A SURVEY ON STEREO MATCHING TECHNIQUES

Mathias J.P.M. Lemmens

Delft University of Technology, Faculty of Geodesy
Institute of Fotogrammetry and Remote Sensing
Thijsseweg 11, 2629 JA Delft, The Netherlands

COMMISSION V/6

Abstract: Tracing of corresponding phenomena (matching) is necessary in many applications, like image sequence analysis, aerotriangulation and 3-D geometric information extraction from stereo pairs of images. An important industrial application is robot stereo vision. In digital images the correspondence problem can be approached by signal, feature or relational matching. Common, at the moment, is the signal approach. Because of its limitations more and more image structures (features) and their mutual relationships are used. The aim of this paper is to review the current stereo matching techniques.

1 Introduction

Collecting information about the real world by means of images is already for a long time well-accepted and widely utilized. There is no unifying theory on which automatic information extraction from digital images can be founded. There are concepts which work well for some vision problems, but they are not general enough to be appropriate for other applications. The lack of theoretical foundation has, among others, led to an overwhelming diversity of approaches to tackle the correspondence problem in matching of image sequences, both by the computer vision community and the photogrammetric community. The aim of this paper is to order and review the existing diversity.

Matching is a very general notion to select corresponding phenomena in two or more observation sets. In advance, it is unknown which phenomena in the one set correspond to the phenomena in the other set. Corresponding phenomena are different mappings of the same object phenomena. Moreover, the counterpart of an element in the one set may even be absent in the other one, and reversely, i.e. the relation is not bijective; elements of the one set may be mapped into the null-space of the other, and reversely. Contrary to classical mapping no transformation is known; it has to be adjusted. This causes the correspondence problem to be severe.

In stereo and motion vision matching refers to identifying corresponding visual phenomena in image sequences, caused by the same phenomena in object space. This survey is devoted to the reconstruction of surfaces from stereo images. There are other passive methods to extract 3-D information from mono images, e.g. shape from (1) shading, (2) texture and (3) focussing. Also active ranging, using laser and radar, is employed. They all are out of the present scope.

The purpose of stereo vision is surface recovery of 3-D object space from conjugate image pairs. It produces a more quantitative depth determination than the passive mono techniques. Its passive state makes it more generally applicable than active ranging (Medioni and Nevatia, 1985). 3-D surface description lies at the basis of a structural description of the real world. It allows, for instance, the compilation of orthophoto's from a single frame or a 3-D impression of spatial mono images (e.g satellite imagery) by superposition. In industrial applications a 3-D surface description defines the entire object structure and real-time processing plays a bottle neck. Digital photogrammetry applies matching techniques for the determination of Digital Elevation Models (DEM) and point transfer in aerotriangulation.

The problem of stereo analysis consists of the following main stages:

1 extraction of phenomena (i.e. elementary descriptors or tokens) in both images;
2 selection of corresponding phenomena and computation of their object space coordinates from triangulation;
3 interpolation to arrive at a full 3-D surface description.

Following the terminology of Marr (1979) these steps lead respectively to (see also fig. 1):

- the primal sketch;
- the $2\frac{1}{2}$-D sketch;
- complete surface description.

Since the $2\frac{1}{2}$-D sketch resembles a DEM, i.e. a 3-D point field ((X, Y, Z)-coordinates), it may be stated that, in general, the photogrammetric task is completed after determination of the $2\frac{1}{2}$-D sketch. To acquire the $2\frac{1}{2}$ sketch from disparaties, a camera model, that is the exterior orientation of the stereo pair, has to be known. Matching is the key step and in course of time a variety of approaches are developed. A first, broad subdivision is obtained from the manner an image may be looked at:

1 as signals, i.e. a 2-D spatial distribution of functions of E.M. intensities;
2 a set of features.

The first view is the classical one and frequently applied. It has the advantage of being fully compatible with the well-defined and thoroughly studied concepts of signal processing, in particular the digital variant. Especially radiometric restoration, enhancement by smoothing, and edge detection gain profit from a signal approach. Corresponding phenomena are identified by correlation techniques using neighbourhood of pixels. As such, signal matching may be regarded as the digital continuation of the analogue approach, using cathode ray tubes, introduced by Hobrough in the late fifties.

Ullman (1979) notices two objections against signal correlation.

1 A correct match is only yield in the very simple case of shift;
2 Grey value distributrions don't correspond to physical entities.

Actually, (1) refers to geometric transformations and (2) to radiometric transformations. Regardless whether

Ullman is right or wrong, currently the developments point from signal matching to feature and even to relational matching.

Regarding an image as a structure is the most natural way to do, because structures is what one wants to describe. The treatment of an image as a 2-D signal is just an attempt to get there. The feature approach didn't enter computer vision until a computational theory of human vision was developed by Marr and co-workers (Marr and Poggio, 1979; Marr and Hildreth, 1980; Marr, 1979; Grimson, 1981). The observation that random dot stereograms, which have no structure at all (see fig. 2), are perceived as depth images by the human visual system, did Marr and co-workers decide that in the process of stereo perception not first meaningful structures in the individual images are detected before junction to stereo impression, but that just elementary tokens are detected (the primal sketch). This approach leaves behind the earlier insights that the human visual system matches at a higher, that is structural, level. The lack of success of the feature approach has led to a revival of the structural approach or more often called, relational matching (c.f. Shapiro and Haralick, 1987). In this approach image structures and their relationships to neighbouring structures are described by a symbolic representation.

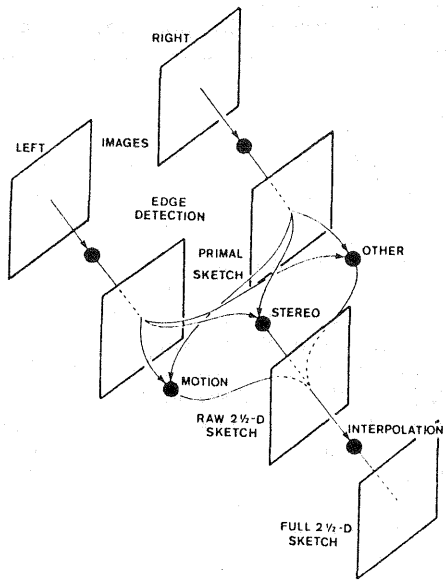

Fig. 1 The three stages of the correspondence problem: (1) extraction of phenomena, (2) selection of corresponding phenomena and (3) full 3-D surface description by interpolation. Both stereo vision and motion vision are indicated.

Based upon the above considerations three classes of matching can be distinguished:
- signal matching
- feature matching
- relational matching.

Signal matching is often referred as area-based matching, but this term doesn't express the principal notion of its background. The classes are treated in detail in separate sections, but first some general notions on matching are outlined in the next section.

2. General considerations on matching

The foundation of stereo vision is the recording of an object space from slightly different view points. The difference in positions causes disparities and from triangulation, using the exterior orientation, 3-D object coordinates are computed. By interpolation the 3-D surface structure can be recovered. So, the three main steps in any stereo vision algorithm are (see also introduction):

1 Detection of items or phenomena.
2 Matching and calculating depth.
3 Surface recovery.

The phenomena may be:

1 non-local (i.e. neighbourhoods of pixels)
2 local (e.g. edges and blobs).
3 locally extended (e.g. line segments and areas).

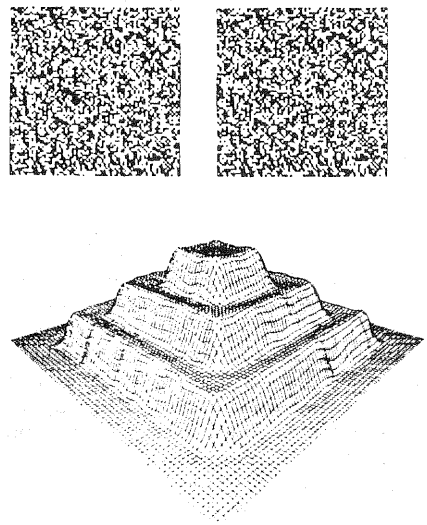

Fig. 2 Random dot stereograms give 3-D impression. This discernment is the foundation of feature matching

A non-local item consisting of the grey values of a neighbourhood of pixels is a target area (also the term mask is in use), a local item is a feature and a locally extended item a structure. So the above subdivison of phenomena defines broadly the three types of matching mentioned in the introduction. Once items are detected in both images the correspondence problem has to be solved. First of all similarity measures are necessary, which depend on the description of the item, e.g. in case of a non-local item a correlation measure is suited. But, a similarity measure gives only a local match, no global consistency, since no similarity measure is immune to misinterpretation. They will, in general, not lead to unambiguous matches. So a global match is necessary. The ambiguity causes, what is called, the false target problem (Marr and Poggio, 1979; c.f. Marr, 1979; Grimson, 1981), i.e. an item in image 1 may match equally well several items in image 2. Fig. 3. shows an example, following Marr, 1979. Each circle L_i, i = 1, ... 4 is similar to any of the circles R_i, i = 1, ...4 in the conjugate image. A priori, any of the 16 possible matches is reasonable. When we match L_1 with R_4, L_2 with R_3, L_3 with R_2 and L_4 with R_1, the circles are seen in a vertical line, but the human visual system never perceives this match. It prefers to make the correspondence between L_1 with R_1, L_2 with R_2, L_3 with R_3, and L_4 with R_4. That is a plane, shown with the filled circles. So, humans prefer the most simple match.

The most likely solution can be found when information about the plausibility of different matches is available. So, an object model and additional information, like maximal disparity are indispensable to decide which matches are correct. Two physical conditions

are relevant to constraint the matches (Marr and Poggio, 1979; c.f. Marr, 1979; Grimson, 1981):

- A given point on an object surface, has a unique position in space at any one time;
- Matter is cohesive, it is separated into objects, which surfaces are generally smooth, i.e. the surface fluctuations are small compared with the viewing distance.

The above has brought Ullman (1979) to the formulation of the minimal mapping theory, which states, roughly, that the matches which generate the simplest surface, are the best. In photogrammetry, planes are taken as surface model (Förstner, 1986).

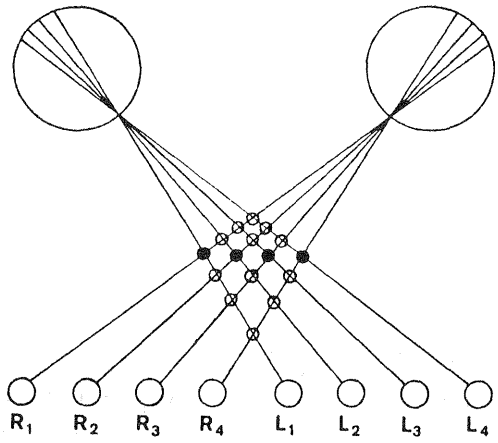

Fig. 3 The False Target Problem of stereo matching. The circles in the one image can match equally well all circles in the other image.

The correspondence problem is constrained by three conditions (Marr, 1979):

1 Compatibility constraint: if two elementary tokens could have arisen from the same object item, they can match, else they can not match;
2 Uniqueness constraint: each elementary token in the one image can only match one element in the other;
3 Continuity constraint: disparities vary smoothly almost everywhere.

There are four main factors responsible for the grey values in an image (Marr, 1979), (see fig. 4):

- illumination;
- reflectivity;
- geometry;
- viewpoint.

To achieve unambiguous matches all four factors should be known in advance. However, only the position and attitude (i.e. exterior orientation) of the viewpoints and the illumination source(s), are, in general, known, and, moreover, the geometry of the object space surface, that is what we are looking for. The reflectivity of the surface is generally also unknown. Because, except in cases of perfect Lambertian reflection, (i.e. the reflections are in all directions the same) the geometry affects the amount of light that arrives at the view point, even when the surface reflectivity is known, no unique image grey value distribution is determinable.

Marr and co-workers evade the reflectance problem by deriving primitive features, like edges and blobs, from the images à priori to matching, assuming that they are independent of surface reflectivity. This approach leads, however, to many possible solutions. To limit the solution space, an object model is necessary. Marr and co-workers use the model of simplest surface.

Signal matching doesn't allow surface reflectivity ignorance. Commonly, implicitely or explicitely, a perfect Lambertian reflection is assumed. Grey value correlation is limited to image pairs with only small perspective changes from one view to the other, i.e. the base-to-height ratio is small (cf: Hannah, 1974), assuming Lambertian reflection and just a shift between the two views. The shift assumption is valid for satellite images, gained by Landsat and SPOT, but in close-range photogrammetry it doesn't hold. The least squares matching (LSM) approach requires an aprroximate surface model, gained e.g. from prediction from previous measures or from feature matching performed in a preprocessing step. The approximate values have to be very precise, even such that for many applications the approximations are good enough as the final matching result. So, LSM is actually a fine correlation method to bring the known matches to a higher precision level.

Let us now describe the above more formally. Let P^0 be the set of items or phenomena p_i^0, i = 1, .., h in object space. Let P^1 be the set of phenomena p_i^1, i = 1, .., k in the first image of the conjugate pair and P^2 the set of phenomena p_i^2, i = 1, .., l in the other image. The nature of imaging causes that the number of phenomena in object space, images 1 and image 2 is different. Let us call the phenomena which are present in image 1 but absent in image 2, the null space of p_o^1 of image 1. The null space p_o^2 of image 2 is defined accordingly.

The task of matching is to select for each p_1^1 one and only p_i^2, under the condition that p_i^1 and p_i^2 are mappings of the same object phenomenon p_i^0. To each image phenomenon a set of property attributes A consisting of the elements a_i, i = 1, .., m, describing some properties, can be assigned, e.g. mean grey value grey value variance, length, shape, area and elongatedness. Further, to each image phenomenon a set of relational attributes R, consisting of the elements r_i, i = 1, .. n, describing the kind of relationship with the neighbouring phenomena, can be assigned, e.g. above, beneath, left, right and nearly.

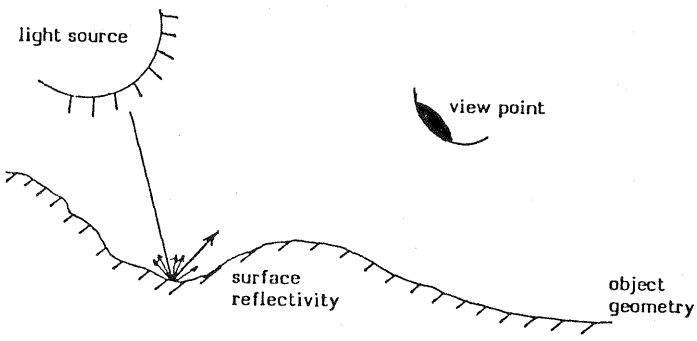

Fig. 4 The four main factors responsible for image grey values.

The attributes are, actually, derived from the grey values of the pixels a phenomenon is built of or/and the neighbourhood. So, the most simple description of a phenomenon is just an image patch, f.i. a 5 x 5 window. The shape of the patch can be regular (e.g. quadrangular and rectangular) or irregular. The most simplest way of phenomena selection is by covering the image by a regular grid and to take patches at the intersection lines. This approach may cause that also

patches are taken with a smooth grey value function, i.e. unsuitable for matching. Therefore, often the images are preprocessed to find pronounced phenomena like corners and blobs. Hannah (1974) selects regions with a steep autocorrelation function and Moravec (1977) employs a directional variance evaluation operator.

An image patch consists of the original greyvalues G with elements g_i , i = 1, .., n, n the number of pixels. From the above follows that each phenomenon can be described by G, A and R. All three sets are employed in the matching process. If just G is used, signal matching is performed. Feature matching needs the set A, but doesn't exclude G. Relational matching needs R and may also make use of G and A.

3 Signal Matching

Neighbourhoods of pixels, G, are used. The similarity is indicated by the resemblance between grey values.

The simplest method is to take some statistical measure, f.i. cross-correlation, to shift the target area, defined in the one image, over a search space in the other image, and to compute the correlation function (see fig. 5). The size of the search space depends on how well the exterior orientation is known and the possible height differences. The mid-pixel of the target area and that of the most similar search area are taken as corresponding points. The similarity measure should exceed a threshold.

Another approach is to define the unknown corresponding coordinates (x_s, y_s) in the search area directly as a function of the grey values of the target image G_t and search area G_s. This leads to the non-linear equation:

$$g_t(x_t, y_t) = g_s(x_s, y_s) + n(x, y).$$

With (x_t, y_t) given coordinates in the target area to which we want to assign the corresponding coordinates (x_s, y_s) in the search area and n the additive noise. If approximate values (x_s^o, y_s^o) are known the equation can be linearized to obtain a solution in the least squares sense. For convergence (x_s^o, y_s^o) should not differ much from the exact value, i.e. very accurate approximate values are necessary. LSM has, however, the distinct advantage of very high subpixel accuracy (e.g. 1/20 pixel size). Other geometric transformations than shifts can be handled too. A model to describe grey value differences can be introduced.

Originally developed as single point matching (Ackermann, 1984; Pertl, 1984, 1985; Förstner, 1984), some refinements are introduced in LSM. Grün and Baltsavias (1987) use multiple views to incorporate geometrical constraints. Rosenholm (1986) introduces multiple point matching. Already mentioned is the approach of Wrobel (1987[a,b]). The next subsection elaborates grey value correlation and LSM.

3.1 Grey value correlation

Grey value correlation employs statistical i.e. covariance(-like) measures as simularity measures. Let g_i^t be the target area and g_i^s the search area in a digital image, i = 1, .., n with n the number of pixels in the area (see fig. 5). The search area is a subimage of the search space. In denoting the similarity measures R_{ts} between target and search area, we will partly follow (Hannah, 1974) where a very nice coherent synopsis is given. The common similarity measure is discrete correlation:

$$R^1_{ts} = \sum_i g_i^t \cdot g_i^s \qquad (3.1)$$

Normalizing R^1_{ts} by the means $\bar{g}^t$ and $\bar{g}^s$ leads to:

$$R^2_{ts} = \sum_i (g_i^t - \bar{g}^t)(g_i^s - \bar{g}^s) \qquad (3.2)$$

Normalizing by the second moments:

$\sum_i (g_i^t)^2$ and $\sum_i (g_i^s)^2$ leads to:

$$R^3_{ts} = \frac{\sum_i g_i^t \cdot g_i^s}{\sqrt{\sum_i (g_i^t)^2 \cdot \sum_i (g_i^s)^2}} \qquad (3.3)$$

Normalizing by both the sample means and the second moments leads to the cross-correlation:

$$R^4_{ts} = \frac{\sum_i (g_i^t - \bar{g}^t)(g_i^s - \bar{g}^s)}{\sqrt{\sum_i (g_i^t - \bar{g}^t)^2 \cdot \sum_i (g_i^s - \bar{g}^s)^2}} \qquad (3.4)$$

With the property: $-1 \leq R^4_{ts} \leq 1$.

The above measures have to be maximized to find the best correlation. Also in use are variance measures, considering:

g_i^t and g_i^s

as two samples of the same observation set. Their measures have to be minimized to find the best fit.

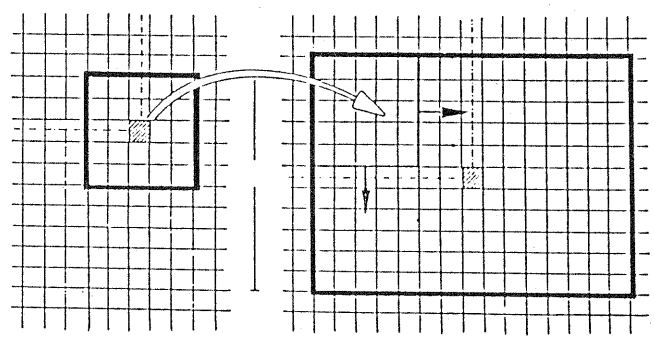

Fig. 5

The Gaussian, i.e. root mean square error, is given by:

$$R^5_{ts} = \sqrt{\frac{1}{n} \sum_i (g_i^t - g_i^s)^2} \qquad (3.5)$$

Normalizing by the sample means:

$$R^6_{ts} = \sqrt{\frac{1}{n} \sum_i \left((g_i^t - \bar{g}^t) - (g_i^s - \bar{g}^s) \right)^2} \qquad (3.6)$$

The Laplacian variance measure, i.e. the sum of the absolute values of the differences, is given by:

$$R^7_{ts} = \frac{1}{n} \sum_i | g_i^t - g_i^s | \qquad (3.7)$$

which can also be normalized by the sample means:

$$R^8_{ts} = \frac{1}{n} \sum_i | g_i^t - g_i^s + \bar{g}^s - \bar{g}^t | \qquad (3.8)$$

Commonly, the computations are carried out using a standard target area, i.e. n is equal for all points, thus can be neglected. Therefore in R^j_{ts}, j = 5, .., 8, n in the denominator can be leaved out. To take account of the geometric differences between the target and search area, a weight function w_i can be introduced, which favours the central parts at the cost of the surrounding portions.

$$R^9_{ts} = \frac{\sum_i w_i \cdot (g_i^t - \bar{g}^t)(g_i^s - \bar{g}^s)}{\sqrt{\sum_i w_i (g_i^t - \bar{g}^t)^2 \cdot \sum_i w_i (g_i^s - \bar{g}^s)^2}} \qquad (3.9)$$

This similarity measure is, for instance applied by Mori et. al. (1973) to match aerial photographs taken at a height of 3000m. In order to become a computationally more efficient expression for the cross correlation R^4_{ts} the terms in R^4_{ts} should be rearranged to:

$$R^{4'}_{ts} = \frac{\sum_i g_i^t \cdot g_i^s - (\sum_i g_i^t \cdot \sum_i g_i^s)/n}{\sqrt{(\sum_i (g_i^t)^2 - (\sum_i g_i^t)^2/n) \cdot (\sum_i (g_i^s)^2 - (\sum_i g_i^s)^2/n)}} \qquad (3.10)$$

In case of multi spectral images each pixel consists of a vector of m grey values:

$g^t_{i,j}$ and $g^s_{i,j}$, j= 1, .., m.

The multispectral cross-correlation can now be defined by (cf: Hannah, 1974):

$$R^{10}_{ts} = \frac{\sum_i \sum_j (g^t_{i,j} - \bar{g}^t_j) \cdot (g^s_{i,j} - \bar{g}^s_j)}{\sqrt{\sum_i \sum_j (g^t_{i,j} - \bar{g}^t_j)^2 \cdot \sum_i \sum_j (G^s_{i,j} - \bar{g}^s_j)^2}} \qquad (3.11)$$

The terms can be rearranged like $R^{4'}_{ts}$.

The search area which has the largest similarity with i.e. the target area is taken as corresponding area. Of course, R_{max}, the maximum similarity value, should exceed a certain threshold. Generally, their mid-pixels are taken as corresponding points. To arrive at subpixel accuracy it is possible to view the distinct R's as discrete samples taken from a continuous correlation function. The neighbouring R's of R_{max} are used to approximate the correlation function by a continuous function. At a local extremum (and this will be the maximum in case of an orderly function) the partial derivatives to x and y, df(r)/dx and df(r)/dy, will be zero. Wiesel (1981) approximates f(r) separately in x and y direction by a second order polynomial:

$$f(r) = a_0 + a_1 x + a_2 x^2$$
$$f(r) = b_0 + b_1 y + b_2 y^2$$

leading to the subpixel position of R_{max} at:

$$x_{max} = -a_1/2a_2 \text{ and } y_{max} = -b_1/2b_2.$$

The precision of the above procedure with regard to the localization of control points in digital remote sensing images is analyzed by (Wiesel, 1981). Claus (1983) employs the method for the determination of a coarse DTM from a scanned aerial stereo photograph, using epipolar geometry to reduce search space. Hannah (1974) approximates f(r) by e^{-q}, with

$$q = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy$$

i.e. log f(r) = q.

A critical point is the choice of the threshold value of the similarity measure, i.e. its lower boundary. This choice seems to be rather subjective, depending on the investigator and its background. Something to go by is to test the hypothesis whether the cross-correlation coefficient R^{cross} between two observation sets, each of sample size N, differs significantly from 0, using student's t test. In table .1, for a series of N, the lower bounds R_0 are listed for the one-sided confidence levels 99.5%, 95% and 90%. In case of a 5 x 5 target, i.e. N = 25, R^{cross} should exceed at least 0.5 to obtain, at the 99.5% confidence level, the certainty that there exists at least some similarity between the signals. In practice, R_0 should be much higher.

Ehlers (1983) has compared five correlation methods (also some not mentioned here) with respect to reliability, precision, stability, convergence and computation time and tested on scanned aerial photographs.

Cross-correlation combined with the vertical line-locus method is applied in digital stereo-photogrammetry by Kern in restitution instruments, equiped with digital cameras (Bethel, 1986). Its precision with respect to kind of texture, Z-spacing, window-size, and so on, is investigated by Almroth and Hendriks (1987) and Hendriks (1988).

In digital remote sensing grey value correlation is employed to find topographic control points (Billingsley, 1983) to automatically rectify satellite images. The target areas are masks taken from previous images and digitally stored.

N	$t_{.995}$	$t_{.95}$	$t_{.90}$
4	0.99	0.90	0.80
5	0.96	0.81	0.69
6	0.92	0.73	0.61
7	0.87	0.67	0.55
8	0.84	0.62	0.51
9	0.80	0.58	0.47
10	0.77	0.55	0.44
15	0.64	0.44	0.35
20	0.56	0.38	0.30
30	0.46	0.31	0.24
40	0.40	0.26	0.21

Table 1

Ho (1985) and Wong and Ho (1986) use cross correlation in combination with epipolar geometry to test the suitability of CCD-camera's and digital matching for close range application. A second order approximation of the correlation function is used to evaluate the quality of the match, the more narrow the correlation function the better the match. A weighted scheme is used to become an approximate value for the parallax of the next point.
The geometric differences between the two image patches are tackled by Nevatia (1976) and Tsai (1983) by a multiframe approach, using a number of progressive, closely spaced views. Disparities between extreme views are determined by chaining through the intermediate views. The reliability increases, at the cost of (although the search space can be kept small) augmented computational effort.

3.2 Least Squares Matching (LSM)

We will develope here the model for a shift:

$$x_s = x_t + dx; \quad y_s = y_t + dy.$$

Let $g_t = g_t(x_t, y_t)$ be, the grey value function of the target area and $g_s = g_s(x_s, y_s)$ the grey value function of the search area. It is assumed that the noise is additive:

$$g_t(x_t,y_t) + n_t(x_t,y_t) = g_s(x_t + dx, y_t + dy) + n_s(x_t + dx, y_t + dy)$$

denoting:

$$n(x_t, y_t) = n_t(x_t, y_t) - n_s(x_t + dx, y_t + dy)$$

the above equation becomes:

$$g_t(x_t, y_t) + n(x_t, y_t) = g_s(x_t + dx, y_t + dy).$$

The equation is obviously non-linear. To solve the equation for (dx, dy), approximate values (dx_o, dy_o) are necessary

$$dx_o = dx - \Delta x; \qquad dy_o = dy - \Delta y;$$

leading to:

$$\Delta g + n = \frac{dg_s}{dx} \cdot \Delta x + \frac{dg_s}{dy} \cdot \Delta y$$

with Δg the difference between the grey values of the target and search area.

Denoting the partitial derivatives

$$\left(\frac{dg_s}{dx}, \frac{dg_s}{dy} \right)$$

(i.e. gradients) in x- and y- direction by (g_x, g_y): Δx and Δy can be computed from a least squares adjustment:

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} \Sigma g_x^2 & \Sigma g_x g_y \\ \Sigma g_x g_y & \Sigma g^2_y \end{pmatrix}^{-1} \begin{pmatrix} \Sigma g_x \, \Delta g \\ \Sigma g_y \, \Delta g \end{pmatrix}$$

Besides shifts, other geometric differences between target and search area may be introduced, and also radiometric differences. Commonly an affine transformation is used:

$$x_t = a_o + a_1 x_s + a_2 y_s$$

$$y_t = b_o + b_1 x_s + b_2 y_s$$

The radiometric differences are often assumed to be linear:

$$g_t = k_o + k_1 g_s$$

Once the transformation parameters are calculated, a representative point in the target area has to be selected and transferred into the search area. The centre coordinates of the mid-pixel of the target window may be appropriate. Pertl (1985) takes the centre of gravity, weighted by the squared gradients:

$$\bar{x}_t = \frac{\sum_i x_{t,i} \, g^2_{x,i}}{\sum_i g^2_{x,i}}; \qquad \bar{y}_t = \frac{\sum_i y_{t,i} \, g^2_{y,i}}{\sum_i g^2_{y,i}}.$$

The precision of signal matching is determined by (Förstner, 1982; Förstner and Pertl, 1986):

- σ_o^2, image noise variance;
- n, the number of involved pixels;
- σ^2_{gx}, σ^2_{gy}, $\sigma^2_{gx\,gy}$, the variances and co-variances of the gradients in x- and y-direction.

An estimate:

$\hat{\sigma}^2_o$ of σ^2_o

can be calculated from the residuals n:

$\sigma^2_o = \sum_i v^2/r$, with r the redundance.

The variances of σx and σy, the positional precision, is given by (Förstner, 1982):

$$\sigma_x^2 = \frac{1}{n} \frac{\sigma_o^2}{\sigma_{gx}^2}; \quad \sigma_y^2 = \frac{1}{n} \frac{\sigma_o^2}{\sigma_{gy}^2}$$

In case the signal to noise ratio is known:

$SNR = \sigma_g / \sigma_o$, with σ_g

the signal variance, the above may be substituted by:

$$\sigma_x^2 = \frac{1}{n\ SNR^2} \frac{\sigma_g^2}{\sigma_{g_x}^2}$$

$$\sigma_y^2 = \frac{1}{n\ SNR^2} \frac{\sigma_g^2}{\sigma_{g_y}^2}$$

In areas with a smooth grey value function, i.e. low texture and little edges, both $\sigma^2 g_x$ and $\sigma^2 g_y$ will be small and consequently σ^2_x and σ^2_y large, which corresponds to the intuitive notion.

Rosenholm (1986) presents a method, called multi-point matching, to achieve matches in homogeneous regions. The method is combined with a bilinear transformation to achieve parallaxes at a regular grid. This method is also appropriate to handle surface discontinuities.

Rosenholm (1987) investigates the effect of windowsize on precision and reliability. Windowsizes of 20 x 20 and 30 x 30 are optimal for precision. Optimal reliability requires larger windows. LSM methods often use the epipolar geometry. LSM can reach precisions up to 1/20 pixelsize. The disadvantages of LSM are:

1. it needs a time-consuming resampling;
2. it needs very accurate approximate values.

So far known, Wrobel 1987[a,b] is the only one who really attempts to describe the surface reflectivity, introducing an optical density function. The mathematical formulation of the 'facet stereo vision' is rather complex, and to obtain a tractable method, simplifications are inevitable.

Discussion

All methods based on signal matching suffer from the following limitations:

- the presents of detectable texture or edges is required; in areas with a smooth grey value function no optimal match will appear;
- repetitive micro structures will cause several equally likely matches;
- linear edges will cause many pronounced matches along the edges;
- surface discontinuities can't be handled;
- the target area may have no counterpart in the search space because of e.g. occlusion;
- they are computationally expensive;
- they are not rotational and scale invariant;

Besides, LSM requires very accurate approximate values. Ambiguities in the match result caused by low structure contents and linear edges may be avoided by examining the target area in advance. For instance, the presence of pronounced texture or corners indicate suitable targets. Hannah (1974) examines the autocorrelation, which should be steep in all directions, of the target prior to correlation. The autocorrelation enables also the setting of a bound on the correlation measure. A bad correlation value will also indicate missing counterparts. Repititive structures will cause a sequence of good correlation values.

Moravec (1977) selects target areas by examining the grey value variances in the four main directions. A target is suitable for correlation if the variances are high in all four directions. Barnard and Thompson (1980) adopted this operator as interest operator for feature matching. Therefore we save its further elaboration for the next section.

Computation time may be reduced by:

- Limitation of the search space by:
  - exertion of the epipolar geometry; the 2-D search space becomes 1-D, both images are rectified to the normal case, the rows become epipolar lines;
  - general knowledge about the object space (e.g. surfaces are smooth);
  - prediction of the next point from previous matches;
- Coarse-fine correlation by a multi resolution approach, matches at higher levels, guide match examination at lower levels;
- Utilisation of less costly similarity measures, like sum of the absolute values of differences.

4 Feature Matching

Contrary to some signal matching procedures, features are detected in both images, leading to the primal sketch. Features can be points, lines and areas. Mainly points in combination with the epipolar geometry is applied. Several characteristic point detectors, i.e. interest operators are developed. We will treat (1) the Marr-Hildreth, (2) Moravec, (3) Dreschler and (4) Förstner operator. For lines and shapes, the basic operators are edge detectors combined with line-following and vectorization methods.

Some criteria, important for the particular selection of feature detectors are (Dreschler-Fischer, 1987):

- Detection: definition of the features which are relevant for the actual correspondence analysis and the statement of their number;
- Localization: definition of the positional precision;
- Attributes: definition of the attributes suitable for matching;
- Robustness: the noise and geometric and radiometric distortion tolerance.

To select corresponding features, first an upper bound is set on the parallaxes to reduce search space. Next, similarity measures are determined and considered as initial weights or costs, depending on further approach. For instance, a 5x5 window around each point feature may be used to compute the cross-correlation with possible counterparts, and its value is used as likelihood measure. Similarity check will generally not lead to unique matches. Additional techniques are needed. the most succesfull techniques are based on relaxation, minimal path computation using dynamic programming, robust statistics and simulated annealing. For line and shape matching other techniques are used.

There are three properties of image pairs which can strongly influence matching (Barnard and Thompson, 1980):

- Discreteness is a property of individual points, giving a measure for the distinction of the point with its neighbourhood;
- Similarity gives a measures of the resemblance of two points;
- Consistency a measure for the conformity of a particular match with surrounding matches, assu-

ming some general object model, e.g. the object surface varies only smooth attended by a limited number of surface discontinuities or the surface is a tilted plane.

The above properties lead to the following three stages in feature matching:

- selection of distinct features by an interest operator (distinction check);
- selection of candidate features which may form possible matches, using one or more similarity measures (similarity check);
- thinning of the list of candidate points, untill unique matches remain, consistent with an object model, i.e. determination of the correct matches (consistency check).

We will first treat the interest operators and similarity measures for point matching. Next the consistency techniques are considered. The last part of this section describes line and shape matching but no special attention will be set on edge detection and line following, neither on vectorization and shape description.

4.1 Point matching

4.1.1 Distinction and similarity check

The environment of an interesting point is characterized by a steep autocorrelation function in all directions, high variances in all directions and steep gradients in all directions. These properties leads to as many as approaches. Autocorrelation considerations are used by Hannah (1974) for signal matching purposes. For feature matching it doesn't seem to be suited (Dreschler, 1981). The statistical variance view leads to the approaches of Moravec (1977) and Förstner (1986). The gradient view has led to the surface curvature examination of Dreschler (1981) and Dreschler and Nagel (1982). Requirements characteristic points should fulfil are (Förstner, 1986):

- Discreteness: the points must be different from neighbouring points;
- Invariance: both the selection and the localization of the points should be unaffected by geometric and radiometric distortions;
- Stability: the point must have a high probability to appear in both images;
- Seldomness: in order to achieve reliable results features derived from repetitive structures should be avoided.

4.1.1.1 Marr-Hildreth operator

Based on the human visual system Marr and Hildreth (1980) have developed an operator to detect grey value changes. It is the second derivative of a Gaussian, in particular the Laplacian of a Gaussian:

$$\nabla^2 S(x,y) * G(x,y)$$

with S(x,y) the 2-D normal distribution in 1-D form given by:

$$S(x) = \frac{1}{\sigma (2\pi)^{\frac{1}{2}}} e^{-x^2/2\sigma^2}$$

and ∇^2. the Laplace operator, executed on the image G it gives:

$$\nabla^2 G = \frac{d^2G}{dx^2} + \frac{d^2G}{dy^2} = g_{xx} + g_{yy}$$

A Gaussian can be approximated by a repeated convolution of a 2x2 unweighted smoothing filter, e.g.:

$$\frac{1}{64}\begin{bmatrix} 1 & 3 & 3 & 1 \\ 3 & 9 & 9 & 3 \\ 3 & 9 & 9 & 3 \\ 1 & 3 & 3 & 1 \end{bmatrix} = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} * \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} * \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} * \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Defining g_x and g_y as the normal gradient images (see for masks table 2.), than the Laplace operator , $g_{xx} + g_{xx}$ is given by:

$$\begin{bmatrix} 1 & -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Convolution of the Gaussian with the Laplacian yields $\nabla^2 S$. The general shape of this operator is illustrated in fig. 6. At the point of inflexion of a curve the Laplace operator will give the response 0. In general on a raster the point of inflexion isn't found directly. It is surrounded by a positive and a negative operator response. The zero-crossing can be found by a linear interpolation. The attributes for similarity test are sign and orientation of the line, the zero-crossing is part of. The sign is determined by the sign of the first operator response (positive or negative) when moving along an epipolar line.

4.1.1.2. Moravec operator

In order to check whether variances are high in all directions, samples in some directions are chosen. In particular the four main directions, defined by rows and columns and the two diagonals are evaluated. The variance in each direction is computed as the sum of squared grey value differences between neighbouring pixels. The original Moravec is defined for a 5x5 window:

$$M_1(i,j) = \frac{1}{20}\sum_{k=-2}^{2}\sum_{l=-2}^{1}\left(g(i+k,j+l) - g(i+k,\ j+l+1)\right)^2$$

$$M_2(i,j) = \frac{1}{20}\sum_{k=-2}^{1}\sum_{l=-2}^{2}\left(g(i+k,j+l) - g(i+k+1,j+l)\right)^2$$

$$M_3(i,j) = \frac{1}{16}\sum_{k=-2}^{1}\sum_{l=-2}^{1}\left((g(i+k,j+l) - g(i+k+1,j+l+1)\right)^2$$

$$M_4(i,j) = \frac{1}{16}\sum_{k=-2}^{1}\sum_{l=-2}^{1}\left(g(i+k,j+l+1) - g(i+k+1,j+l)\right)^2$$

The operator response M for single grey value images i defined by:

$$M = \min(M_i),\ i = 1,..,4.$$

If M exceeds a certain threshold M_t then the point (i,j) is excepted as characteristic point. Characteristic points will cause a series of operator responses above the threshold. To achieve distinct points non-maximum suppression has to be implemented.

For colour images (multispectral data) the operator can be modified in two manners (Dreschler-Fischer 1987):

- The single grey value Moravec operator is applied to the distinct bands, M_k, k=1,..n. The colour response is the maximum of the responses in the distinct bands, i.e.: $M_c = \max(M_k)$;
- The directional variances may also be computed for the spectral band vectors. The squared differences of the vectors are calculated. Like in the single value approach, the response should exceed a threshold. In case of visible bands (blue, green and red) this colour-vector operator has proven to be significantly better than the other Moravec operators (Dreschler Fischer, 1987).

The Moravec operator is easy to implement, but its ad hoc character has several drawbacks (Dreschler, 1981):

- not the real corner is found, but a shift is introduced, the larger the size of the operator, the larger the shift;
- it is sensitive to low resolution features, i.e. small points cause an extended operator signal;
- the operator is non-rotational invariant.

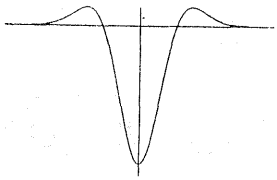

Fig. 6

Barnard and Thompson (1980) use as similarity measure between characteristic points of conjugate images the sum of the squares of the differences of the surroundig windows.

4.1.1.3 Dreschler operater

The grey value function may be looked at as a curved plane, much in the same way as a digital elevation model. From differential geometry it is known that the Gaussian curvature: $K = k_1 \times k_2$ is invariant against geometric transformation and such a measure is very appropriate for stereo matching. The general equation of a smooth, i.e. second differentiable, surface is represented by:

$$r(u,v) = (x(u,v), y(u,v), z(u,v))$$

with (x,y,z) Carthesian coordinates and (u,v) surface coordinates. The explicite expression of a smooth surface g = f(x,y) becomes in parameter description, with $(x,y) \leftrightarrow (u,v)$:

$$r(u,v) = (u, v, g(u,v))$$

The second partial derivatives of r(u,v) are given by:

$$r_{uu} = (0, 0, g_{uu})$$
$$r_{uv} = (0, 0, g_{uv})$$
$$r_{vv} = (0, 0, g_{vv})$$

The curvature information is now enclosed in the symmetric tensor:

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

with:

$$L = g_{uu}/C$$
$$M = g_{uv}/C$$
$$N = g_{vv}/C$$
$$C = \sqrt{1 + g_u^2 + g_v^2}$$

Since, only the third coordinate of r_{uu}, r_{uv} and r_{vv} differs from 0 and just the kind of curvature and not its absolute value is of interest, the principal curvatures are found from an eigenvalue analysis of the symmetric matrix:

$$\begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix}$$

The principle curvatures k_1 and k_2 ($k_1 > k_2$) become:

$$k_{1,2} = \tfrac{1}{2} \left(g_{xx} + g_{yy} \pm \sqrt{(g_{xx} + g_{yy})^2 - 4 (g_{xx} g_{yy} - g_{xy}^2)} \right)$$

The principal direction D is given by:

$$D = \tfrac{1}{2} \operatorname{atan} (2g_{xx} / (g_{xx} - g_{yy}))$$

The attributes for similarity check are the signs of k_1 and k_2, e.g. k_1 is positive and k_2 is negative indicates a saddle point.

4.1.1.4 Förstner operator

Förstner (1986) describes an operator, which evaluates the covariance matrix C of the gradient images g_x and g_y of a neighbourhood, e.g. a 5x5 window:

$$C = \begin{pmatrix} \Sigma g_x^2 & \Sigma g_x g_y \\ \Sigma g_x g_y & \Sigma g_y^2 \end{pmatrix}$$

g_x and g_y may be computed from the normal gradient, the Roberts gradient or one of the other operators in table 2. The distinct advantage of this approach is that the covariance matrix of the gradients determines the precision of the match, i.e. features can be a priori selected on their suitability to give precise matches. To avoid edges, where the match isn't defined in the direction along the edge, the error ellipse must be nearly a circle. Further the error ellipse should be small.

The eigenvalues λ_1 and λ_2, $\lambda_1 > \lambda_2$, determine the shape of the ellipse, e.g. the elongatedness is given by $E = (\lambda_1 / \lambda_2)^{\frac{1}{2}}$. The eigenvalue computation can be avoided by taking a direct measure:

$$q = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 = \frac{4 \lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2}$$

Since:

$$\lambda_1 \lambda_2 = C_{xx} C_{yy} - C_{xy}^2$$

and:

$$\lambda_1 + \lambda_2 = C_{xx} + C_{yy}$$

$$q = 4 \det C / tr^2 C$$

The size of the ellipse is computed from:

$$w = \det C / tr\, C$$

The feature can be computed up to subpixel level by using the centre of gravity. Extracted points in an artificial stereo pair found by this operator are shown in fig. 7. Contrary to the Moravec operator, the Förstner operator is rotational invariant. It has also shown very nice properties for selecting other features, like edges and circles (Förstner and Gülch, 1987). As similarity measure the cross-correlation between small windows around the points are used.

4.1.2 Consistency Check

The interest operator together with the similarity check will lead to an initial set of possible matches. The list will contain many ambiguities, because of imperfections in operators and similarity measures and because e.g. occlusion and shadows will bring about that points are just detected in one image. A consistency check has to be undertaken using an object model. Commonly, the epipolar constraint is employed to reduce search space. Ullman (1979) reformulates his surface model of minimal mapping as a linear programming problem, i.e. a path of minimal costs is searched. Commonly, the costs are defined reciprocal to the similarity measures. Many authors have adopted the idea of minimal cost path search, mostly combined with concepts of dynamic programming (DP), developed by Bellman (1957) and extensively expounded in (Kaufmann, 1967). In photogrammetric stereo matching the DP approach is explored by (Benard et. al., 1986) and (Kölbl et. al., 1987). A further refinement is to constraint the solution by edge connectivity across the epipolar lines, e.g. Lloyd et. al. (1987) and Ohta and Kanade (1985) use DP combined with consistency constraints defined by vertically connected edges. Ohta and Kanade (1985) use both DP for the search along epipolar lines (intra-scanline search) and across the epipolar lines (inter-scanline search). Lloyd et. al. (1987) employ DP to produce candidate matches along epipolar lines. The solution is constrained by relaxation labelling using the connectivity of edges.

Relaxation labelling iteratively updates an initial probability by an amount proportial to the estimate of its consistency with the labelling over its neighbourhood. A labelling consistent with neighbouring allotments is remunerated with an increased probability. The process is repeated until a steady state is achieved. The relaxation approach is investigated by Barnard and Thompson (1980), using a smooth object model, i.e. the dispartities in a neighbourhood don't vary abruptly. The problem with this approach is that some labels may be left unmatched whereas others may be double or even more matched. This is caused by the fact that the assignment of image 1 to image 2 will not give the same results as when the assignement is performed in reverse direction. Dreschler (1981) signalizes this problem and modifies the Barnard and Thompson approach by introducing a symmetric assignment. Relaxation labelling is also used when the features are shapes. A brief view is given in the next section.

Förstner (1986) introduces the concepts of robust statistics, like they can be found in (Hampel et. al., 1986) to tackle the consistency constraint. The object model is a plane. As initial weights of the consistency of the match between two points cross-correlation is employed. Using an influence function, weights are decreased when matches show large residuals in the least squares approach. In an interative procedure a steady state is achieved. In fig. 7 the correspondening points found by this procedure are indicated.

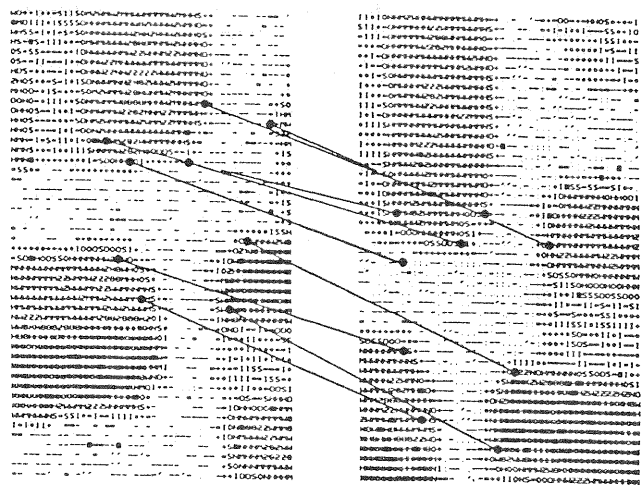

Fig. 7 Artificial stereo pair. The isolated points are extracted with the Förstner operator. Correspondence analysis is performed with robust estimation, assuming a tilted plane as object model (adopted from Förstner, 1986)

Another approach, which has a physical background, is the one of Barnard (1986, 1987). It is an interative method too, based on annealing and therefore called simulated annealing. It is employed to overcome the disadvantages of relaxation, which are (Anily and Federgruen, 1987):

- the final solution is heavely dependent on the starting point;
- the solutions tend to relax in local optima.

Simulated annealing methods attempt to avoid these problems by randomizing the procedure, at the cost, however, of much more computation time.

4.2 Line and Shape Matching

The segmentation techniques, like edge detection, line following and region growing, to detect shapes, are leaved out of consideration. The masks of some common edge detectors are just listed in table 2. The Förstner operator is also suited for edge detection; an edge is indicated by a small roundness measure q. Since line segements are the elementary parts of shapes, we will view line matching as part of shape matching.

Shape matching is tackled by many approaches. One of the earliest procedures is to decompose shapes in a chaincode. If $a_1,..,a_n$ is the chain code of the one curve and $b_1,..,b_m$ that of the other, with $n > m$, than the chain correlation function C(j), is given by:

$$C(j) = \frac{1}{m} \sum_{i=1}^{m} \cos\left((b_i - a_{i+j}) \bmod 8\ \pi/4 \right)$$

C(j) = 1 if both curves match exactly. Although computationally very efficient (O(n,m) operations), chain

code correlation is of limited value, since:
- it is not rotation and scale invariant;
- it is sensitive to noise.

Many shape matching techniques are based on the view that shapes can be represented by a set of measures, defined by the particular occurance of the shape, and matched against each other using statistical methods. To the variety of features, which may describe a shape, belong: elongatedness, compactness, perimeter, area, moments and Fourier descriptors.

Another approach is that shape boundaries are represented by a polygonal approximation. The length of the segements depend on the curvature. Davis (1979) uses the angles between adjacent segments as basic descriptors, i.e. a shape is viewed as a sequence of angles. The similarity between pairs of angles on the two shapes are evaluated and indicated by figures of merit. Although the problem is formulated as an optimal path search, i.e. the most suitable approach would be dynamic programming, a relaxation method is applied to select the best matches, because of the computational costs associated with dynamic programming.

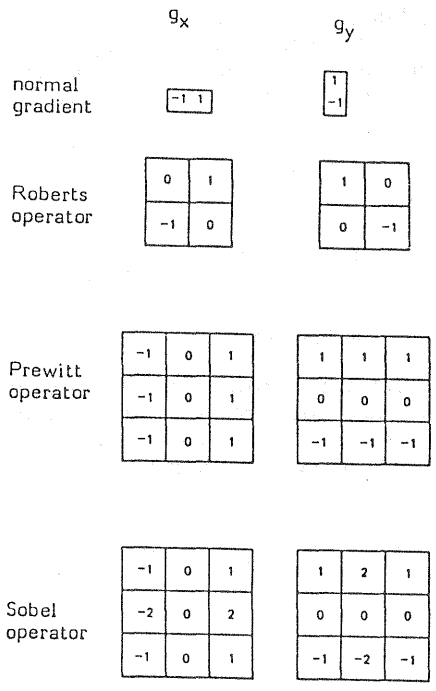

Table 2

Medioni and Nevatia (1984) develop a shape matching method based on a graph representation using geometrical descriptors. These descriptors are:
- coordinates of the end points;
- orientation;
- width,

and refer to segments, i.e. groups of connected line features. Additionally, constrast is used. The segmentation is carried out by first detecting local edges, using step edge masks in various orientations, and next thinning and linking the edges and fitting the curves by piece-wise linear segments. Matching is performed by discrete relaxation, i.e. the geometrical descriptions which give the best resemblance are searched in an iterative process.

For time varying images, Costabile et. al (1985) develop a method of shape matching using a graph description combined with a tree search. Once the boundary of an area is approximated by linear features, the polygon is decomposed into convex parts. To each convex part, two attributes are assigned:
1 the arc-to-chord ratio;
2 the area between arc and chord.

The arc is defined by the sequence of lines between the start and end point of a convex part and the cord by the line between start and end point. The correspondence search is carried out by a tree search.

5 Relational Matching

Relational matching takes, besides the descriptive attributes of phenomena, also relations between the phenomena into account. So, a set of elementary tokens, like points, blobs, line fragments and regions, are detected, characterized by attributes, like length, area, shape and average grey value, and assigned to each other. To these relationships, which describe only the spatial connextion, also attributes can be assigned describing their properties (Shapiro and Haralick, 1987). The primitives form an entity. It is the task to find a mapping of the primitives of the one entity to those of the second entity, that best perceives the characteristics and relationships. The basis of entity description is found in graph theory. The description yields long lists. With the aid of searching algorithms, like backtracking tree search or one of its variants, the best mapping of the one entity onto the other is determined. Because of the different view points of stereo images, the same object structure has different graphs in both images. So, the matching procedures must tolerate these differences, leading to inexact matching techniques. Shapiro and Haralick (1981) formulate the concepts of exact and inexact matching of graph representation of structures. Because relational matching of complex scenes is still in its infancy and subject of current research, we will not elaborate it here. A basic introduction to relational matching can be found in Ballard and Brown (1982). More advanced are the expositions of Shapiro and Haralick (1987) and Boyer and Kak (1988).

6 Conclusions

A survey on stereo image matching techniques is presented. Ideally we would like to match each pixel in the one image with a pixel in the other image to arrive at a dense surface description. But the individual grey values or vector of grey values don't contain enough information to arrive at reliable matches, because noise, illumination differences and so on, will cause ambiguities. Using a neighbourhood of the pixels can limit the ambiguity, leading to signal matching, but ambiguities can't be avoided entirely, because of geometric differences and occlusions.

Methods, highly proof against image differences, operate on a high level, i.e. pattern recognition is performed prior to matching. The thematic phenomena, like houses and roads, and their interrelationships are matched. But such a relational performance requires extraction of many objects in complex scenes, which is beyond the present state of computer vision. Besides, without further refinement techniques no high geometric precision will be achieved. A technique in between signal and relational matching is the extraction of radiometric independent phenomena, like points and linear features, solving the correspondence problem by matching these phenomena. A dense surface description is achieved by interpolation.

To reduce search space and to find matches more reliable, frequently a coarse to fine matching strategy is applied. But the three matching methods may be viewed in an hierarchical way, too. Relational matching, to find rough (i.e. global) matches, next feature matching will give precise (i.e. local) matches. For high precision measurements these feature matches may be considered as approximate values for signal matching. Signal matching and feature matching have already found their way to photogrammetry. Relational matching is on its way to get there.

References

Ackermann, F., 1984, Digital Image Correlation: Performance and Potential Application in Photogrammetry, Photogrammetric Record, vol. 11(64), October, pp. 429-439.

Almroth, U., Hendriks, L., 1987, High accuracy digital matching of close-range objects on the analytical plotter Kern DSR11, Proc. ISPRS Intercomm. Conf. on fast processing of photogrammetric data, Interlaken, pp 193-203.

Anily, Federgruen, 1987, Simulated annealing methods, Journal of applied probability, vol. 28, nr. 3. pp. 657-666.

Ballard, D.H., Brown, Ch.M., 1982, Computer Vision, Prentice Hall, Inc., Englewood Cliffs, New Jersey.

Barnard, S.T., 1986, A stochastic approach to stereo vision, SRI International, Technical note 373.

Barnard, S.T., 1987, Stereo matching by hierarchical microcanonical annealing, SRI International, Technical note 414.

Barnard, S.T. Thompson, W.B., 1980, Disparity Analysis of Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-2, no. 4, July, pp. 333-340.

Bellman, R., 1957, Dynamic Programming, Princeton Univ. press.

Benard, M., Boutaleb, A.K., Kölbl, O., Penis, C., 1986, Automatic stereophotogrammetry: implementation and comparison of classical correlation methods and dynamic programming based techniques, ISPRS Comm. III, Rovaniemi.

Bethel, J., 1986, The DSR11 Image Correlator, Technical papers, ACSM-ASPRS, Annual Convention, vol. 4, Washington D.C., U.S.A, pp. 44-49.

Billingsley, F.C. 1983, Data processing and reprocessing, Manual of Remote Sensing, ch. 17, pp. 719-792.

Boyer, K.L., Kak, A.C., 1988, Structural stereopsis for 3-D vision, IEEE Trans. on PAMI, vol. 10, nr. 2, pp. 144-166.

Claus, M. 1983, Korrelationsrechnung in Stereobildpaaren zur Automatischen Gewinnung von digitalen Geländemodellen, Orthophotos und Höhenlinienplänen, Ph.D. Thesis, University of München.

Costabile, M.F., Guerra, C., Pieroni, G.G., 1985, Matching Shapes: A Case Study in Time-Varying Images, Computer Vision, Graphics, and Image Processing vol. 29, pp. 296-310.

Davis, L.S., 1979, Shape Matching Using Relaxation Techniques, IEEE Trans. on PAMI, vol. 1, no. 1, pp. 60-72.

Dreschler, L., 1981, Ermittlung markanter Punkte auf der Bildern bewegter Objekte und Berechnung einer 3-D Beschreibung auf dieser Grundlage, Ph.D. Thesis, University of Hamburg.

Dreschler-Fischer, L.S., 1987, A blackboard system for dynamic stereo matching, in: Hertzberger, L.O., Groen, F.C.A. (edt.), Intelligent Autonomous Systems, Elsevier Science Publishers B.V. Amersterdam, pp. 189-202.

Dreschler, L., Nagel, H.-H., 1982, Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene, Computer Graphics and Image Processing, vol. 20, pp. 199-228.

Ehlers, M., 1983, Untersuchung von digitalen Korrelationsverfahren zur Entzerrung von Fernerkundungsaufnahmen, Ph.D. Thesis, University of Hannover.

Förstner, W. 1982, On the geometric precision of digital correlation, ISPRS Int. Arch. of Photogrammetry, vol. XXIV, Comm. III, Helsinki, pp. 176-189.

Förstner, W., 1984, Quality assessment of object location and point transfer using digital image correlation techniques, invited paper to Commission III, 15th ISPRS Congress, Rio de Janeiro, pp. 169-191.

Förstner, W., 1986, A Feature Based Correspondence Algorithm for Image Matching, Int. Arch. of Photogr., vol. 26-III, Rovaniemi, pp. 1-17.

Förstner, W., Pertl, A., 1986, Photogrammetric standard methods and digital image matching techniques for high precision surface measurements, Pattern Recognition in Practice II, pp. 57-72.

Förstner, W., Gülch, E., 1987, A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features, Proc. ISPRS Intercomm. Conf. on fast processing of photogrammetric data, Interlaken, pp. 281-305.

Grimson, W.E.L., 1981, From images to surfaces: a computational study of the human early visual system, MIT press.

Grün, A.W., Baltsavias, E.P., 1987, High-Precision Image Matching for Digital Terrain Model Generation, Photogrammetria (PRS), vol. 42, pp. 97-112.

Hampel, F.R., Ronchetti, E.M. Rousseeuw, P.J. Stahel, W.A., 1986, Robust statistics, John Wiley & Sons.

Hannah, M., 1974, Computer matching of areas in stereo images, Ph.D. Thesis, University of Stanford.

Hendriks, L.F.G.M., 1988, Investigation into the correlation algorithm of the analytical plotter DSR11, afstudeerscriptie, Faculty of Geodesy, Delft University of Technology.

Ho, W.-H., 1984, The potential of a low resolution digital camera in close-range photogrammetry, Ph.D. Thesis, University of Illinois.

Kaufmann, A., 1967, Graphs, dynamic programming, and finite games, Academic Press, New York-London.

Kölbl, O., Boutaleb, A.K., Penis, C., 1987 A concept for automatic derivation of a digital terrain model with the Kern DSR11, Proc. ISPRS Intercomm. Conf. on fast processing of photogrammetric data, Interlaken, pp. 306-317.

Lloyd, S.A., Haddow, E.R., Boyce, J.F., 1987, A Parallel Binocular Stereo Algorithm Utilizing Dynamic Programming and Relaxation Labelling, Computer Vision, Graphics, and Image Processing 39, pp. 202-225.

Marr, D., 1979, Vision, Freeman and Compagny, San Francisco.

Marr, D., Hildreth, E., 1980, Theory of edge detection, Proc. R. Soc. Lond. B. Vol. 207, pp.187-217.

Marr, D., Poggio, T., 1979, A theory of human stereo vision, Proc. R. Soc. Lond. B. 204, pp. 301-328.

Medioni, G., Nevatia, R., 1984, Matching Images Using Linear Features, IEEE Trans. on PAMI, vol. 6, no. 6, pp. 675-685.

Medioni, G., Nevatia, R. 1985, Segment-based stereo matching, Computer vision, Graphics and Image Processing, vol. 31, pp. 2-18.

Moravec, H.P., 1977, Towards automatic visual obstacle avoidance, Proc. 5th Joint Conf. Art. Intell., Cambridge, p. 584.

Nevatia, R., 1976, Depth measurement by motion stereo, Computer Graphics and Image Processing, vol. 5, pp. 203-214.

Ohta, Y., Kanade, T., 1985, stereo by intra-and Inter-scanline search unsing
dynamic programming, IEEE Trans. on PAMI, vol. 7, No. 2, pp. 139-154.

Pertl, A. 1984, Digital image correlation with the analytical plotter Planicomp C100, Int. Arch. of Photogr., 25, III, pp. 874-881.

Pertl, A., 1985, Digital image correlation with an analytical plotter, Photogrammetria, vol. 40, pp. 9-19.

Rosenholm, D. 1986, Accuracy improvement of digital matching for evaluation of digital terrain models, Int. Arch. of Photogr., 26, 3, pp. 573-587.

Rosenholm, D., 1987, Empirical Investigation of Optimal Window Size Using the Least Squares Image Matching Method, Photogrammetria (PRS), vol. 42, pp. 113-125.

Shapiro, L., Haralick, R.M., 1981, Structural descriptions and inexact matching, IEEE Trans. on PAMI, vol 3, no. 5.

Shapiro, L.G., Haralick, R.M., 1987, Relational Matching, Applied Optics, vol. 26, pp. 1845-1851.

Tsai, R.Y., 1983, Multiframe image point matching and 3-D surface reconstruction, IEEE Trans. on PAMI, vol. 5, no. 2.

Ullman, S., 1979, The interpretation of visual motion, MIT press.

Wiesel, W.J. 1981, Passpunktbestimmung und geometrische Genauigkeit bei der relativen Entzerrung von Abtastdaten, Ph. D. Thesis, DGk, vol. C., nr. 268.

Wong, K.W., Ho, W.H., 1986, Close-Range Mapping with a Solid State Camera, Photogrammetric Engineering and Remote Sensing, vol. 52, no. 1, January, pp. 67-74.

Wrobel, B.P., 1987[a] Facets Stereo Vision (FAST Vision) - A new approach to computer stereo vision and to digital photogrammetry, Proc. ISPRS Intercomm. Conf. on fast processing of photogrammetric data, Interlaken, pp. 231-258.

Wrobel, B.P. 1987[b], Einige Überlegungen über die theoretischen Grundlagen der digitalen Photogrammetrie, Bildmessung und Luftbildwesen, vol. 55, pp. 129-140.