4
29

4

Matching Techniques

Felicitas Lang, Wolfgang Förstner
Institut für Photogrammetrie
Universität Bonn
Nußalle 15
D – 53115 Bonn

February, 1995

Contents

1 Motivation

One of the central tasks in Photogrammetry and Computer Vision is the localization and reconstruction of objects in the scene. *Localization* aims at determining the *pose*, i. e. the position and the orientation of an object. It assumes the form of the object to be known or at least to be known up to some structural or numerical parameters and the mutual relation between the reference frames, i. e. coordinate systems of the object and the cameras to be determined. *Reconstruction*, on the other hand aims at determining the *form* possibly also the structure of the object. The form description need not, but may be related to a e. g. object centred, reference coordinate system. In all cases the central tasks is to match the description of one or several images to the description of the images or objects, i.e. to establish correspondence.

1.1 Notions

In all cases automating localization and reconstruction requires to establish the *correspondence* or *match* between several images or between one or several images and a model. We therefore may distinguish several cases:

1. Image Matching

   aims at establishing the relation between two images. Examples are the classical point transfer used in relative orientation (cf. SCHICKLER chap. 5) or - applied to several image points - in image sequences or bundle blocks. Image matching also is applied when features, others than points, e.g. lines or segments are transferred from one image to an other. When transferred manually these usually are features with a label or an interpretation attached to it, as the gable of a roof (line) or a group of trees (segment). It is essential that *no explicit* reference to a scene model is required but only the appearance of the scene is needed for establishing correspondence.

2. Object Localization

   aims at establishing the relation between an image or a set of images and a given object. Examples are the localization of fiducial marks, of numbers in a map, of targeted points or natural points in an aerial image that means localization of objects for which an explicit model is available such as a canal cover or a house roof (cf. SCHICKLER chap. 5) . The location may be the position in the image (translation), possibly rotation and scale or the pose in 3D (translation and rotation). There are close relations between image matching and object localization, especially if the expected *appearance* of the object, i. e. the projection of the object is used for its localization within the image. It is typical for object location that only a few datum parameters are of primary interests, though some form parameters may have to be determined in order to establish a good match between the object and the image(s).

3. Object Reconstruction

certainly is the most demanding task in this context as a relation between the image(s) and a parametric or generic (structural) model of the object has to be established. Examples are the mensuration of Digital Elevation Models (cf. KRZYSTEK chap. 7) including the detection and localization of breaklines or the extraction of man made objects such as buildings, bridges or roads, which in a first instance may be represented by polyhedra and conics (cf. BRAUN chap. 10) . The reconstruction may include the recovery of physical properties such as the albedo or parameters of a model for the reflectance function. In all cases it is essential that rich geometric, physical, possibly structural information is extracted, however without explicit reference to the meaning of the parts of the objects, which would lead to an image interpretation. This would require a much deeper modelling and up to now only is realised in prototypes.

1.2 Formalisation

We may formalise the different tasks in order to provide the basis for the more technical description of their solutions.

Fig. 1 shows the standard set up: The surface O of the object is projected into the images $I^1, I^2, \ldots$. These projections formally are mappings

$$T_o^k : O \rightarrow I^k. \tag{1}$$

Object reconstruction is the process of inverting this mapping process:

$$(I^1, I^2, ...) \rightarrow \hat{O} \tag{2}$$

which of course can only yield an estimate $\hat{O}$ for the objects surface, however including its geometry and its reflectance properties. Due to the projections T_o^k, the large number of unknown parameters and of course due to the imperfect measurements the problem is underconstrained.

In case of two images we obtain the diagram of Fig. 2 which suggests to directly relate the images I^1 and I^2 by a mapping $T_1^2 : I^1 \rightarrow I^2$.

This may be derived from

$$T_1^2 = T_o^2 \cdot T_1^o = T_o^2 \cdot (T_o^1)^{-1} \tag{3}$$

in case the projection T_o^1 can be inverted.

No explicit reference to the object O is visible any more. However it is hidden in the usually much higher complexity of T_1^2, when compared with T_o^k:

E. g. if T_o^k are the usual perspective relations and if the objects surface is nearly flat and diffuse, then T_1^2 may be represented by a smooth vector field which in the normal case of a stereo image pair even only consist of the locally varying x-parallaxes. In special cases, e. g. locally flat terrain, the parallax field, moreover, may very well be approximated by a constant or at least a linear function which is the justification for using the simple approaches of image matching.

However, slight height variations may lead to (local) occlusions making the parallax field T_1^2 discontinuous, thus much more complicated than the surface itself (cf. Fig. 3)

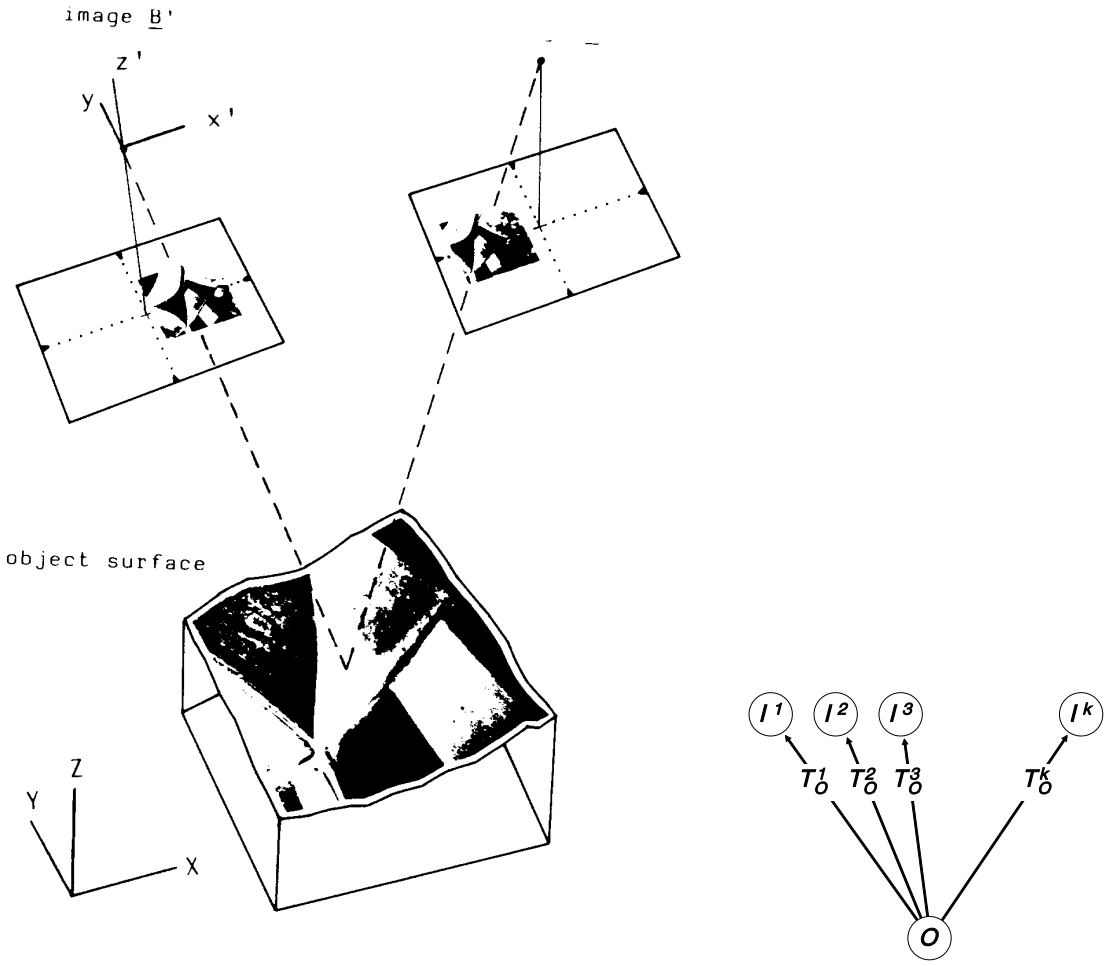


Figure 1: illustrates the setup if image matching or surface reconstruction for two perspective images. The reconstruction of the surface is either possible by a direct solution of the parameters describing the surface's geometry and reflectance properties, or by forward intersection after having established correspondencies between homologeous image features and interpolation (taken from Wrobel 1987). This would require a much deeper modelling and up to now is only realised in prototypes.

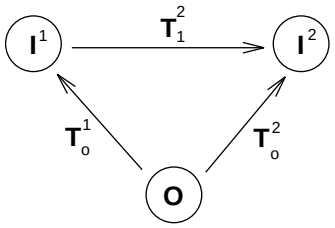


Figure 2: illustrates the general setup for object reconstruction and image matching. The object is mapped to I^1 and I^2 via the transformations T_o^1 and T_o^2 resp. In case enough images I^k are available the reconstruction of the whole imaging process is possible in principle. In case the reflectance properties of the surface O can be derived from one image, the compound transformation T_1^2 only contains geometric parameters which may be derived using image matching techniques.

In both cases – reconstruction and image matching – the estimation may be based on higher level descriptions D^k of the images I^k or the surface.

Thus the general *object reconstruction* procedure formally looks like

$$(I^1, I^2, ...) \rightarrow (D^1, D^2, ...) \rightarrow \hat{O} \tag{4}$$

whereas the *image matching* procedures may look like:

$$(I^1, I^2) \rightarrow (D^1, D^2) \rightarrow T_1^2. \tag{5}$$

As the second image may be replaced by an object model O, *object localization* formally reads as

$$(I^k, O) \rightarrow (D^k, D^o) \rightarrow T_o^k. \tag{6}$$

In case the descriptions D are iconic, i. e. in raster format $D \equiv I$, the reconstruction and the matching algorithms are termed *area based*, in case D is composed of lists of features possibly including their relations, the algorithms are *feature based.*

The goal of image matching may be described in two ways:

1. Establish the correspondence between the primitives (features) P_1 and P_2 of the two descriptions $D_1 = (P_1, R_1)$ and $D_2 = (P_2, R_2)$ of the two images I_1 and I_2. The relations R_1 and R_2 may be used to restrict the correspondence or may not be available.

2. Establish the geometric transformation T_1^2 between the two images.

The solutions of 1. and 2. are not completely equivalent: The solution to 1. not automatically implies the mapping T_1^2 as an interpolation scheme may not be defined. But the solution to 1. may be derived from the solution of 2. in case the mapping T_1^2 is one-to-one and known for the complete image area.

In the following we first want to discuss area based matching, as here the mapping function at least for a small window is determined, thus following the second type of approach. Feature based methods follow both approaches.

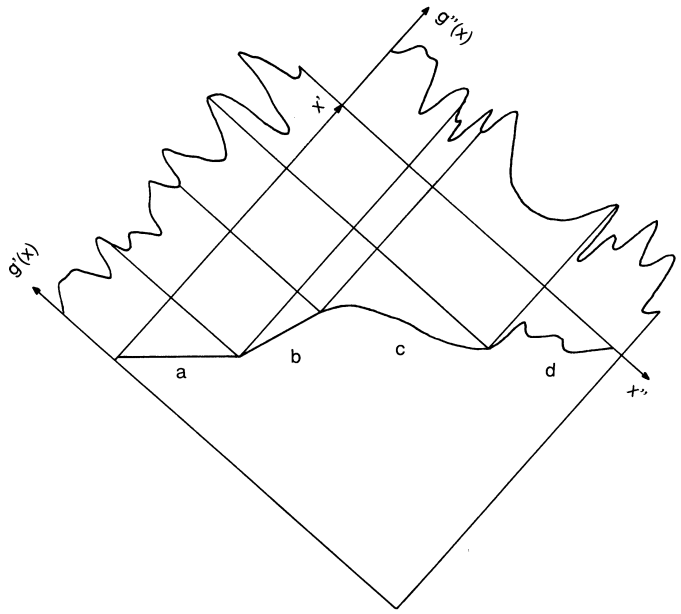


Figure 3: shows a one-dimensional profile observed from two line cameras using orthogonal projection: a) horizontal, b) sloped, c) smooth and d) piecewise smooth with occlusions.

1.3 Problem areas

There are a number of subproblems to be solved in image matching. Up to now they have no commonly accepted solution which is general enough. But for several well defined subtasks solution strategies are known.

1. Similarity Measure

   The similarity of features within the descriptions has to be measured. The similarity measure has to take into account the uncertainty of the features, the projection T_o^k and the type of the object. There exist proposals for measuring the similarity of intensities (cf. Hannah 1974, Witkin *et al.* 1987, Huang 1981), points (cf. Nagel and Enkelmann 1981, Förstner 1986), straight line segments (cf. Baker and Binford 1982, Benard 1983, Grimson 1981, Ohta and Kanade 1987), and relational descriptions (cf. Boyer and Kak 1988, Price 1985, Vosselman 1992). In all cases they are embedded in a statistical framework. With the exception of relational descriptions, simple means for deriving the influence of the sensor noise on the similarity measure and for evaluating the sensitivity with respect to model errors are available.

2. Regularization

   All matching problems are ill-posed, that is they are underconstrained, have no unique solution or have an instable solution. Therefore they require regulariza-

tion (cf. Terzopoulos 1986b). E. g. least squares estimation is a regularization method for in general underconstrained problems which have no unique solution. In image matching the object model may be useful for regularization (cf. Rosenholm 1987), e. g. when requiring the resultant parallax field to be smooth due to the assumption of a smooth object surface. Other regularization techniques are parametric or structural constraints as the assumption of the object to be planar or polyhedral. The problem with regularization, as with all types of imposed constraints, is the proper link with the similarity measure. Generally speaking, the matching may be treated as an optimization problem minimizing a cost function:

$$C = \alpha_d C_d + \alpha_m C_m \tag{7}$$

where the total costs C are a weighted sum of the costs C_d due to dissimilarity and C_m due to mismatch with the model structure. The relation α_s/α_m determines whether the solution is closer to the data or closer to the model. It may be determined by introducing α_s/α_m as additional unknown, which under certain conditions, is equivalent to variance component estimation (cf. Förstner 1991, Weidner 1994).

3. Algorithmic Complexity

   In most matching and reconstruction problems the number of unknowns and therefore the search space is huge. Only under very restricted conditions the problem can be formulated as an estimation task with a few parameters. Generally the number of unknown parameters is in the order of $10^4 - 10^7$. In case of discrete parameters, e. g. when matching features, search methods have to be applied with search spaces which generally are at least in the order of again $10^4 - 10^7$ but may easily grow as fast as $n!$ with n being the number of features.

   Thus in all cases the algorithmic complexity has to be taken care of. Any type of a priori knowledge therefore may be included into the algorithmic solution. Strong constraints, e. g. like the epipolar geometry (cf. Kreiling 1976) of course are of highest value. But also weak constraints as approximate values, average depth, conditional probabilities (close to a house should be a road) etc. may significantly reduce the expected computing times. In case these constraints can be evaluated they may be at the same time used for regularization .

4. Approximate Values.

   The optimization function C from eq. 7 is highly irregular with many local optima. This requires specific strategies for finding the global optimum in order to cope with far off approximate values. Scale Space techniques (cf. Witkin *et al.* 1987), e.g using image pyramids are powerful for solving this problem: They follow a coarse-to-fine strategy, also known from the densification of geodetic networks. In our context the strategies which follow this paradigm are: multigrid methods (cf. Terzopoulos 1986a), graduated convexity algorithms (cf. Blake and Zisserman 1987) and the use of image pyramids (cf. Ackermann and Hahn 1991), which is the most transparent one.

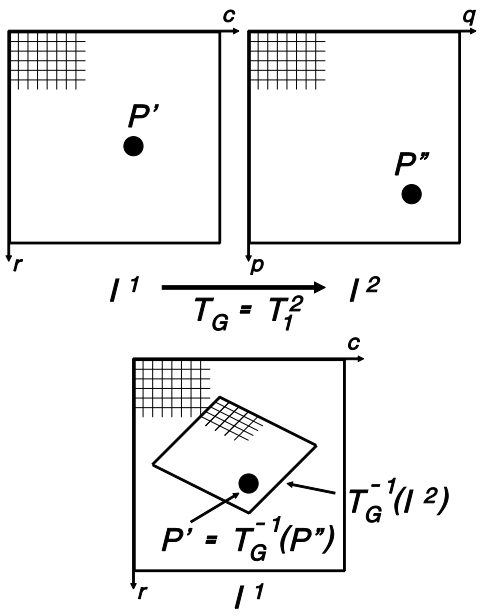


Figure 4: shows the principle of image matching: The image I_2 is warped such that the intensities best fit to I_1. A point P^1 may then be transferred to P^2 using the estimated geometric transformation T_1^2.

We will touch all these aspects in the following sections. Section 2 discusses approaches for area based image matching, especially cross correlation (sect. 2.1) and least squares matching (sect. 2.2). Some feature based matching approaches are presented in section 3, namely pose clustering (sect. 3.1), string matching (sect. 3.3), and relational matching (sect. 3.4).

2 Area Based Matching

Area based image matching starts from two descriptions D_1 and D_2 which are identical to the images or which are pixelwise (pointwise) functions $\overline{I_1} = f(I_1)$ and $\overline{I_2} = f(I_2)$ of the images, thus $\overline{I_1}$ and $\overline{I_2}$ are images themselves. In both cases the images may be vector valued, e. g. color images. In the following we will restrict the discussion to scalar valued intensity images.

The idea of area based matching is to shift and possibly warp one of the images such that its intensities best fit to the intensities of the other image:

$$\max_T(similarity(I^1, T(I^2))) \rightarrow \hat{T} \tag{8}$$

$$\min_T(distance(I^1, T(I^2))) \rightarrow \hat{T} \tag{9}$$

The best fit can be achieved using either a similarity or a distance measure.

In case of only one channel $g_k(z_i)\,(k = 1, 2,\ z_i = (r_i, c_i))$ per image k, we may write

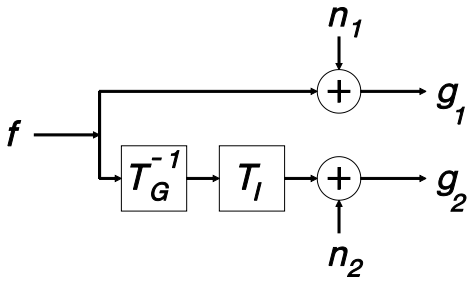


Figure 5: shows the assumed generation process: The original image f is observed leading to a noisy version g_1 on one hand and is warped, pointwise transformed and observed independently leading to g_2 on the other hand.

the following model ($g' \doteq g_1, g'' \doteq g_2$, etc.):

$$g'(z_i') = f(z_i') + n'(z_i') \tag{10}$$
$$g''(z_i'') = h(f(z_i'; p_G); p_I) + n''(z_i'') \tag{11}$$

with the parametric geometric transformation

$$T_G = f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \tag{12}$$

depending on the geometric parameters p_G and the intensity transformation

$$T_I = h : \mathbb{R} \rightarrow \mathbb{R} \tag{13}$$

depending on the radiometric parameters p_I.

The model contains the following components (cf. fig. 5):

$f(z_i')$	=	$f(r_i, c_i)$	the ideal image
$g'(z_i')$	=	$g_1(r_i, c_i)$	the first image
$n'(z_i')$	=	$n_1(r_i, c_i)$	its noise
$g''(z_i'')$	=	$g_1(p_i, q_i)$	the second image
$n''(z_i'')$	=	$n_1(p_i, q_i)$	its noise
T_G	=	f	a geometric transformation from the (r, c)-coordinate system of G_1 to the (p, q)-coordinate system of g_2
p_G			the parameters of T_G
T_I	=	h	an intensity transformation, in general it relates the values of I_1 to I_2
p_I			the paramters of T_I

There exist quite some variations how to setup and handle this problem, depending on T, on the chosen similarity measure and on the algorithmic solution:

1. The mapping function T_G

   implicitely refers to the object model, specifically the type of the surface. Examples for the geometric transformation and its implicit assumption about the surface, related to the normal setup of a stereo model are (cf. Fig. 4)

- translation (horizontal plane)
- translation and rotation
- translation, rotation and scale
- affinity (locally planar surface)
- projectivity (globally planar)
- smooth (smooth surface without occlusions)
- piecewise smooth (piecewise smooth, possibly with occlusions)

Translation is used for cross correlation e. g. for point transfer on surfaces which are flat enough, affinity is used for least squares matching, e. g. for point transfer on locally smooth surfaces and - the most general model - piecewise smooth T_G is used in the "multipoint matching" procedure of Rosenholm 1987 for matching SPOT images, where T_G is parametrically represented by finite elements.

2. The similarity measure

   has to measure how well the two images match with each other. This usually is based on the normalized cross correlation or on a function of the intensity differences of the values themselves, actually measuring the distance between the images. The following measures are frequently used

   - sum of products (mixed 2nd moments)
   - sum of products, reduces by the mean (covariance)
   - sum of squared differences (least squares, L_2-norm)
   - sum of absolute differences (L_1-norm)
   - normalized cross correlation

   All can be shown to be the optimum for a different given radiometric and noise model.

3. The algorithmic solutions

   - complete search
   - sequential search
   - heuristic search
   - iterative least squares
   - simplex algorithms
   - dynamic programming

   highly depend on the mapping function and the similarity measure and reflect the a priori knowledge.

The lists are by no way complete. Many, but not all combinations are meaningful - reveiling the wide field to be covered.

We will representatively discuss some of these techniques in the further sections.

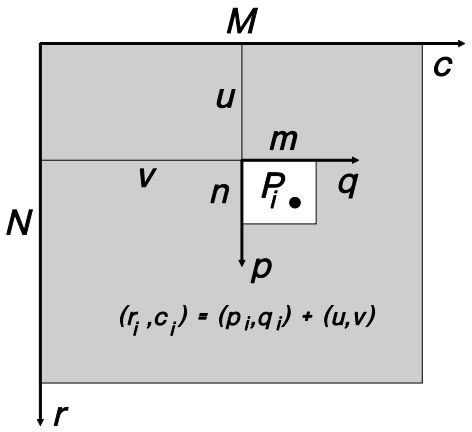


Figure 6: shows the principle of cross correlation.

2.1 Cross Correlation

Cross Correlation is a powerful technique to get the correspondence between digital images (cf. Jähne 1989). It is based on two assumptions:

1. the two images geometrically differ only due to translation.
2. the two images radiometrically differ only due to brightness and contrast

Thus the geometric transformation

$$T_G : \begin{pmatrix} p \\ q \end{pmatrix}_i = \begin{pmatrix} r \\ c \end{pmatrix}_i - \begin{pmatrix} u \\ v \end{pmatrix} \tag{14}$$

only contains the two unknown shift parameters $p_G = (u, v)^T$. The radiometric transformation

$$T_i : h(f) = a + bf \tag{15}$$

is linear with the parameters $p_I = (a, b)^T$. The principle is illustrated in Fig. 6.

An initial estimate $\hat{p}_G^{(0)} = (\hat{u}, \hat{v})^{0^T}$ may be obtained from

$$\boxed{\max_{u,v} \rho_{12}(u, v) \rightarrow (\hat{u}, \hat{v})^{(0)}} \tag{16}$$

with

$$\begin{aligned}
\rho_{12}(u, v) &= \frac{\sigma_{g_1 g_2}(u, v)}{\sigma_{g_1}(u, v)\sigma_{g_2}} \\
\sigma_{g_1 g_2}(u, v) &= \frac{1}{m-1}\left[\sum_{i=1}^{m} g_1(r_i - u, c_i - v)g_2(r_i, c_i) - \frac{1}{m}\sum_{i=1}^{m} g_1(r_i - u, c_i - v)\sum_{i=1}^{m} g_2(r_i, c_i)\right], \\
\sigma_{g_1}^2(u, v) &= \frac{1}{m-1}\left[\sum_{i=1}^{m} g_1^2(r_i - u, c_i - v) - \frac{1}{m}\left(\sum_{i=1}^{m} g_1(r_i - u, c_i - v)\right)^2\right],
\end{aligned}$$

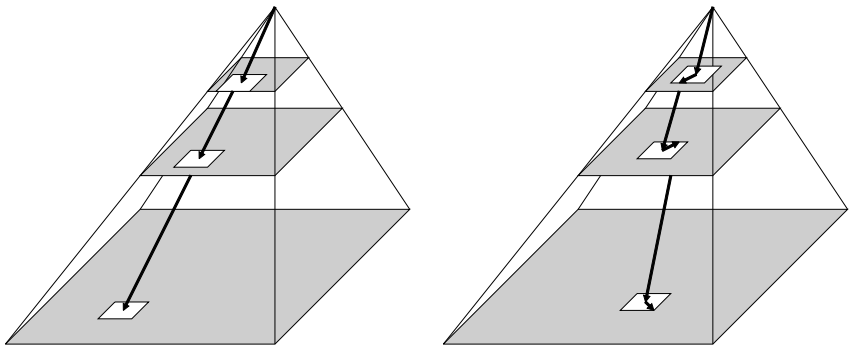


Figure 7: illustrates a hierarchical coarse-to-fine approach for image matching. The result of the match in each level of the image pyramid is used as approximate value for the match in the next lower level.

$$\sigma_{g_2}^2 = \frac{1}{m-1}\left[\sum_{i=1}^{m} g_2^2(r_i,c_i) - \frac{1}{m}\left(\sum_{i=1}^{m} g_2(r_i,c_i)\right)^2\right] \tag{17}$$

The estimate is independent on differences in brightness and contrast due to the normalization with respect to mean and standard deviation. The estimation process usually exists of a spiral type of search starting from an approximation value for the shift, stopping in case a relative maximum of ρ_{12} is found.

In case the approximation values are very coarse, a complete search in a search window may be performed. However it is better to perform the correlation using a coarse to fine strategy. This is based on an image pyramid (cf. WEIDNER chap. 2) of the two images I^1 and I^2 (cf. Fig. 7). The correlation starts at the top of the pyramid which may consist of two small windows of 10×10 to 40×40 pixels, say. A window in the left image is selected and searched for in the right image. The estimated positions are transferred into the next lower level. At this level the transferred window in the left image is then searched for in the right image. This procedure is repeated until the lowest level of the pyramid is reached. In all cases only a small search area is necessary, making this coarse to fine strategy very fast (cf. Hannah 1974).

The estimate $(\hat{u},\hat{v})^{(o)}$ from eq. (17) is integer valued. This is sufficient for the intermediate steps in the hierarchy and in case no high precision is aimed at. Usually the precision then is governed by the rounding error being approximately $1/3[pel]$.

A subpixel estimate can be obtained by approximating the surface of the two dimensional correlation function $\rho_{12}(u,v)$ in the vicinity of $(\hat{u},\hat{v})^{(o)}$ by a second order polynomial and determining its local maximum.

This leads to the final estimate

$$(\hat{u},\hat{v})^T = (\hat{u},\hat{v})^{(0)T} - \left[H\rho\mid_{(\hat{u},\hat{r})^{(0)}}\right]^{-1}.\nabla\rho\mid_{(\hat{u},\hat{r})^{(0)}} \tag{18}$$

with

$$H\rho\,|_{(\hat{u},\hat{v})^{(0)}} = \begin{pmatrix} \rho_{rr} & \rho_{rc} \\ \rho_{cr} & \rho_{cc} \end{pmatrix} |_{(\hat{u},\hat{v})^{(0)}} \tag{19}$$

$$\nabla\rho\,|_{(\hat{u},\hat{v})^{(0)}} = \begin{pmatrix} \rho_r \\ \rho_c \end{pmatrix} |_{(\hat{u},\hat{v})^{(0)}} \tag{20}$$

and e. g.

$$\rho_r = \frac{\delta}{\delta r}\rho = \frac{1}{8}\begin{pmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & 1 \end{pmatrix} * \rho \tag{21}$$

$$\rho_{rr} = \frac{\delta^2}{\delta r^2}\rho = \frac{1}{4}\begin{pmatrix} -1 & 2 & -1 \\ -2 & 4 & -2 \\ -1 & 2 & -1 \end{pmatrix} * \rho \tag{22}$$

...

(cf. WEIDNER chap. 2) .

The precision of this estimate is given by its covariance matrix

$$D\begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} = \frac{1}{m}\cdot\frac{1-\rho_{12}}{\rho_{12}}\left[H\rho\,|_{(\hat{u},\hat{v})^{(0)}}\right]^{-1}\cdot\Delta x^2 \tag{23}$$

with

m	the number of pixels involved in the correlation
ρ_{12}	the similarity of the two windows at $(\hat{u},\hat{v})^{(0)}$
$H\rho$	the roughness of the texture of the signal, a high value of $H\rho$ leads to a small covariance matrix, in case the texture within the window shows an orientation this will reflect a higher uncertainty of the estimate in that direction.
Δx	the pixel size, assumed to be identical in row and column direction.

Eq. (23) reveils that merely evaluating the correlation coefficient is not sufficient.

Also the noise variance σ_n^2 of the two images can be estimated in case $\sigma_{n_1} \approx \sigma_{n_2}$ and the contrast of both images does not differ too much ($a \approx 1$) :

$$\hat{\sigma}_n^2 = \sigma_g^2(1-\rho_{12}) \tag{24}$$

where the mean variation of g is

$$\sigma_g^2 = \frac{1}{2}(\sigma_{g_1}^2 + \sigma_{g_2}^2) \tag{25}$$

Discussion:

1. Cross Correlation can be shown to be optimal for estimating a shift between two signals provided that the noise is white, i. e. the intensities are independent random variables with the same standard deviation. Otherwise a withening process has to be performed which, in case of images, where the intensity is proportional to the number of received photons, leads to a squareroot point function.

2. If the signal is homogeneous, i.e. with constant variance $\sigma_{g_1}(u) = \sigma_{g_1}$, normalization with $\sigma_{g_1}, \sigma_{g_2}$ can be omitted, reducing the estimate for maximizing the covariance $\sigma_{g_1 g_2}(u, v)$.

3. If, in addition to the variance being constant, the mean of the signal can be assumed to be zero, only the sum of the products $\sum g_1(r_i - u, c_i - v) g_2(r_i, c_i)$ of g_1 and g_2 within the window has to be maximized. Both normalizations can be achieved by replacing g by $(g - \mu)/\sigma_g$ where μ and σ_g are local means and standard deviations. Then the products can be taken from a look-up table increasing the efficiency of the procedure.

4. Hardware chips for determining the cross correlation of $64 \times 64\,[pel]$ windows in a few milliseconds are available. Cross correlation also can be realized optically thus allowing real time applications on large images.

5. The equation for one dimensioned cross correlation can be derived by replacing (r, c) by x, (p, q) by y, (u, v) by u leading to

$$\begin{aligned} \hat{u} &= \hat{u}^{(0)} - \frac{\rho_+ - \rho_-}{2(\rho_+ + 2\rho_0 - \rho_-)} \cdot \Delta x \\ \sigma_{\hat{u}}^2 &= \frac{1}{m} \cdot \frac{1 - \rho_{12}}{\rho_{12}} \cdot \frac{\Delta x^2}{-\rho_+ + 2\rho_0 - \rho_-} \end{aligned} \tag{26}$$

   with $\rho_0 = max_u \rho_{12}(u)$ to integer position and ρ_- and ρ_+ being the correlation coefficients at $\hat{u}^{(0)} - \Delta x$ and $\hat{u}^{(0)} + \Delta x$ resp.

6. The applications of cross correlation are manifold:

   - point transfer in aerial triangulation and registration of satellite images (cf. Tsingas 1992, ACKERMANN chap. 6)
   - locating fiducials or targeted control or tie points (SCHICKLER chap. 5, FUCHS chap. 3)
   - mensuration of digital elevation models (cf. KRZYSTEK chap. 7) .
   - recognition of standardized letters or numbers (e. g. of aerial images)
   - tracking objects in image sequences (cf. Hahn 1994)

7. Cross correlation has a number of disadvantages

   - The precision decreases rapidely if the geometric model is violated. Rotations greater 20^o and scale differences greater 30 % should be avoided (cf. Förstner 1984).
   - The extension of the search space when determining other parameters (e. g. rotation or scale) is prohibitive. Therefore no real extension of the method is vailable.
   - Cross correlation cannot handle occlusions in a straight forward manner.

The least squares approach discussed in the next section aims at reducing some of these restrictions.

2.2 Least Squares Matching

Least squares matching (cf. Förstner 1982, Ackermann 1984) is a generalization of cross correlation. It has the following crucial features:

- any parametric type of mapping function can be assumed,
- any parametric type of radiometric relation between the two images may be dealt with,
- the estimation process is efficient,
- the evaluation tools from least squares estimation are available.

We first want to explain the principle of least squares matching of two one-dimensional signals, where one is given, taking the part of the object, the other one being a shifted and noisy version of the object. Then we will discuss the case of linear T_G and T_I.

2.2.1 Principle of Least Squares Matching

The model in the one-dimensional situation with the only unknown parameter, the shift, is given by:

$$y = x - u \tag{27}$$

$$g(x_i) = f(y_i) + n(x_i), \qquad i = 1, ..., m \tag{28}$$

The observed discrete signal $g(x_i)$ is related to the given function $f(y)$ by an unknown shift u. The noise $n(x_i)$ is assumed to be white, i. e. uncorrelated and with variance σ_n^2. The number of observations is denoted by m.

With an approximate value u_0 we may write

$$u = u_0 + du \tag{29}$$

where du is the unknown correction to the approximate value. Thus the nonlinear model can be rewritten as

$$g_i = f(x_i - u) + n_i \tag{30}$$

$$= f(x_i - u_0 - du) + n_i \tag{31}$$

with the abbreviations $g_i = g(x_i)$ and $n_i = n(x_i)$. We now linearize around the point $x_i - u_0$ to obtain

$$g_i = f(x_i - u_0) - f'(x_i - u_0)du + \frac{1}{2}f''(x_i - u_0 - \zeta du)du^2 + n_i \tag{32}$$

for some $\zeta \in [0, 1]$ and where $f'(y) = df/dy$ and $f''(y) = d^2f/dy^2$. In the following we assume f' not to vanish and the second order terms to be negligible.

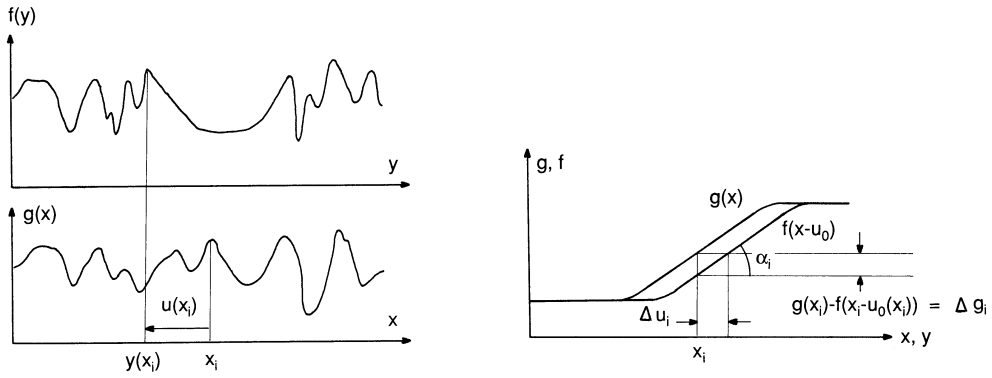


Figure 8: illustrates the principle of the differential least squares approach to estimate the local shift $du(x_i)$ (left). The functions g and f (right) are assumed to be locally approximable by a linear function. Then one is able to derive the shift $du_i = -\Delta g_i / f_i'$ from the difference Δg_i of f and g at the corresponding positions x_i and $x_i - u_0$ which depends on the approximate shift u_0 and on the slope $f_i' = \tan\alpha$.

With the differences

$$\Delta g_i = \Delta g(x_i) = g(x_i) - f(x_i - u_0) \tag{33}$$

and the derivative

$$f_i' = f'(x_i - u_0) \tag{34}$$

we then obtain the *linearized model*

$$\Delta g_i = -f_i' \, du + n_i, \qquad i = 1, ..., m \tag{35}$$

or explicitely

$$g(x_i) - f(x_i - u_0) = -\frac{df(y)}{dy}|_{y=x_i-u_0} du + n(x_i) \qquad i = 1, ..., m \tag{36}$$

We first give the solution for a single observed $g(x_i)$, thus for $m = 1$

$$\widehat{du_i} = -\frac{\Delta g_i}{f_i'}|_{y=x_i-u_0} \tag{37}$$

thus

$$\widehat{u_i} = u_0 + \widehat{du_i} \tag{38}$$

This estimate thus is based on the assumption, that the local intensity change is only caused by the unknown shift (cf. Fig. 8)

We obtain for its standard deviation

$$\sigma_{\widehat{du_i}} = \frac{\sigma_n}{|f_i'|} \tag{39}$$

Obviously the precision depends on the local slope f_i' and is high at edges with high slope, which is to be expected. No estimate for u is available if $f_i' = 0$. In case one choses

$$\sigma_0 = \sigma_n \tag{40}$$

the weight of this estimate is

$$w_i = f_i'^2 \tag{41}$$

again intuitively going with the square of the slope of f.

There is no redundancy in this estimate making the estimate unreliable.

Using m observations leads to the estimate

$$\widehat{du} = -\frac{\sum_{i=1}^m f_i' \Delta g_i}{\sum_{i=1}^m f_i'^2} = -\frac{\sum_{i=1}^m (g(x_i) - f(x_i - u_0)) f'(x_i)}{\sum_{i=1}^m (f'(x_i - u_0))^2} \tag{42}$$

This is equivalent with the weighted correction of the shift

$$\widehat{du} = \frac{\sum_{i=1}^m w_i \widehat{du_i}}{\sum_{i=1}^m w_i} \tag{43}$$

Obviously observations with weight being 0 thus $f_i' = 0$ do not at all contribute to the estimation.

The precision of the total shift

$$\hat{u} = u_0 + \widehat{du} \tag{44}$$

is given by

$$\sigma_{\widehat{du}}^2 = \frac{\sigma_n^2}{\sum_{i=1}^m f_i'^2} \tag{45}$$

If we now define the *mean squared gradient* of f by

$$\sigma_{f'}^2 \doteq \frac{\sum_{i=1}^m f_i'^2}{m} \tag{46}$$

we can represent the standard deviation of the estimated shift by

$$\sigma_{\hat{u}} = \frac{\sigma_n}{\sqrt{m}\sigma_{f'}} \tag{47}$$

As expected it depends on the number m of observations, on the noise standard deviation σ_n and the texture or the edge buisiness measured by the squared gradient $\sigma_{f'}$.

This simplified formula may be used for predicting the precision of least squares matching.

Example: In case two profiles with noise standard deviation $\sigma_n = 2[gr]$, $m = 10$ common pixels and an average gradient $\sigma_{f_x} = 10[gr/pel]$ we can expect the precision of the shift to be $\sigma_x = \sqrt{(1/10)} \cdot 2[gr]/10[gr/pel] \approx 0.06[pel]$. Clearly subpixel accuracy is reachable.

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

$$\hat{\sigma}_n^2 = \frac{1}{m-1}\sum_{i=1}^{m}(g(x_i) - f(x_i - \hat{u})^2 \tag{48}$$

whose standard deviation is

$$\hat{\sigma}_{\hat{\sigma}} = \frac{1}{\sqrt{2(m-1)}}\hat{\sigma}_n. \tag{49}$$

in case the simple model really holds.

2.2.2 2D-Least Squares Matching with Linear Geometric and Radiometric Model

We now want to explicitely give the linearized observation equation for the case of linear T_G and T_I (cf. Ackermann 1984):

$$\begin{aligned} T_G : \begin{pmatrix} p \\ q \end{pmatrix}_i &= \begin{pmatrix} a_1 & a_2 \\ a_4 & a_5 \end{pmatrix} \begin{pmatrix} r \\ c \end{pmatrix}_i + \begin{pmatrix} a_3 \\ a_6 \end{pmatrix} \\ T_I : h(f) &= a_7 \cdot f + a_8 \end{aligned} \tag{50}$$

Starting from initial values

$$a_{0k} = 1, \qquad k = 1, 5, 7 \tag{51}$$

$$a_{0k} = 0, \qquad k = 2, 3, 4, 6, 8 \tag{52}$$

we obtain the m observation equations:

$$\begin{aligned} \Delta g_i = \; & f_{r_i} \cdot r_i \cdot \widehat{da_1} + f_{r_i} \cdot c_i \cdot \widehat{da_2} + f_{r_i} \cdot \widehat{da_3} + f_{c_i} \cdot r_i \cdot \widehat{da_4} \\ + \; & f_{c_i} \cdot c_i \cdot \widehat{da_5} + f_{c_i} \cdot \widehat{da_6} + f_i \cdot \widehat{da_7} + 1 \cdot \widehat{da_8} + \bar{n}_i \end{aligned}$$

with

$$\begin{pmatrix} p \\ q \end{pmatrix}_{0i} = \begin{pmatrix} a_{01} & a_{02} \\ a_{04} & a_{05} \end{pmatrix} \begin{pmatrix} r \\ c \end{pmatrix}_i + \begin{pmatrix} a_{03} \\ a_{06} \end{pmatrix} \tag{53}$$

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

estimates

$$f_{r_i} = \frac{\partial \hat{g}_1(r_i, c_i)}{\partial r}$$

$$f_{c_i} = \frac{\partial \hat{g}_1(r_i, c_i)}{\partial c} \tag{54}$$

$$f_i = \hat{g}_1(r_i, c_i) \tag{55}$$

where $\hat{g}_1$ is a *restored*, i. e. *smoothed* version of g_1. Of course the restored version $\hat{g}_1$ may also use $\hat{g}_2$ in order to obtain a better estimate. But then the geometric and radiometric transformations have to be taken into account.

Observe that the transferred coordinates (p_{0i}, q_{0i}) according to eq. (53) are not integer values which therefore requires resampling when calculating the expressions in eq. (54). At least linear interpolation is always necessary. This is to avoid instabilities in the iteration scheme.

The 8×8 normal equation system

$$\mathbf{N}\hat{\boldsymbol{\beta}} = \mathbf{h} \tag{56}$$

for the 8 unknown parameters $\hat{\boldsymbol{\beta}} = (\widehat{da_k})$ yield the 6 corrections $d\hat{p}_G$ for the approximate values $a_k^{(\nu-1)}$ of the geometric transformation and the corrections $\widehat{da_7}$ and $\widehat{da_8}$ for contrast and brightness. Thus

$$\hat{a}_k^{(\nu)} = \hat{a}_k^{(\nu-1)} + \widehat{da_k}^{(\nu)} \tag{57}$$

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

The estimated noise variance

$$\hat{\sigma}_{\overline{n}}^2 = \frac{1}{m-8} \sum_i \overline{n}_i^2 \tag{58}$$

measures the average noise difference between g_1 and g_2 which in case $a_7 \approx 1$ and $\sigma_{g_1} \approx \sigma_{g_2}$ leads to the estimate

$$\hat{\sigma}_n = \hat{\sigma}_{\overline{n}}/\sqrt{2} \tag{59}$$

for the image noise, which is assumed to be constant, i. e. signal independent ($w_i = const.$).

The standard deviations of the unknowns then can be derived from

$$\hat{\sigma}_{\hat{a}_k} = \hat{\sigma}_{\overline{n}} \sqrt{(\mathbf{N}^{-1})_{kk}} \tag{60}$$

In case a point (r_0, p_0) is transferred to (p_0, q_0) via the estimated geometric transfromation T_G its 2×2 covariance matrix may be used for an evaluation. This has to be derived by error propagation.

Fig. 9 shows 3 small windows from which the interior 16×16 pixels are used to evaluate the expected precision of a match. Assuming a noise standard deviation of $\sigma_n = 5[gr]$ the confidence ellipses shown below the template are obtained. With 99 % probability the true shift will lie within an area depicted by the ellipse around the estimated shift. The pixel size is $20\mu m$. The smallest confidence ellipses have major axes of below $1\mu m$ thus below 1/20 of a pixel. The largest confidence ellipse is obtained at the edge point indicating that one cannot expect a good accuracy for the position along the edge. With respect to the expected matching precision the other ellipses reflect the image content reasonably well. As the form of the ellipse essentially depends on the squared gradient, no image interpretation is performed when predicting the precision.

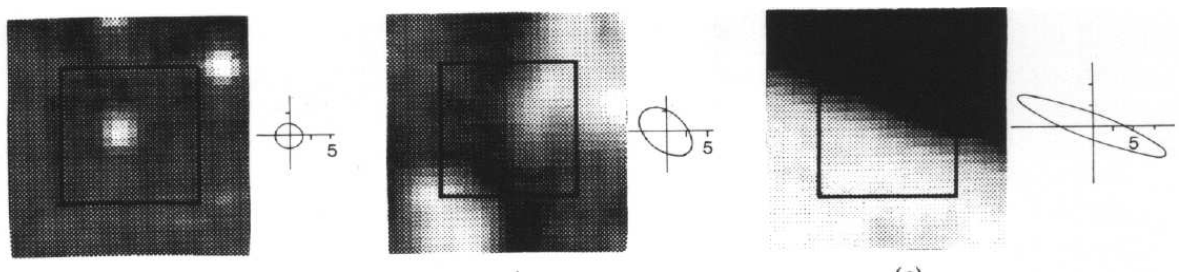


Figure 9: Precision of matching to be expected for 3 small windows. Although 32 × 32 pixels are shown, only the interior 16 × 16 pixel are assumed to be used. The noise standard deviation is assumed to be $\sigma_n = 5[gr]$. The 99 % confidence ellipses given in μm refer to a pixel size of $20\mu m$.

In the most simple case where only a shift is to be estimated the 2 × 2 normal equation system reads as

$$\begin{pmatrix} \sum_i f_{r_i}^2 & \sum_i f_{r_i} f_{c_i} \\ \sum_i f_{c_i} f_{r_i} & \sum_i f_{c_i}^2 \end{pmatrix} \cdot \begin{pmatrix} d\hat{r} \\ d\hat{c} \end{pmatrix} = \begin{pmatrix} \sum_i f_{r_i} \Delta g_i \\ \sum_i f_{c_i} \Delta g_i \end{pmatrix} \tag{61}$$

$$\mathbf{N}\hat{\boldsymbol{\beta}} = \mathbf{h} \tag{62}$$

where the sums are to be taken over all pixels in concern. Obviously the normal equation matrix $\mathbf{N}$ contains the average $\overline{\mathbf{\Gamma} f}$ of the squared gradient $\mathbf{\Gamma} f = \nabla f \nabla f^T$, with $\nabla f = (\partial f/\partial r, \partial f/\partial c)$. Thus the normal equations may be written as

$$m\,\overline{\mathbf{\Gamma} f}\,\hat{\boldsymbol{\beta}} = m\,\overline{\nabla f\,\Delta g} \tag{63}$$

where m is the number of observations.

This reveils close relations to the information preserving smoothing (cf. WEIDNER chap. 2) and to the interest operator (cf. FUCHS chap. 3) .

In order to give an insight into the structure of the precision to be expected from least squares matching we analyse the covariance matrix of the estimated shifts $(\hat{r}, \hat{c})$

$$D\begin{pmatrix} \hat{r} \\ \hat{c} \end{pmatrix} = \frac{\hat{\sigma}_n^2}{m} \left(\overline{\mathbf{\Gamma} f}\right)^{-1} \tag{64}$$

The precision depends on:

1. the number m of pixels used, i. e. the window size,
2. the noise variance σ_n and
3. the average squared gradient $\overline{\mathbf{\Gamma} f}$.

$\overline{\mathbf{\Gamma} f}$ reflects the characteristics of the local texture. Three cases are of special interest:

1. In case the window contains an irregular texture, a corner or a distinct point, $\overline{\mathbf{\Gamma} f}$ will be large with the diagonal elements being small and approximately equal. Thus the error ellipse will be small and round.

2. In case the window contains an edge or oriented texture, the elements of $\overline{\mathbf{\Gamma} f}$ will be large, but $\overline{\mathbf{\Gamma} f}$ will be nearly singular. This indicates that the estimation of the shifts is uncertain, namely in the direction of the edge or the texture. The shift across the edge, however, may be very precise.

3. In case the window contains no or only weak information, the average gradient will be small. This leads to a large error ellipse, indication no shift can be estimated.

The discussion confirms intuition. It also motivates the information preserving filter and the feature extraction procedures discussed in WEIDNER chap. 2 and FUCHS chap. 3.

Often the small windows do not contain enough detail to enable a determination of all 8 parameters. Especially the scales a_1 and a_5 and the shears a_2 and a_4 often are not estimable. Therefore *a priori* knowledge about the transformations may be introduced in a Bayesian manner by using additional observations (possibly fictitious ones)

$$da_k = \widehat{da}_k + n_{a_k} \qquad w_{a_k} = \frac{\sigma_n^2}{\sigma_{a_k}^2} \qquad k = 1, ..., 8 \tag{65}$$

with individual weights depending on the quality, specifically on the standard deviations of the corrections to the *a priori* values da_k. This leads to the modified and stabilized normal equation system (cf. eq. (62)).

$$[\mathbf{N} + Diag(w_{a_k})] \cdot \boldsymbol{\beta} = \mathbf{h} \tag{66}$$

The right hand side remains unchanged $\mathbf{h}$, because of the corrections assumed to be $da_k = 0$. The following standard deviations can be recommended:

scales, shears	$\sigma_{a_k} = 0.1 - 1$	k = 1, 2, 4, 5, 7
geometric shifts	$\sigma_{a_k} = 1 - 10\,[pel]$	k = 3, 6
radiometric shift	$\sigma_{a_k} = 10 - 100\,[gr]$	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.

Discussion:

1. In case the geometric transformation is restricted to a shift and the radiometric transformation is linear, least squares matching and cross correlation conceptually are equivalent.

2. In case the images are in normal position, i. e. the image planes and the $x'-$ and $x''-$axis are parallel to the baseline, the geometric model reduces to

$$x' = a_1 x'' + a_2 y'' + a_3 \tag{67}$$
$$y' = y'' \tag{68}$$

Thus, together with p_I, only 5 parameters have to be determined. The a_k intuitively correspond to the depth (a_3), to the slope of the surface along the baseline (a_1) and to the slope across the baseline (a_2).

3. The case, that the unknowns are restricted to a locally varying shift – thus no radiometric parameters are involved, is important in image sequence anlysis, where the time difference dt between neighbouring images is small. Then the observation equation

$$\Delta g_i = f_{r_i}\, da_3 + f_{c_i} da_6 \tag{69}$$

after division by dt can be rewritten as the socalled *optical flow equation*

$$\frac{dg}{dt} = \frac{\partial f}{\partial r}\cdot\frac{\partial r}{\partial t} + \frac{\partial f}{\partial c}\cdot\frac{\partial c}{\partial t} \tag{70}$$

or

$$\dot{g} = \nabla f \cdot \mathbf{v} \tag{71}$$

with the temporal change $\dot{g}$ of g, the virtual velocity $\mathbf{v} = (\partial r/\partial t, \partial c/\partial t)^T$ of the point in concern and the gradient ∇f of the intensity function. Observe that $(da_3, da_6) = (dr, dc)$ represents the parallax field between the neighbouring images. The optical flow equation just states, that the (temporal) change of the intensity is solely due to the optical flow, i. e. the velocity of the projection of the object feature in the image plane.

4. Eq. (69) reveils that the local shift cannot be determined from *one* intensity change alone. This is the socalled *aperture problem.* Therefore the estimation of a parallax field is underdetermined, illposed, thus requires regularization. In the approach above this is achieved by requiring the parallaxes to follow a linear transformation. In general one could require the parallax field to be locally smooth.

5. Eqs. (61) and (64) were the basis for the "interest operator" selecting distinct image "points" used in feature based matching (cf. Paderes *et al.* 1984, Förstner 1986). The variance of the predicted match leads to an optimal choice of a window for matching.

3 Feature Based Matching

Feature based matching uses symbolic descriptions of the images for establishing correspondence. It is assumed that such a symbolic description can replace the original image sufficiently well and all information which is necessary for matching is contained in the attributes of the features and possibly their relations. Using features instead of the original intensities allows to select a representation which is much more invariant with respect to distortions such as illumination, reflectance or geometry. This makes feature based algorithms, also in general, more robust than intensity based procedures. These advantages are only partly payed for by a loss in precision, much more, however, by a loss in spatial resolution as the features do not replace the complete image in all details. In many applications such as object location or surface reconstruction feature based matching algorithms have proven to be very effective which is the reason to discuss them in this chapter.

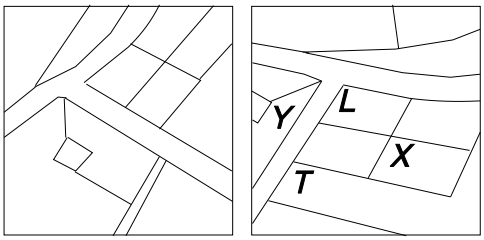


Figure 10: shows the graph of two symbolic image descriptions one derived from a satellite image the other from a map. The points are characterized as T-, L-, Y-, or X-junctions

We want to discuss 4 representative techniques in order to show the wide range of applications of feature based matching algorithms:

1. pose clustering

   is a powerful tool for for object location and recognition in case only a few parameters have to be estimated and the percentage of spurious features is high (cf. Stockman 1987, Stockman *et al.* 1982, SCHICKLER chap. 5).

2. robust estimation

   has been proved to be quite powerful in case approximate values for the transformation T_g are available (cf. Förstner 1986)

3. string matching

   is favorable in case the symbolic representation is a sorted list of attributed features. Dynamic Programming is a tool for finding optimal matches and is used quite often (cf. Baker and Binford 1982, Benard 1983, Ohta and Kanade 1987).

4. relational matching

   probably is the most general tool for finding correspondencies between descriptions as any type of attributed features and relations can be handled. This makes this procedure also useful for image interpretation tasks (cf. Boyer and Kak 1988, Price and Huertas 1992, Vosselman 1992).

3.1 Pose Clustering

Pose clustering is a powerful technique for determining the transformation between two image or object descriptions. The representation of the images or objects is assumed to consist of a list of attributed geometric primitives.

We want to describe the approach for the case of matching an image and a 2D-object (cf. Stockman 1987). The procedure is the same for matching two images or two objects.

The idea of pose clustering is to combine every primitive of the image with every primitive of the object, determine the parameters of the geometric transformation and select the parameter set where there is a cluster in parameter space.

Fig. 10 shows the sketches of a symbolic description of a satellite image and of a map. It is assumed that the image feature extraction is able to produce points which may be classified as X-, L-, Y- or T-junctions depending on the number and directions of the adjacent edges and that a similar description can be derived from the map.

We first assume image and map to be of approximately the same scale (e. g. derived from the sensor model) and only a translation to be determined. Thus the representation of the image and the map looks like:

$$D_I = P_I = \{(type, (r, c))\} \tag{72}$$

$$D_M = P_M = \{(type, (p, q))\} \tag{73}$$

The geometric model is identical to eq.(14). The similarity measure simply states two primitives (points) are similar if and only if they are of the same type. Then all pairs (P_i, P_m) with $P_i \in P_I$ and $P_m \in P_M$ and $type_i = type_j$ give an estimate

$$u_{im} = r_i - p_m \qquad v_{im} = c_i - q_m \tag{74}$$

for the unknown shift. These estimates have the following property:

1. If the match is correct then

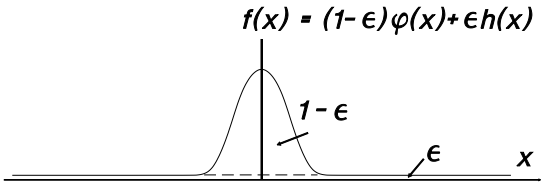
$$\begin{pmatrix} u \\ v \end{pmatrix}_{im} \sim N(\begin{pmatrix} u \\ v \end{pmatrix}, \mathbf{\Sigma}) \tag{75}$$

   where $(u, v)^T$ is the true shift and $\mathbf{\Sigma}$ the covariance matrix of the differences $(u, v)_{im}$.

2. If the match is incorrect the $(u, v)_{im}$ are equally (E) distributed in a certain range, depending on the size of the image and the map, say in a square of size L for simplicity:

$$\begin{pmatrix} u \\ v \end{pmatrix}_{im} \sim E(-L/2, L/2) \times E(-L/2, L/2) \tag{76}$$

   The density function in 1-D looks like



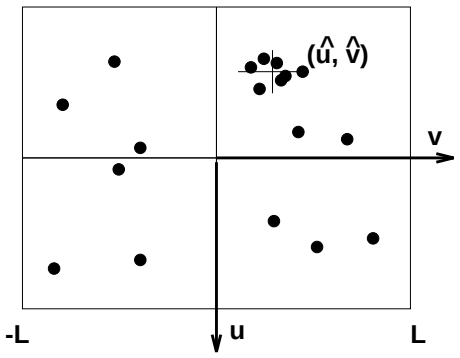


Figure 11: illustrates the result within the accumulating process.

Thus the joint distribution function of $(u,v)_{im}$ would be

$$F = (1-\varepsilon)N + \varepsilon E \tag{77}$$

where $100{\cdot}\varepsilon$ is the percentage of wrong matches, which usually is large. Then the mode of the density function can be used as an estimate for the unknown shift.

The estimation procedure therefore consists of the following steps:

1. provide an accumulator for the unkowns. In our case this is a matrix `acc` representing the discretized (u,v)-plane in the expected range. Fill the accumulator with zeros.

2. For each pair P_i, P_m which is similar determine the transformation T_i^m, here the shift $(u,v)_{im}$ and increase the corresponding element of `acc` by 1.

3. Search for the maximum of `acc` which gives the desired estimate $(\hat{u},\hat{v})$ for the unknown shift.

The result may look like in Fig. 11.

Discussion:

1. The similarity measure is decisive for the success of the procedure, as with a poor similarity measure the accumulator would show a full background making the detection of the peak very difficult.

2. Pose clustering for determining a shift is a different way to realize cross correlation.

3. In case the geometric model is violated the peak is blurred.

4. In case the discretization of the accumulator is too coarse the result is inaccurate. If the discretization is too fine the peak will not show.

5. The last two effects can be significantly reduced if instead a 1, i.e. a unit impulse, a blurred version of a unit impulse, e. g. a 3×3 or 5×5 Binomial kernel is added to the accumulator (cf. Sester and Förstner 1989).

6. A higher precision can be achieved by interpolation around the peak e. g. by a second order polynomial and determining the maximum (cf. eq. 18).

7. The procedure can be generalized to more complex transformations (cf. Stockman 1987). Then more complex primitives have to be selected in order to be able to obtain individual estimates $\hat{T}_i^m$.

   An example is the joint estimation of translation, rotation and scale. Here the primitives may be point pairs. When using the same type of points one could select pairs of different type, e. g. XT, XL, XY, TL, TY or LY, where the order of the points is important. Point pairs of this kind could be termed *abstract edges* as, when drawing them, they look like edges though they in general to not represent real image edges.

8. The algorithmic complexity of the algorithm is high. The storage and the computing are of order $O(n^{2p})$, where n is the number of the primitives (assumed to be equal in image and map) and p is the number of parameters.

The procedure can be used to advantage if:

- the number of unkowns is small
- a direct solution for T from pairs of primitives is available
- large noise, i. e. a large amount of spurious features is present and
- several solutions are possible. Then the search for the second solution can be started after the primitive pairs corresponding to the first are eliminated from the accumulator.

Pose clustering is used in the absolute orientation procedure presented by (cf. SCHICKLER chap. 5) .

3.2 Robust Estimation

The principle of pose clustering is to replace the matching problem by the estimation of the transformation parameters. In case of a large number of such parameters a different technique for parameter estimation has to be applied which however is capable of rejecting wrong matches. Robust maximum likelihood type estimators, short M-estimators have shown to be a powerful tool in this situation (cf. Paderes *et al.* 1984, Förstner 1986).

The model eq. (75) then may be replaced by a general relation

$$\mathsf{E}\left[\begin{pmatrix} u \\ v \end{pmatrix}_{im}\right] = \mathbf{g}(\hat{\boldsymbol{\beta}}) \tag{78}$$

or after linearization with the groups $\mathbf{y}_i$ of observations containing the $(u,v)_i m$'s

$$\mathsf{E}(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta} \qquad \mathbf{W}_i = \boldsymbol{\Sigma}_{y_i y_i}^{-1} \tag{79}$$

representing a Gauß-Markov-Model. Observe that correlations between the observations within each group may be be present.

In case of wrong matches eq. (75) holds.

The principle of robust M-estimation is to use a less increasing function $\rho(\mathbf{e}_i)$ in the minimization $\Omega = \sum_i \rho(\mathbf{e}_i) \rightarrow min.$. This may be realized by iteratively reweighting the observations

$$\mathbf{W}_i^{(\nu)} = \mathbf{W}_i^{(0)} \cdot f(A_i^{(\nu)}) \tag{80}$$

i.e. reducing the weights of the group if the residuals $\hat{\mathbf{e}}_i^{(\nu)}$ or a function $A^2(\hat{\mathbf{e}}_i^{(\nu)})$ is large. The reweighting function $f(x)$ is choosen to be

$$f(x) = \frac{1}{\sqrt{1+x^2}} \tag{81}$$

for the first few (3, say) iterations as correspondence is guaranteed if the model eq. (78) really is linear. Later iterations use

$$f(x) = exp(-x^2/2) \tag{82}$$

The argument $x = A(\hat{\mathbf{e}}_i^{(\nu)})$ can be choosen in various ways:

1. $A_1(\hat{\mathbf{e}}_i^{(\nu)}) = \hat{\mathbf{e}}_i^{(\nu)}$
   This is the most practical choice, with the disadvantage not taking different weights of $\mathbf{y}_i$ and the varying geometry of the mensuration design into account.

2. $A_2(\hat{\mathbf{e}}_i^{(\nu)}) = \hat{\mathbf{e}}_i^{(\nu)T}\, \mathbf{W}_i^{(0)}\, \hat{\mathbf{e}}_i^{(\nu)} / n_i \quad , \quad n_i = rank(\mathbf{W}_i^{(0)})$
   is the most simple choice, as no error propagation for the $\hat{\mathbf{e}}_i^{(\nu)}$'s is required, and in case the geometry is homogeneous, i.e. the redundancy is homogeneous distributed and leads to quite satisfactoring results.

3. $A_3(\hat{\mathbf{e}}_i^{(\nu)}) \doteq \|\hat{\mathbf{e}}_i^{(\nu)}\| / n_i = \hat{\mathbf{e}}_i^{(\nu)T}\, \mathbf{W}_{\hat{e}_i\hat{e}_i}^{(\nu)}\, \hat{\mathbf{e}}_i^{(\nu)} / n_i$, with $n_i = rank(\mathbf{W}_{\hat{e}_i\hat{e}_i}^{(\nu)})$
   only seems to be a natural choice, taking the geometry via the weights of the residuals into account. However $\mathsf{E}[A(\hat{\mathbf{e}}_i^{(\nu)})] = 1$ therefore even for large errors A is not large leading to no reduction of the weights.

4. $A_4(\hat{\mathbf{e}}_i^{(\nu)}) = \|\widehat{\Delta\mathbf{y}_i}^{[i]\nu}\| / n_i$
   seems to be the best choice. It is the estimated error $\Delta\mathbf{y}_i$ if observational group $\widehat{\Delta\mathbf{y}_i}$ is not taking part in the estimation. $\widehat{\Delta\mathbf{y}_i}^{[i]\nu}$ is invariant with respect to $\mathbf{W}_i^{(0)}$.

The prerequisits for this type of matching techniques are:

1. The relative number ε of errorneous matches should be small, when taking the individual weights into account. The percentage $100 \cdot \varepsilon$ may practically reach 80 % or even 90 %, but then refers to matches which partially are of low weight.

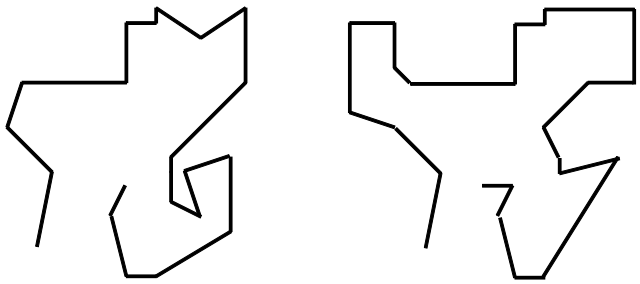


Figure 12: shows two segmented lines with straight segments. The symbolic description could be a sequence of length and directions.

2. The size of the errors should not be large, thus not exceed the size of the observations.

3. approximate values are necessary. Good approximate values, of course, speed up convergence. Wrong approximate values may prevent a valid solution.

Robust estimation using M-estimators has been successfully applied in DEM-Generation (cf. KRZYSTEK chap. 7) , spatial resection for exterior orientation (cf. SCHICKLER chap. 5) , automatic aerial triangulation (cf. ACKERMANN chap. 6) or model based building detection (cf. BRAUN chap. 10) .

3.3 String Matching

In many cases the description of an object may be an ordered list of features which has to be compared to an ordered list of features of a reference object.

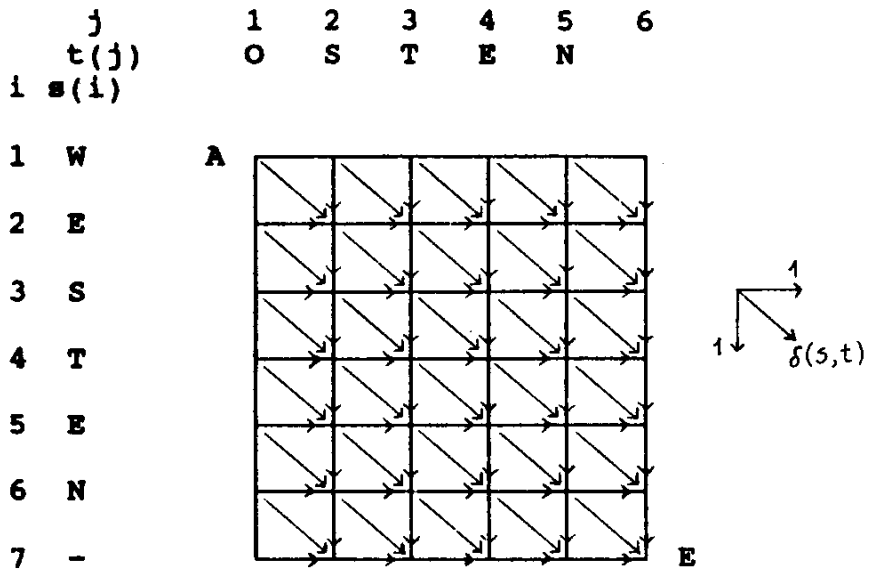
A typical example is the boundary description of a planar figure, e. g. a line with straight and curved parts. Such a line could be described by a sequence of -1's, 0's and 1's describing the sign of the curvature. Fig.12 shows two similar lines which may be compared. There are missing and additional parts and one line obviously has additional parts at the beginning or end.

Another example could be two ordered lists representing two intensity profiles which are to be matched. The lists $\{(x, sign(f_x), f_x^2)_i\}$ may consist of the edge points within the profiles with attributes e. g. being the position x and the sign $sign(f_x)$ and size f_x^2 of the intensity change.

In the extreme case the pixels of an intensity profile themselves could be interpreted as features with their intensity value as attribute (cf. the example below)

These problems can be interpreted as a string matching problem which occurs when searching a word in a dictionary assuming misspellings. Then the task is to find the word in the dictionary which is closest to the given word. Errors are:

- replacement
- deletion and



1. d(1,1) = 0
2. for i = 2 to |s|+1
   d(i,1) = d(i-1,1) + 1
3. for j = 2 to |t|+1
   d(1,j) = d(1,j-1) + 1
4. for i = 2 to |s|+1
   for j = 2 to |t|+1
   d1(i,j) = d(i-1,j) + 1
   d2(i,j) = d(i,j-1) + 1
   d12(i,j) = d(i-1,j-1) + δ(s(i-1),t(j-1))
   d(i,j) = min (d1(i,j), d2(i,j), d12(i,j))
5. D(s,t) = d(|s|+1, |t|+1)

Figure 13: a.shows the determination of the Levenshtein-distance as shortest path in a directed attributed graph (cf. Förstner 1991); b. shows the algorithm.

```
Match
   54323 31221 asdfq wetzw #eruw oryfl vgöcj ghkkf +h.12 341
3 11010 ----- ----- ----- ----- ----- ----- ----- ----- ---
4 -0-11 1---- ----- ----- ----- ----- ----- ----- ----- ---
1 --111 10--- ----- ----- ----- ----- ----- ----- ----- ---
2 ----- --0-- ----- ----- ----- ----- ----- ----- ----- ---
2 ----- ---0- ----- ----- ----- ----- ----- ----- ----- ---
1 ----- ----0 ----- ----- ----- ----- ----- ----- ----- ---
q ----- ----- 11110 ----- ----- ----- ----- ----- ----- ---
w ----- ----- ----- 0---- ----- ----- ----- ----- ----- ---
e ----- ----- ----- -0--- ----- ----- ----- ----- ----- ---
r ----- ----- ----- --1-- ----- ----- ----- ----- ----- ---

z ----- ----- ----- ---0- ----- ----- ----- ----- ----- ---
w ----- ----- ----- ----0 ----- ----- ----- ----- ----- ---
e ----- ----- ----- ----- 10--- ----- ----- ----- ----- ---
r ----- ----- ----- ----- --0-- ----- ----- ----- ----- ---
h ----- ----- ----- ----- ---1- ----- ----- ----- ----- ---
w ----- ----- ----- ----- ----0 ----- ----- ----- ----- ---
o ----- ----- ----- ----- ----- 0---- ----- ----- ----- ---
r ----- ----- ----- ----- ----- -0--- ----- ----- ----- ---
y ----- ----- ----- ----- ----- --0-- ----- ----- ----- ---
f ----- ----- ----- ----- ----- ---0- ----- ----- ----- ---

d ----- ----- ----- ----- ----- ---11 ----- ----- ----- ---
f ----- ----- ----- ----- ----- ---01 ----- ----- ----- ---
d ----- ----- ----- ----- ----- ---11 ----- ----- ----- ---
f ----- ----- ----- ----- ----- ---01 ----- ----- ----- ---
d ----- ----- ----- ----- ----- ---11 ----- ----- ----- ---
f ----- ----- ----- ----- ----- ---01 ----- ----- ----- ---
d ----- ----- ----- ----- ----- ---11 ----- ----- ----- ---
f ----- ----- ----- ----- ----- ---01 ----- ----- ----- ---
ß ----- ----- ----- ----- ----- ---11 ----- ----- ----- ---
g ----- ----- ----- ----- ----- ---11 ----- ----- ----- ---

i ----- ----- ----- ----- ----- ---11 ----- ----- ----- ---
f ----- ----- ----- ----- ----- ---01 ----- ----- ----- ---
l ----- ----- ----- ----- ----- ----0 ----- ----- ----- ---
v ----- ----- ----- ----- ----- ----- 0---- ----- ----- ---
g ----- ----- ----- ----- ----- ----- -0--- ----- ----- ---
ö ----- ----- ----- ----- ----- ----- --0-- ----- ----- ---
c ----- ----- ----- ----- ----- ----- ---0- ----- ----- ---
j ----- ----- ----- ----- ----- ----- ----0 ----- ----- ---
g ----- ----- ----- ----- ----- ----- ----- 0---- ----- ---
x ----- ----- ----- ----- ----- ----- ----- -1111 1---- ---

c ----- ----- ----- ----- ----- ----- ----- --111 11--- ---
v ----- ----- ----- ----- ----- ----- ----- ---11 111-- ---
1 ----- ----- ----- ----- ----- ----- ----- ----1 1110- ---
2 ----- ----- ----- ----- ----- ----- ----- ----- ----0 ---
3 ----- ----- ----- ----- ----- ----- ----- ----- ----- 0--
1 ----- ----- ----- ----- ----- ----- ----- ----- ----- -10
Distance =  32.0
```

Figure 14: shows the match of two strings. The optimal path is goes via the nodes denoted by 1 or 0. 1 denotes different symbols, 0 identical symbols.

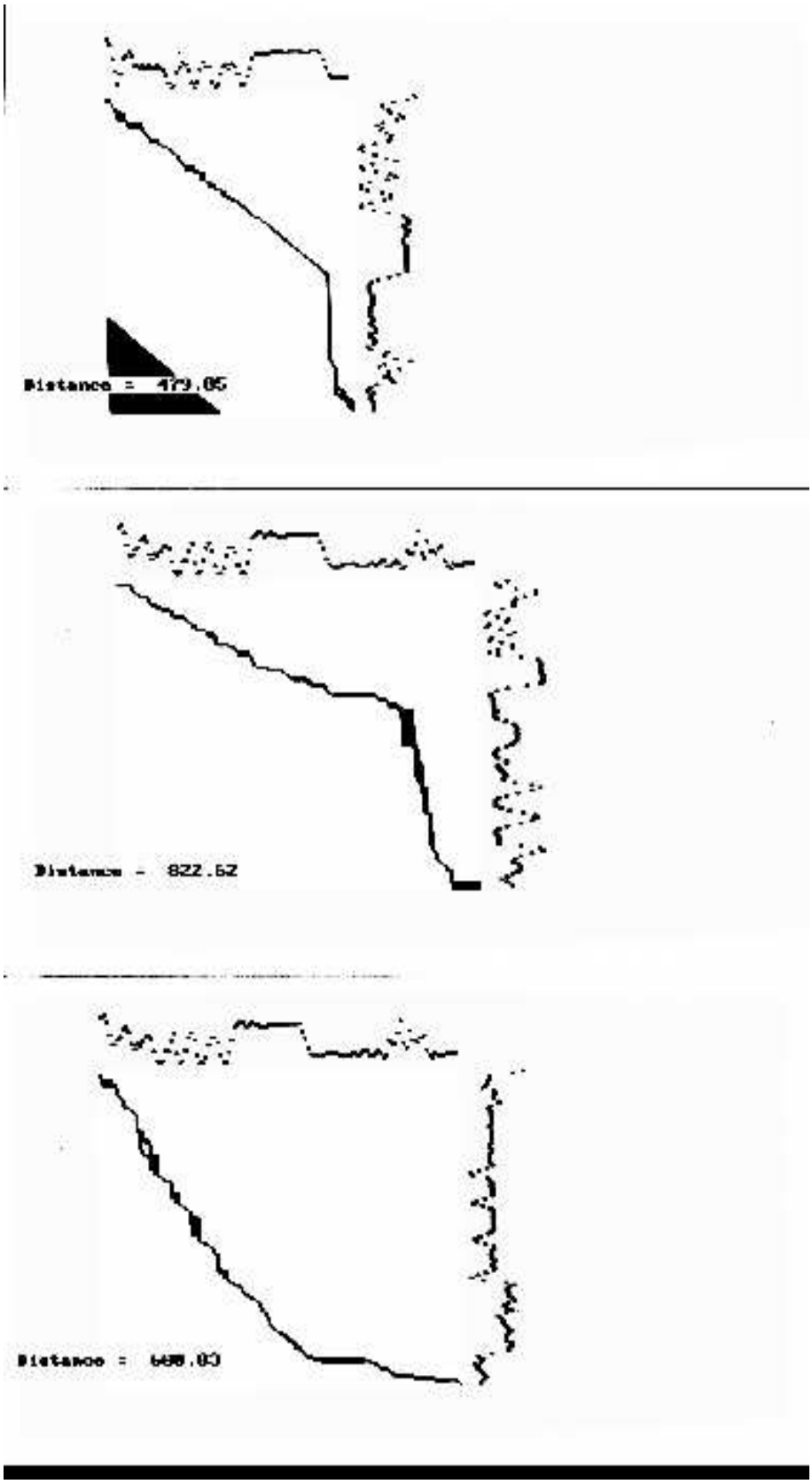


Figure 15: shows three samples of two grey value profiles and their match determined by a modified version of the algorithm in Fig.13 b.

- inclusion

of a letter.

The similarity of two strings $\mathsf{s} = \{s_i\}$ and $\mathsf{t} = \{t_j\}$ can be based on the *Levenshtein*-distance (cf. Niemann 1990). It is defined as the minimum number of the sum of deletions, insertions and replacements:

$$D(\mathsf{s},\mathsf{t}) = min(n_d + n_r + n_i) \tag{83}$$

where n_d: the number of symbols to be left out in the first string
n_r: the number of symbols to be replaced
n_i: the number of symbols to be left out in the second string

E. g. the Levenshtein-distance of s = OSTEN and t = WESTEN is 2, as 2 letters have to be replaced.

In the previous examples this procedure corresponds to replacing, deleting or including a segment in a line, an edge in the edge list, or a pixel in the profile. Here deletions and insertions may be necessary to cope with occlusions.

The matching algorithm itself can rely on the powerful tools of dynamic programming as in all these cases the optimization function, i. e. the similarity measure can be written as a sum of locally computable parts. This allows a recursive formulation of the optimization procedure which is the precondition for using dynamic programming techniques. In this context the distance can also be determined as the shortest path in a directed attributed graph (cf. fig.13 a.)

Fig. 15 shows the match of two intensity profiles, where the cost function takes into account the difference of the intensity values and the cost for a high slope in the profile.

The path from A to B following the edges corresponds to the match of the symbols of the strings s and t. A vertical edge represents a deletion of the previous symbol in s, a horizontal edge represents the deletion of the previous symbol in the string t, while a diagonal edge represents no deletion. The costs of passing a horizontal or a vertical edge are 1. The cost of passing a diagonal edge from $(i-1, j-1)$ to (i, j) depends on whether the symbols s_{i-1} and t_{j-1} are equal or not. Incase they are equal the cost of the edge is 1, else 0. The algorithm is given in fig.13 b.

The solution needs not to be unique. By backpropagation one is able to determine the corresponding symbols. Fig. 14 shows the match of two longer sequences. The symbols may represent a code for textures in intensity profiles of two epipolar lines. The optimal path goes via the 1's and 0's. It obviously is not unique. From the match the parallax profile may be derived.

This type of matching has been used for matching images which have a large number of discontinuities (cf. Baker and Binford 1982, Ohta and Kanade 1987).

3.4 Relational Matching

Relational matching is the most powerful matching technique available as it refers to a symbolic description of the image which in addition to the primitives also contains

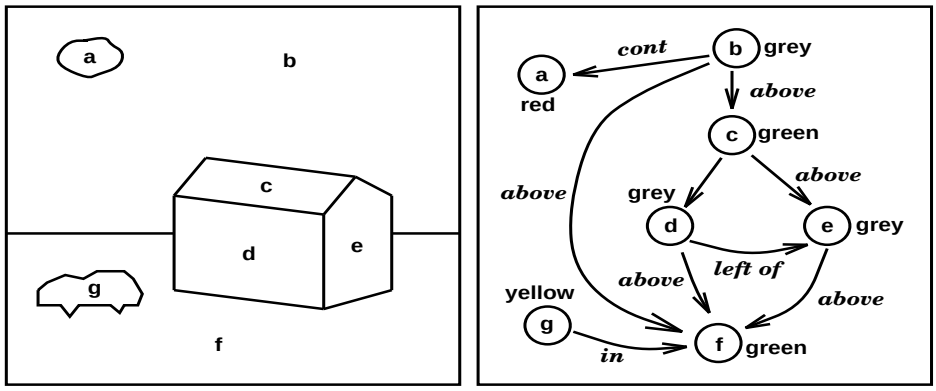


Figure 16: a. shows a segmented image with named regions as features, b. shows the relations between the segmented regions

relations between these primitives. We want to introduce the idea by giving an example, discuss one basic technique for the algorithmic solution and discuss the properties of the matching approach.

Assume you have a segmented image, as e. g. in Fig. (16). Of course this segmentation is idealised[1]. We have no difficulty in interpreting this scene, which means labelling the different region with names: 'sun', 'sky', 'roof', 'wall', 'grass', 'ground'. This labelling is a matching problem. Assume the descriptions of this image to consist of the region identifiers, their color, their form and their relations as shown in Table 1.

- Possible form attributes are:
  circular (circ), round, partially round (p-round), irregular (irreg), parallel (paral), rectangle (rect), triangle (triang), polygon(poly), pentagon (penta), oval.

- Possible relations are:

'no'	denotes	'no'	'cont'	denotes	'contains'
'in'	denotes	'in'	'neig'	denotes	'neigbour of'
'abov'	denotes	'above'	'any'	denotes	'any'
'belo'	denotes	'below'	'fron'	denotes	'in front of'
'left'	denotes	'left of'	'behi'	denotes	'behind of'
'righ'	denotes	'right of'	'clos'	denotes	'close'

On the other side you have a model of an outdoor scene, which consists of a list of possible objects with their expected colour, form and possible neighbourhood relations (Table 2 and 3).

Based on this information a correspondence can be established: (a, sun), (b, sky), $(c, roof)$, $(d, wall)$, $(e, wall)$, $(f, ground)$, (g, car). This is a mapping $T_0^i : O \rightarrow L$ where

[1] but remember the illustrations in children's books

id	COLOR	FORM	*a*	*b*	*c*	*d*	*e*	*f*	*g*
a	red	irreg							
b	grey	irreg	cont		abov			abov	
c	green	irreg				abov	abov		
d	grey	irreg					left	abov	
e	grey	irreg						abov	
f	green	irreg							
g	yellow	round						in	

Table 1: shows the extracted features (regions), their attributes and mutual relations

class	possible COLOR	possible FORM
sun	yellow, orange, red, white	circ, round, p-round, irreg
sky	blue, grey, black, white	irreg
roof	red, brown, grey, white, green, black	parall, rect, triang, poly, irreg
wall	white, grey, brown, black, pink	parall, rect, penta, poly, irreg
car	—	round, oval
ground	brown, grey, white, green	irreg

Table 2: shows the possible attributes of the object classes

	sun	*sky*	*roof*	*wall*	*car*	*ground*
sun	no	in, belo	abov, left, righ	abov, left, righ	abov	abov
sky	cont, abov		abov	abov	abov	abov
roof	belo, left, righ	belo	neig, left, righ	abov	any	abov
wall	belo	belo	belo	n, left, righ	any	abov
car	belo	belo	any	fron, behi, left, righ	any	abov, in
ground	belo	belo	belo	belo		close

Table 3: shows the possible neighbour relations.

$O = \{a, b, c, d, e, f, g\}$ and $L = \{sun, sky, roof, wall, car, ground\}$. Observe, that there could be multiple solutions.

Formally the matching problem can be stated as follows. Let the two descriptions of the images be

$$D_1 = (P_1, R_1) \qquad D_2 = (P_2, R_2) \tag{84}$$

The lists P_1 and P_2 contain attributed primitives like in the previous example, e. g. ((COLOR red),(FORM irreg)) representing the color and form of the object a.

Further these lists contain relations between the primitives. In the example above this e. g. could be (inside, sun, sky) stating the region sun lying inside the region sky.

This type of matching problem, especially in more complex situations, heavily relies on the expected relations between the features.

There are several ways to solve this matching problem. The most important are

- discrete relaxation (cf. Price 1985, Hummel and Zucker 1983)
- tree search (cf. Shapiro and Haralick 1987)

To set an example for the algorithmic solution we will explain the principle of tree search. In this case the search space of the correspondence problem is represented in a tree consisting of nodes and edges (Nilsson 1971). Each node stands for a possible mapping between one primitive element of P_1 and one primitive element of P_2. The edges represent the connections between different nodes. Thus starting with the root node, representing the initial problem state, each path from the root node to the bottom leaf node is a possible mapping of the descriptions. The depth of the tree is equal to the number of primitives in the list P_1.

The most general way to find a mapping is to completely scan the tree in a systematic manner, called blind search, namely Depth-First and Breadth-First Search.

The Depth-First Search always expands first the node with the deepest level. If there doesn't remain any successor node, backtracking starts up to that predecessor node, which has an alternative successor. The search will stop if a solution is found.

An alternative possibility to proceed is the Breadth-First Search, which always axpands every successor node on the actual level before moving down in the tree. Observe that the size of the search tree exponentially grows with the number of levels, making a complete search unfeasible already for middle size problems.

In contrast to blind search, it is therefore useful to include information, which path is the most likely one to lead to a solution. This kind of informed search can be done by associating costs or merits to the edges of the tree. The best solution is then defined as the path from the root node to a leaf node, which has the lowest sum of costs or the maximum merit respectively with the edges of the path (cf. Hill Climbing Method or Best First Search).

Including some heuristics, these evaluation functions more generally can be defined as the sum $f(n) = g(n) + h^*(n)$ of the minimum cost path $g(n)$ from the root node to the actual node n plus the estimated cost of a minimal cost path from the actual node to a goal node. That is $f(n)$ is an estimate of the cost of a minimal cost path constrained to go through node n (cf. A* Algorithm). Thus it is possible to avoid the

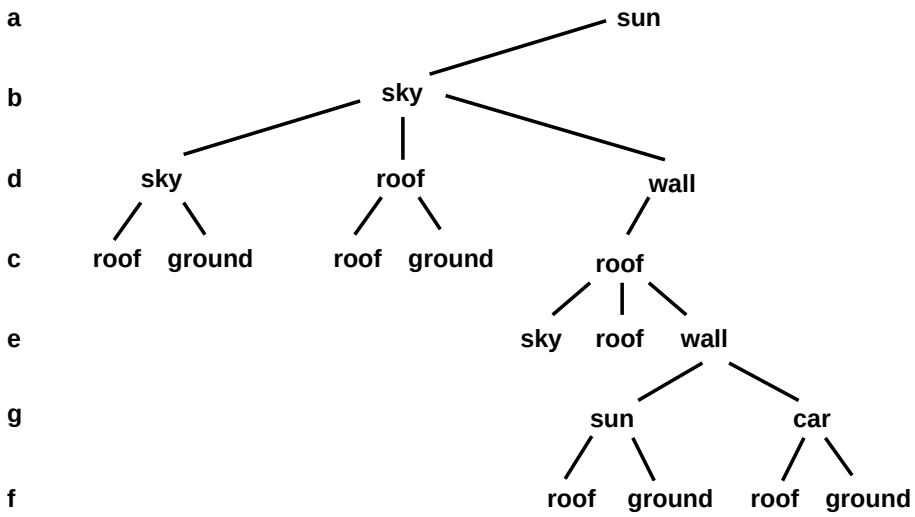


Figure 17: shows the tree which is build up during depth-first-search for the given example Tab. 1-3

search to stay on an upper level because of the costs being low, as it can happen with Best First Search.

Due to the exponential complexity of tree search, these methods can only be applied to sparse data sets, like high level descriptions. As all possible combinations between primitives of the descriptions P_1 and P_2 can be evaluated, no a priori knowledge about the spatial transformation between the 2 descriptions is required. This makes the tree search method suitable for tasks like object recognition and object localization where no such information is available (Vosselman 1992, Vosselman 1994).

To give an example for Depth-First Search we want to illustrate the tree which is built up during the search if the list orderings are $P_1 = \{a,\ b,\ d,\ c,\ e,\ g,\ f\}$ and $P_2 = \{sun,\ sky,\ roof,\ wall,\ car,\ ground\}$.
The solution is given by the labelling $\{\ (a, sun),\ (b, sky),\ (d, wall),\ (c, roof),\ (e, wall),\ (g, car),\ (f, ground)\}$ and is found by 19 trials (cf. Fig. 17).
Which branches of the tree are to be expanded during the search, mainly depends on the ordering of the matching units within the lists. This can be used for advantage, if durig the built up of the unit orderings, those units with the least possible corresponencies are put at the root of the tree. Thus the search tree will be reduced by always first labelling the unit with the fewest possible labels. In case the ordering of list P_1 is $P_1 = \{f,\ g\ ,e,\ d\ ,c\ ,b\ ,a\ \}$ certainly the same solution will be found, but the search time requires up to 25 trials.

Another possibility to reduce search space is by forward checking, which is looking forward to the units of a lower level, to check if there remain possible labels after labelling on the higher levels.

Discussion:

1. As the symbolic image descrition being invariant with respect to different geometric transformations, even distortions, under such circumstances relational matching techniques are very useful.

2. As the numbers of possible correspondencies is very large, it is necessary to reduce the search space as soon as possible. Besides the algorithmic techniques this can be done including constraints due to the geometric model or scene knowledge.

3. In contrast to area based or feature based matching using a local similarity measure, within relational matching the similarity measure is a globel one, including the hole image information (primitives, their attributes and mutual relations). Thus relational matching doesn't need any approximate values and thus is not liable to find a wrong match due to insufficient approximate values.

4 Conclusions

This contribution gave an introduction in different matching techniques which can be used to establish the correspondence or match between several images or between one or several images and a model as it has to be done in image matching (cf. ACKERMANN chap. 6, KRZYSTEK chap. 7), object localization (cf. SCHICKLER chap. 5) and object reconstruction (cf. GÜLCH chap. 12). We introduced the two fundamental techniques area based matching and feature based matching, using different abstraction levels of the images and models.

References

Ackermann, F.; Hahn, M. (1991): Image Pyramids for Digital Photogrammetry. In: Ebner, Fritsch, Heipke (Ed.), *Digital Photogrammetric Systems*. Wichmann-Verlag, 1991.

Ackermann, F. (1984): High Precision Digital Image Correlation. In: *39. Photogrammetric Week*. Schriftenreihe des Instituts für Photogrammetrie, Heft 9, Universität Stuttgart, 1984.

Baker, H.H.; Binford, T.O. (1982): Depth from Edges and Intensity Based Stereo. In: *International Joint Conference on Artificial Intelligence, Vancouver*, pages 631–636, 1982.

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

Blake, A.; Zisserman, A. (1987): *Visual Reconstruction*. MIT Press, Cambridge, 1987.

Boyer, K.L.; Kak, A.C. (1988): Structural Stereopsis for 3D Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10:144–166, 1988.

Förstner, W. (1982): On the Geometric Precision of Digital Correlation. *International Archives of Photogrammetry and Remote Sensing, Helsinki*, 24-III:176–189, 1982.

Förstner, W. (1984): Quality Assessment of Object Location and Point Transfer Using Digital Image Correlation Techniques. *International Archives of Photogrammetry and Remote Sensing, Rio de Janeiro*, 25-A3a:197–219, 1984.

Förstner, W. (1986): A Feature Based Correspondence Algorithm for Image Matching. In: *International Archives of Photogrammetry and Remote Sensing, Vol.26-3/3, Rovaniemi*, pages 150–166, 1986.

Förstner, W. (1991): *Statistische Verfahren für die automatische Bildanalyse und ihre Bewertung bei der Objekterkennung und -vermessung*, Band 370 der Reihe C. Deutsche Geodätische Kommission, München, 1991.

Grimson, W.E.L. (1981): *From Images to Surfaces: A Computational Study to the Human Early Visual System.* MIT Press, Cambridge, MA, 1981.

Hahn, M. (1994): *Bildsequenzanalyse für die passive Navigation.* PhD thesis, Institut für Photogrammetrie, Universität Stuttgart, Stuttgart, 1994.

Hannah, M.J. (1974): *Computer Matching of Areas in Stereo Images.* PhD thesis, Stanford University, Stanford, CA, 1974.

Huang, T.S. (1981): *Image Sequence Analysis.* Springer-Verlag, New York, 1981.

Hummel, R.; Zucker, S. (1983): On the foundations of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:267–287, 1983.

Jähne, B. (1989): *Digitale Bildverarbeitung.* Springer-Verlag, 1989.

Kreiling, W. (1976): Automatische Auswertung von Stereobildern durch digitale Korrekation. *International Archives for Photogrammetry and Remote Sensing, Helsinki*, 23-3, 1976.

Nagel, H.-H.; Enkelmann, W. (1981): Iterative Estimation of Displacement Vector Fields from TV-Frame Sequences. In: *Second European Signal Processing Conference, Erlangen , Germany*, 1981.

Niemann, H. (1990): *Pattern Analysis and Understanding.* Springer-Verlag, 1990.

Nilsson, N. J. (1971): *Problem-solving in Artificial Intelligence.* McGraw-Hill, New York, 1971.

Ohta, Y.; Kanade, T. (1987): Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7:139–154, 1987.

Paderes, F.C.; Mikhail, E.M.; Förstner, W. (1984): Rectification of Single and Multiple Frames of Satellite Scanner Imagery Using Points and Edges as Control. In: *NASA Symposium on Mathematical Pattern Recognition and Image Analysis , Houston*, 1984.

Price, K.; Huertas, A. (1992): Using Perceptual Grouping to detect Objects in Aerial Scenes. In: *Internat. Archives for Photogrammerty, Comm. III, Washington*, Band 29, pages 842–855, 1992.

Price, K.E. (1985): RelaxationMatching Techniques – A Comparision. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-7:617–623, 1985.

Rosenholm, D. (1987): Multi-point matching using the least-squares technique for elevation of three-dimensional models. *Photogrammetric Engeneering and Remote Sensing*, 53(6):621–626, 1987.

Sester, M.; Förstner, W. (1989): Object Location Based on Uncertain Models. In: *11. DAGM-Symposium, Hamburg*, 1989.

Shapiro, L.G.; Haralick, R.M. (1987): Relational Matching. *Applied Optics*, 26:1845–1851, 1987.

Stockman, G.C.; Kopstein, S.; Bennett, S. (1982): Matching Images to Models for Image Registration and Object Location via Pose Clustering. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-4:229–241, 1982.

Stockman, G.C. (1987): Object Recognition and Localization via Pose Clustering. *Computer Vision, Graphics, and Image Processing*, 40:361–387, 1987.

Terzopoulos, D. (1986): Image Analysis Using Multigrid Relaxation Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:129–139, 1986.

Terzopoulos, D. (1986): Regularization of Inverse Visual Problems Involving Discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(2):129–139, 1986.

Tsingas, V. (1992): *Automatisierung der Punktübertragung in der Aerotriangulation durch mehrfache digitale Bildzuordnung*, Band 392 der Reihe C. Deutsche Geodätische Kommission, München, 1992.

Vosselman, G. (1992): *Relational Matching.* Lecture Notes in Computer Science 628. Springer-Verlag, 1992.

Vosselman, G. (1994): On The Use Of Tree Search Methods in Digital Photogrammetry. In: Ebner, Heipke, Eder (Ed.), *ISPRS Vol. 30, 3/2*, pages 886–893. SPIE, 1994.

Weidner, U. (1994): Parameterfree Information-Preserving Surface Restoration. In: Eklundh, J.-O. (Ed.), *Computer Vision - ECCV 94, Vol. II, Proceedings*, pages 218–224, 1994.

Witkin, A.; Terzopoulos, D.; Kass, M. (1987): Signal Matching through Scale Space. *International Journal on Computer Vision*, 1:133–144, 1987.

Wrobel, B. (1987): A New Approach to Computer Stereo Vision and to Digital Photogrammetry. In: *Intercommunication Conference on Fast Processing of Photographs, Interlaken, Switzerland*, pages 231–258, 1987.