engineering in Medicine and Biology Society,* M. Merickel and M. McCarthy (eds.), Vol. 1, 1985, pp. 616–620.

Mikhail, E. M., and F. Ackermann, *Observations and Least Squares,* Dun-Donelly, New York, 1976.

Milgram, D. L., "Computer Methods for Creating Photomosaics," *IEEE Transactions on Computers, Correspondence,* 1975, pp. 1113–19.

Moravec, H., *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover,* Technical Report CMU-Ri-TR3, Carnegie-Mellon University, Pittsburgh, 1980.

Munteau, C., "Evaluation of the Sequential Similarity Detection Algorithm Applied to Binary Images," *Pattern Recognition,* Vol. 13, 1981, pp. 167–175.

Nagel, H.-H., "Displacement Vectors Derived from Second Order Intensity Variations in Image Sequences," *Computer Vision, Graphics, and Image Processing,* Vol. 21, 1981, pp. 85–117.

Nagel, H.-H., and W. Enkelmann, "Iterative Estimation of Displacement Vector Fields from TV-Frame Sequences," *Proceedings of the Second European Signal Processing Conference,* Erlangen, Germany, 1983, pp. 299–302.

———, "An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-8, 1986, pp. 565-593.

Negahdaripour, S., and B. K. P. Horn, "A Direct Method for Locating the Focus of Expansion," *Computer Vision, Graphics, and Image Processing,* Vol. 46, 1989, pp. 303–326.

Ogawa, H., "Labeled Point Pattern Matching by Delaunay Triangulation and Maximal Cliques," *Pattern Recognition,* Vol. 19, 1986, pp. 35–40.

Ohta, Y., and T. Kanade, "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-7, 1985, pp. 139–154.

Ohta, Y., K. Takano, and K. Ikeda, "A Highspeed Matching System Based on Dynamic Programming," *Proceedings of the First International Conference on Computer Vision,* London, 1987, pp. 335–342.

Paderes, F. C., E. M. Mikhail, and W. Förstner, "Rectification of Single and Multiple Frames of Satellite Scanner Imagery Using Points and Edges as Control," *Proceedings of the NASA Symposium on Mathematical Pattern Recognition and Image Analysis,* Houston, 1984.

Pagano, M., "Estimation of Models of Autoregressive Signal Plus White Noise," *Annals of Statistics,* Vol. 2. 1984, pp. 99-108.

Panton, D. J., "A Flexible Approach to Digital Stereo Mapping," *Proceedings of the DTM Symposium, American Society of Photogrammetry,* St. Louis, 1978, pp. 32–60.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, New York, 1965; 2d ed., 1984.

Park, S. K., and R. A. Schowengerdt, "Image Reconstruction by Parametric Cubic Convolution," *Computer Vision, Graphics, and Image Processing,* Vol. 23, 1983, pp. 258–272.

Pentland, A. P. (ed.), *From Pixels to Predicates,* Ablex, Norwood, NJ, 1986.

Pereira, J. A. G., and N. D. A. Mascarenhar, "Digital Image Registration by Sequential Analysis," *Computers and Graphics,* Vol. 8, 1984, pp. 247–253.

Pertl, A., "Digital Image Correlation with the Analytical Plotter Planicomp C100," *International Archives of Photogrammetry and Remote Sensing,* Vol. 25-A3a, Rio de Janeiro, 1984.

Price, K. E., "Relaxation Matching Techniques—A Comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-7, 1985, pp. 617–623.

Rice, T. A., and L. H. Jamieson, "Scaling and Rotational Registration," in *Computing Structures and Image Processing,* M. J. B Duff et al. (eds.), Academic Press, San Diego, 1985.

Rosenfeld, A., (ed.), *Image Modeling,* Academic Press, New York, 1981.

Ryan, T. W., R. T. Gray, and B. R. Hunt, "Prediction of Correlation Errors in Stereo Pair Images," *Optical Engineering,* Vol. 19, 1980.

Sadjadi, F. A., "Performance Evaluation of Correlations of Digital Images Using Different Separability Measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-4, 1982, pp. 436–441.

Schachter, B., "Long Crested Wave Models," *Image Modeling*, Rosenfeld (ed.), Academic Press, New York, 1981, pp. 327–342.

Schalroff, R. J., and E.S. McVey, "Algorithms Development for Real Time Automatic Video Tracking Systems," *Proceedings of the Third International Computers and Applications Conference,* Chicago, 1979, pp. 504–511.

Schewe, H., "Automatische photogrammetrische Erfassung von Industrie oberflächen," *IDENT/Vision,* Eindelfingen, 1989.

Schewe, H., and W. Förstner, "The Program PALM for the Automatic Line and Surface Measurement Using Image Matching Techniques," *International Archives of Photogrammetry and Remote Sensing,* Vol. 26-3/2, Rovaniemi, 1986, pp. 608–622.

Schulte, S., "Modellierung von Beobachtungsreihen durch ein erweitertes Autoregressives Modell," *Deutsche Geodätische Kommision,* München C 327, Munich, 1987.

Shapiro, L. G., and R. M. Haralick, "A Metric for Comparing Relational Descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-7, 1985, pp. 90-94.

Sharp, J. V., R. L. Christensen, and W. L. Gilman, "Automatic Map Compilation Using Digital Techniques," *Photogrammetric Engineering,* Vol. 31, 1965, pp. 223–239.

Smith, G. B., and H. C. Wolf, "Image-to-Image Correspondence: Linear-Structure Matching," *Proceedings of the NASA Symposium on Mathematical Pattern Recognition and Image Analysis,* Houston, 1984.

Spiegelhalter, D. J., "A Statistical View of Uncertainty in Expert Systems," in *Artificial Intelligence and Statistics,* W. A. Gale (ed.), Addison-Wesley, Reading, MA, 1985.

Steiner, D., "Digital Geometric Picture Correction Using a Piecewise Zero-Order Transformation," *Remote Sensing of Environment,* Vol. 3, 1974, pp. 261–283.

Stockman, G. C., "Object Recognition and Localization via Pose Clustering," *Computer Vision, Graphics, and Image Processing,* Vol. 40, 1987, pp. 361–387.

Stockman, G. C., S. Kopstein, and S. Bennett, "Matching Images to Models for Image Registration and Object Location via Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-4, 1982, pp. 229–241.

Svedlov, M., C. McGillem, and P. Anuta, "Analytical and Experimental Design and Analysis of an Optimal Processor for Image Registration," LARS Inf. Note 090776, Purdue University, West Lafayette, IN, 1976.

Terzopoulos, D., "Image Analysis Using Multigrid Relaxation Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-8, 1986a, pp. 129–139.

——— "Regularization of Inverse Visual Problems Involving Discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-8, No. 2, 1986, pp. 129-139.

——— "The Computation of Visible-Surface Representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-10, 1988, pp. 417–438.

Thorpe, C. E., *An Analysis of Interest Operators for FiDO,* Technical Report CMU-Ri-TR-83-19, Carnegie-Mellon University, Pittsburgh, 1983.

Thurgood, J. D., and E. M. Mikhail, "Subpixel Mensuration of Photogrammetric Targets in Digital Images," School of Civil Engineering, Purdue, University, West Lafayette, IN, CH-PH-82-2, 1982.

Tian, Q., and M. N. Huhns, "Algorithms for Subpixel Registration," *Computer Vision, Graphics, and Image Processing,* Vol. 35, 1986, pp. 220–233.

Venot, A., J. F. Lebruchec, and J. C. Roucayrol, "A New Class of Similarity Measures for Robust Image Registration," *Computer Vision, Graphics, and Image Processing,* Vol. 28, 1984, pp. 176–184.

Voorhees, H., and T. Poggio, "Detecting Blobs as Textons in Natural Images," *Proceedings of the Image Understanding Workshop,* Los Angeles, 1987, pp. 892–899.

Vosselman, G., "An Investigation into the Precision of a Digital Camera," Engineering Thesis, TH Delft, Department of Geodesy, Delft, 1986.

Vosselman, G., and W. Förstner, "The Precision of a Digital Camera," *International Archives of Photogrammetry and Remote Sensing,* 27-B1, Kyoto, 1988, pp. 148–157.

Wang, C. Y., et al., "Some Experiments in Relaxation Image Matching Using Corner Features," *Pattern Recognition,* Vol. 16, 1983, pp. 167–182.

Witkin, A., D. Terzopoulos, and M. Kass, "Signal Matching through Scale Space," *International Journal on Computer Vision,* Vol.1, 1987, pp. 133-144.

Wrobel, B., "A New Approach to Computer Stereo Vision and to Digital Photogrammetry," *Proceedings of the Intercommunication Conference on Fast Processing of Photographs,* Interlaken, Switzerland, 1987, pp. 231-258.

Zimmermann, G., and R. Kories, "Eine Familie von Bildmerkmalen für die Bewegungsbestimmung in Bildfolgen," *Informatik-Fachberichte,* Vol. 87, Springer, 1984, pp. 147–153.