Geometrically Constrained Multiphoto Matching

Armin W. Gruen and *Emmanuel P. Baltsavias*
Institute of Geodesy and Photogrammetry, Federal Institute of Technology, ETH-Hoenggerberg, CH-8093 Zurich, Switzerland

ABSTRACT: The Adaptive Least-Squares Correlation, combining gray level matching with geometrical constraints, is applied for *X, Y, Z* object coordinate determination. The constraints used are the collinearity conditions. A new aspect is the simultaneous use of multiple (more than two) scenes. This paper outlines the mathematical model and highlights some essential features of the algorithm with practical data.

Tests using CCD camera data in a close-range environment were performed on the aspects of pull-in range, occlusions, and reliability (multiple solutions, mismatch). In all cases remarkable advantages result from the use of geometrical constraints (conditional one-dimensional search) and multiple scenes. Depth errors of 5 ‰ average depth (d_o) (6 pixels pull-in range) and 10 ‰ d_o (12 pixels pull-in range) were examined, with 100 percent and 70 percent success rate, respectively. In case of occlusions the success rate is increased and the occluded image patches can be labeled in most cases. Multiple solutions and the danger of undetected mismatches are considerably reduced.

INTRODUCTION

THE LIMITS of stereovision systems, as used for three-dimensional object-space reconstruction, are well-known. The small redundancy of the point positioning phase leads to low reliability of the object point coordinates, the precision suffers very often from bad geometrical conditions, there is only little control of mismatches in case of repetitive object patterns, and occlusions may cause the loss of object points. Little effort has been made in computer vision and robot vision to overcome this situation. This paper offers a solution to the problems.

In Gruen and Baltsavias (1985), the method of Adaptive Least-Squares Correlation was outlined and extended towards a simultaneous matching/point positioning system by incorporating geometrical constraints which are defined by the sensor imaging geometry and the kind of knowledge available with respect to the sensor orientation parameters and different object space elements (such as points, planes, distances, parallel and straight lines, etc.).

This paper focuses on the problem of multiphoto matching with the collinearity conditions as additional constraints for *X, Y, Z* determination. The mathematical model is presented and some features of this method are outlined using practical data. This mathematical model can always be applied when the interior and exterior orientations of the various scenes are known and when the sensor geometry is defined by the collinearity condition. Deviations from the strict geometry, caused by systematic errors resulting from lens distortion, video signal generation, etc., are also incorporated. Particular emphasis in the practical tests is put on the aspects of reliability (mismatch, multiple solutions), goodness of approximate values, and occlusions.

The practical experiences are based on a few data sets only; they highlight some aspects, but do not approach the problem in a fully comprehensive and systematic manner.

THE ESTIMATION MODEL

The statistical estimation model used is a combined least-squares adjustment. The observation equations consist of two parts, gray level matching and imposed geometrical constraints, whereby the two parts are related through the shift parameters of the image patches. Here a multiphoto approach is used, whereby the gray level matching equations are formulated such that they allow for simultaneous, local image shaping, for the inclusion of additional geometrical constraints as well as for discrete corrections, which are required because of the deviation of the actual sensor geometry from the theoretical model. Radiometric corrections are not part of the estimation model. These corrections are applied prior to the adjustment.

Assume a sequence of $n+1$ images of an object (Figure 1). The object is defined in a three-dimensional cartesian coordinate system (*X, Y, Z*), the images in a three-dimensional cartesian coordinate systems (x, y, $z = 0$). The images are discrete two-dimensional approximations of continuous functions. One image function $f(x,y)$ serves as the template, a reference, and the remaining n images $g_1(x,y), \ldots, g_n(x,y)$ as the pictures. An ideal situation gives

$$f(x,y) = g_i(x,y),\ i = 1,\ldots,n. \tag{1}$$

Taking into consideration noise and assuming that the template noise is independent of the picture noise, Equation 1 becomes

$$f(x,y) - e_i(x,y) = g_i(x,y), \tag{2}$$

where $e_i(x,y)$ is a true error vector.

Equation 2 can be considered as a nonlinear observation equation which models the observation vector $f(x,y)$ with functions $g_i(x,y)$ whose locations in the pictures $1,\ldots,n$ need to be estimated. The location is described by Δx_i, Δy_i shifts with respect to an approximate position of each function $g_i^o(x,y)$.

To account for systematic image deformations caused by projection and object reflection effects and to achieve a better match, additional geometric ("image shaping") transformation parameters are included.

Each function $g_i^o(x,y)$ forms a grid of gray values and is located in a larger search window $w_i(x,y)$. The image shaping is achieved by transforming the x_o, y_o coordinates of $g_i^o(x,y)$ and resampling over $w_i(x,y)$. The geometrical transformation can be modeled by a bivariate polynomial

$$x_i = \mathbf{t}_y^T \bar{\mathbf{A}}_i \mathbf{t}_x \tag{3a}$$

$$y_i = \mathbf{t}_y^T \bar{\mathbf{B}}_i \mathbf{t}_x \tag{3b}$$

with

$$\mathbf{t}_x^T = (x_o^0\ x_o^1\ x_o^2 \ldots x_o^{m-1}) \tag{4a}$$

$$\mathbf{t}_y^T = (y_o^0\ y_o^1\ y_o^2 \ldots y_o^{m-1}) \tag{4b}$$

0099-1112/88/5405–633$02.25/0

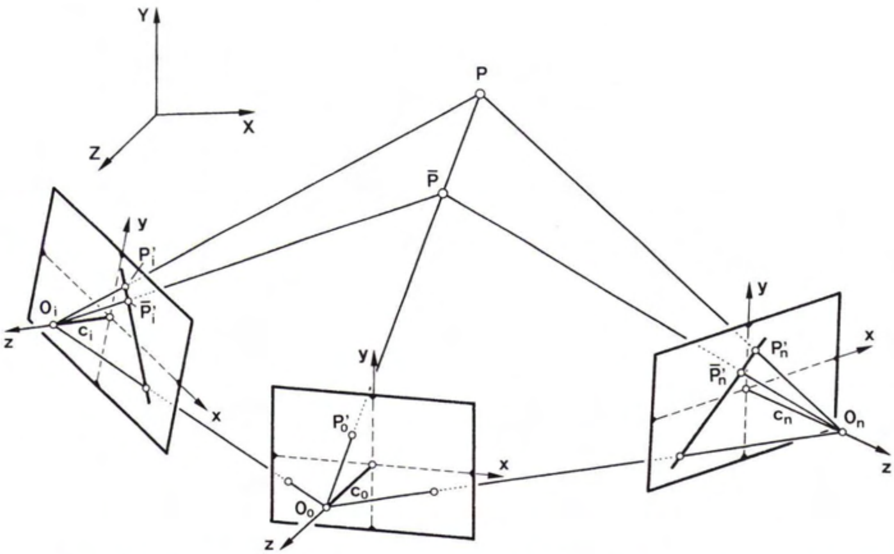


FIG. 1. Multiphoto arrangement for matching and point positioning. P. . . .correct (matched position of object point. $\bar{P}$. . . .initial position of object point.

and the parameter matrices $\bar{\mathbf{A}}_i$, $\bar{\mathbf{B}}_i$ are given by

$$\bar{\mathbf{A}}_i = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & \cdot & a_{1m} \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ a_{m1} & \cdot & \cdot & \cdot & \cdot & \cdot & a_{mm} \end{bmatrix}_i , \qquad (5)$$

$$\bar{\mathbf{B}}_i = \begin{bmatrix} b_{11} & b_{12} & \cdot & \cdot & \cdot & \cdot & b_{1m} \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ b_{m1} & \cdot & \cdot & \cdot & \cdot & \cdot & b_{mm} \end{bmatrix}_i ,$$

where x_o, y_o are the grid coordinates of the function values $g_i^o\ (x,y)$, and m-1 is the degree of the polynomial.

The transformation parameters $a_{11}, \ldots, a_{mm}, b_{11}, \ldots, b_{mm}$ need to be estimated from Equation 2. Because the image patches to be matched are usually rather small and as such generated by a very narrow bundle of rays, the perspective projection might be locally replaced by a parallel projection. Therefore, an affine transformation (six parameters) is a valid approximation for the geometrical image forming process. So the parameter set in $\bar{\mathbf{A}}_i$, $\bar{\mathbf{B}}_i$ is specified to be

$$\bar{\mathbf{A}}_i = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \end{bmatrix}_i , \quad \bar{\mathbf{B}}_i = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & 0 \end{bmatrix}_i . \qquad (6)$$

Because Equation 2 is nonlinear, it is linearized according to

$$f(x,y) - e_i(x,y) = g_i^o(x,y) + \frac{\partial g_i^o(x,y)}{\partial x_i} dx_i + \frac{\partial g_i^o(x,y)}{\partial y_i} dy_i . \qquad (7)$$

With the notations

$$g_{x_i} = \frac{\partial g_i^o(x,y)}{\partial x_i} , \ g_{y_i} = \frac{\partial g_i^o(x,y)}{\partial y_i} \qquad (8)$$

Equation 7 becomes finally

$$\begin{aligned} f(x,y) - e_i(x,y) = {} & g_i^o(x,y) + g_{x_i} da_{11i} + g_{x_i} x_o da12i \\ & + g_{x_i} y_o da_{21i} \\ & + g_{y_i} db_{11i} + g_{y_i} x_o db_{12i} + g_{y_i} y_o db_{21i} . \end{aligned} \qquad (9)$$

With

$$\mathbf{x}_i^T = (da_{11}\ da_{12}\ da_{21}\ db_{11}\ db_{12}\ db_{21})_i , \qquad (10a)$$

$$\mathbf{l}_i = f(x,y) - g_i^o(x,y), \text{ and} \qquad (10b)$$

$$\mathbf{A}_i = (g_x\ g_x x_o\ g_x y_o\ g_y\ g_y x_o\ g_y y_o)_i , \qquad (10c)$$

Equation 9 results in

$$- \mathbf{e}_i(x,y) = \mathbf{A}_i x_i - \mathbf{l}_i \ \ ; \ \ \mathbf{P}_i$$

$$\mathbf{P}_i . . . \text{weight coefficent matrix of } \mathbf{l}_i \qquad (11)$$

with the assumption

$$E(\mathbf{e}_i) = 0 \quad , \quad E(\mathbf{e}_i \mathbf{e}_i^T) = \sigma_{oi}^2 \mathbf{P}_i^{-1} . \qquad (12)$$

Equations 11 form a system of n sets of gray level correlation equations, with each set consisting of $n_1 \cdot n_2$ correlation equations (n_1, n_2. . . dimensions of pixel patch used for the match). In Equation 11, the n sets are orthogonal to each other; they do not have any joint parameter. This is equivalent to solving the n sets independently of each other using the standard least-squares technique.

If the image forming process followed the law of perspective projection, a set of $n+1$ collinearity conditions can be formulated for each imaged object point p as

$$\vec{\mathbf{x}}_{pk} = \frac{1}{\lambda_{pk}} \mathbf{R}_k^T (\vec{\mathbf{X}}_p - \vec{\mathbf{X}}_{ok}), \qquad (13)$$

with $k = 0,1,2,. . ., n$

p = object point index

$$\vec{\mathbf{x}}_{pk} = \begin{bmatrix} x_p \\ y_p \\ -c \end{bmatrix}_k ,$$

x_{pk}, y_{pk} = image coordinates of point p in scene (k), reduced to the principal point

c_k = camera constant of sensor of scene (k)

$\mathbf{R}_k = (r_1\ r_2\ r_3)_k$, $\mathbf{R}_k$ = rotation matrix of scene (k)

$$\vec{\mathbf{X}}_p = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_p \quad , \quad \vec{\mathbf{X}}_p = \text{vector of object coordinates of point } p$$

$$\vec{\mathbf{X}}_{ok} = \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix}_k \quad , \quad \vec{\mathbf{X}}_{ok} = \text{vector of object coordinates of the perspective center of scene } (k)$$

λ_{pk} = scale factor for point p imaged in scene (k)

The x,y components of Equations 13 are

$$x_{pk} = -c_k \frac{\mathbf{r}^T_{1k} (\vec{\mathbf{X}}_p - \vec{\mathbf{X}}_{ok})}{\mathbf{r}^T_{3k} (\vec{\mathbf{X}}_p - \vec{\mathbf{X}}_{ok})} \triangleq - F^x_k, \tag{14a}$$

$$y_{pk} = -c_k \frac{\mathbf{r}^T_{2k} (\vec{\mathbf{X}}_p - \vec{\mathbf{X}}_{ok})}{\mathbf{r}^T_{3k} (\vec{\mathbf{X}}_p - \vec{\mathbf{X}}_{ok})} \triangleq - F^y_k. \tag{14b}$$

For simplification, the point index p is ignored in the following.

With

$$x_k = x^o_k + \Delta x_k \quad , y_k = y^o_k + \Delta y_k \quad , \tag{15}$$

where Δx_k, Δy_k are the shift parameters in Equations 11 and $\Delta x_o = \Delta y_o = 0$ (for the template), Equations 14 become

$$\Delta x_k + F^x_k + x^o_k = 0, \tag{16a}$$

$$\Delta y_k + F^y_k + y^o_k = 0. \tag{16b}$$

It is assumed that the exterior orientation parameters of each sensor k, $(X_o, Y_o, Z_o, \omega, \phi, \kappa)_k$, $k = 0, \ldots, n$ are either given or can be derived from object information (control points, control elements), and the interior orientation parameters of the sensors (x_H, y_H. . .principal point coordinates, c. . .camera constant)$_k$ are given as well. Then the parameters to be estimated in Equations 16 are the shift values Δx_k, Δy_k and the coordinates of the object point X, Y, Z. Equations 16 are nonlinear in X, Y, Z. Linearization of Equations 16 with respect to the object coordinates X, Y, Z results in

$$\Delta x_k + \frac{\partial F^x_k}{\partial X} dX + \frac{\partial F^x_k}{\partial Y} dY + \frac{\partial F^x_k}{\partial Z} dZ + F^{x(o)}_k + x^o_k = 0, \tag{17a}$$

$$\Delta y_k + \frac{\partial F^y_k}{\partial X} dX + \frac{\partial F^y_k}{\partial Y} dY + \frac{\partial F^y_k}{\partial Z} dZ + F^{y(o)}_k + y^o_k = 0. \tag{17b}$$

In order to take care of a possible systematic image deformation, the linearized collinearity conditions can be augmented by correction terms Δx_{sk}, Δy_{sk}. Their use is important, especially for close-range applications, where the deformations are significant; e.g., for CCD images, they were 140 μm at the image border. These corrections may be introduced according to the effects of previously determined self-calibration parameters.

With

$$\mathbf{t}_k = \begin{bmatrix} F^{x(o)} + x^o + \Delta x_s \\ F^{y(o)} + y^o + \Delta y_s \end{bmatrix}_k ,$$

$\mathbf{x}$ = vector of all parameters, and (18)

$\mathbf{B}_k$ = design matrix of the coefficients of the parameters for scene (k),

the extended Equations 17 yield

$$\mathbf{B}_k \, \mathbf{x} + \mathbf{t}_k = 0 \; ; k = 0, \ldots, n \tag{19}$$

Equations 11 and 19 are connected by means of the shift parameters Δx_k, Δy_k that appear in both equations.

The Equations 19 are not treated as functional constraints but as a set of observation equations.

In this case, Equations 19 become

$$-\mathbf{e}_t = \mathbf{Bx} + \mathbf{t} \; ; \mathbf{P}_t \tag{20}$$

$\mathbf{P}_t$. . . .weight coefficient matrix of $\mathbf{t}$

with

$$\mathbf{e}_t \sim N(0, \sigma^2_o \mathbf{Q}_{tt}) \; ; \mathbf{Q}_{tt} = \mathbf{P}_t^{-1} \tag{21}$$

The least-squares solution for the joint system Equation 11, Equation 20 gives

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{PA} + \mathbf{B}^T\mathbf{P}_t\mathbf{B})^{-1} (\mathbf{A}^T\mathbf{Pl} - \mathbf{B}^T\mathbf{P}_t\mathbf{t}). \tag{22}$$

The weight matrices $\mathbf{P}$, $\mathbf{P}_t$ are used as diagonal matrices. Therefore, the normal Equations 22 are of great sparsity (Figure 2), which is favorably exploited when solving the system for $\hat{\mathbf{x}}$. With the solution vector $\hat{\mathbf{x}}$ we get

$\mathbf{v} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{l}$, residual vector for intensity observations,

$\mathbf{v}_t = \mathbf{B}\hat{\mathbf{x}} + \mathbf{t}$, residual vector for collinearity constraint observations, and (23)

$$\hat{\sigma}^2_o = \frac{\mathbf{v}^T\mathbf{Pv} + \mathbf{v}_t^T\mathbf{P}_t\mathbf{v}_t}{r}, \text{ variance factor,} \tag{24}$$

where

$r = n - u$, redundancy,

n = number of observations, and

u = number of parameters.

Because of the nonlinearity of the joint system, the final solution is obtained iteratively, whereby approximate values for the nonlinear parameters (six transformation parameters for each picture patch, object coordinates of point p) are required. The iteration is stopped if each element of the solution vector $\hat{\mathbf{x}}$ in Equation 22 falls below a certain limit.

A key element of this algorithm is the conditional one-dimensional transformation of the images. The word "conditional" refers to the fact that the transformation follows certain rules. These rules relate the transformation of the images to each other and is one important reason for the improved performance of this algorithm, as shown in the section on Practical Tests.

More specifically, for the X, Y, Z version of the algorithm, the x,y-shifts of all patches are functionally related to the alter-

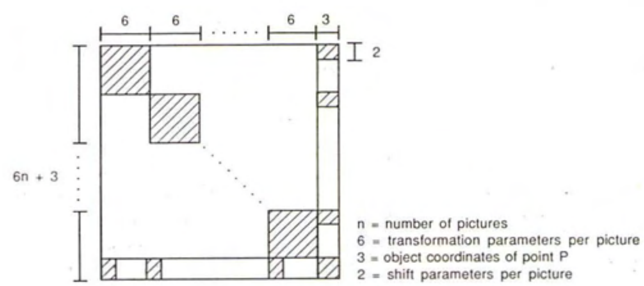


FIG. 2. Matrix sparsity pattern of normal equations (Equation 22).

ations of X, Y, Z object coordinates and are, thus, related to each other. Following the notation of Equations 17, this relation is given by

$$\begin{aligned}\Delta x_k &= -\frac{\partial F_k^x}{\partial X}\,dX - \frac{\partial F_k^x}{\partial Y}\,dY - \frac{\partial F_k^x}{\partial Z}\,dZ \\ \Delta y_k &= -\frac{\partial F_k^y}{\partial X}\,dX - \frac{\partial F_k^y}{\partial Y}\,dY - \frac{\partial F_k^y}{\partial Z}\,dZ \\ k &= 0,1\ldots n\end{aligned} \tag{25}$$

Because the template ($k = 0$) is not transformed, $\Delta x_o = \Delta y_o = 0$. Thus, from Equations 25 for $k = 0$, we can express, dX, dY as functions of dZ. Using the notations

$$F^x_{X_k} = \frac{\partial F^x_k}{\partial X}\; F^x_{Y_k} = \frac{\partial F^x_k}{\partial Y}\; F^x_{Z_k} = \frac{\partial F^x_k}{\partial Z}$$

and similarly for y

$$F^y_{X_k} = \frac{\partial F^y_k}{\partial X}\; F^y_{Y_k} = \frac{\partial F^y_k}{\partial Y}\; F^y_{Z_k} = \frac{\partial F^y_k}{\partial Z}$$

this procedure yields

$$\begin{aligned}dX &= \frac{F^y_{Z_o}\,F^x_{Y_o} - F^x_{Z_o}\,F^y_{Y_o}}{F^x_{X_o}\,F^y_{Y_o} - F^y_{X_o}\,F^x_{Y_o}}\,dZ = CX\,dZ \\ dY &= \frac{F^y_{X_o}\,F^x_{Z_o} - F^x_{X_o}\,F^y_{Z_o}}{F^x_{X_o}\,F^y_{Y_o} - F^y_{X_o}\,F^x_{Y_o}}\,dZ = CY\,dZ\end{aligned} \tag{26}$$

where CX, CY are known.

Substituting Equation 26 in Equation 25 for $k = 1,2\ldots,n$ leads to

$$\begin{aligned}\Delta x_k &= (-CX\,F^x_{X_k} - CY\,F^x_{Y_k} - F^x_{Z_k})\,dZ = Cx_k\,dZ \\ \Delta y_k &= (-CX\,F^y_{X_k} - CY\,F^y_{Y_k} - F^y_{Z_k})\,dZ = Cy_k\,dZ\end{aligned} \tag{27}$$

where Cx_k, Cy_k are known.

Equations 26 to 27 show that all parameters can be expressed as functions of one parameter (e.g., in this case of dZ). They form a system of $2n+2$ equations in $2n+3$ unknowns. The coefficients CX, CY, Cx_k, Cy_k that appear in these equations depend only on the X, Y, Z derivatives of the collinearity equations. These derivatives again depend on the exterior and interior orientations of the sensors (which are assumed to be known) and the location of the match points in the image system. Small changes in the point locations leave the derivatives almost unaffected, so practically they can be considered constant throughout the iterations. Their recomputation, though, is a trivial task. From Equations 27 it is apparent that the image patches cannot move arbitrarily along the epipolar lines, but they must follow a certain direction and step ratio. If we assume, for example (see Figure 3), that patch i moves along its epipolar line with a step S_i

$$S_i = (\Delta x_i^2 + \Delta y_i^2)^{1/2} = (Cx_i^2 + Cy_i^2)^{1/2}\,dZ$$

and any other patch j ($j = 1,\ldots,n$, $j \neq i$) with a step S_j,

$$S_j = (Cx_j^2 + Cy_j^2)^{1/2}\,dZ,$$

then the ratio $R_{ij} = S_i/S_j$ is given by

$$R_{ij} = \left[\frac{Cx_i^2 + Cy_i^2}{Cx_j^2 + Cy_j^2}\right]^{1/2}. \tag{28}$$

This ratio (Equation 28) is practically constant throughout the iterations for the reasons explained above. That means that, if the patch i moves with step S_i at point A_i, then any other patch j must move with step $S_j = S_i/R_{ij}$ at point A_j. So, these "step conditions" force the images to move simultaneously at "fixed step ratios" (e.g., A_k, $k = 1,\ldots,n$) along the epipolar lines.

Because the transformations of the images are related to each other, the target function to be minimized by the matching algorithm (e.g., squared sum of residuals) is a single global function, i.e., depends on the position of all images simultaneously, as compared to the unconstrained matching, where the transformation of the images is independent from each other, and there are n(n = number of transformed images) local target functions to be minimized. This is a considerable advantage in comparison to other algorithms (even epipolar matching), where there is no relation and mutual support among the images.

SOFTWARE IMPLEMENTATION

The previous estimation model is FORTRAN 77-coded and is installed at the Digital Photogrammetric Station (DIPS) of the Institute of Geodesy and Photogrammetry, ETH Zurich. The DIPS is described in Gruen (1986). As a software module, it is part of a larger program, which includes several mathematical models, which solve for different problems spanning from simple gray level stereo matching to a combined matching/bundle solution for any number of photographs. The program uses only the host computer and the graphic display facilities of DIPS. Although a pipeline processor (MIAP) with a large number of fast image processing functions is available at DIPS, all necessary image preprocessing routines (like filtering with variable masks), spatial resampling, geometrical warping, radiometric corrections, etc., have been rewritten for the host computer. The following functional features and options, apart from those that can be set interactively, are realized in the program:

- The matching parameters are introduced as observed quantities. Each individual parameter can carry its own weight. Thus, parameters in any combination can be constrained or even excluded before or during the iteration process.
- The collinearity conditions are introduced as observed quantities. Their weight may be chosen in the range from

  $P_{ti} \rightarrow 0$: The ith condition is not effective; the corresponding image patch is subject to gray level matching only

  to

  $P_{ti} \rightarrow \infty$: The ith condition is strictly valid; the match point of the corresponding image patch must slide along its image epipolar line during the iterations.
- In order to save computing time and to gain flexibility, the radiometric corrections (gain and offset transformations, histogram range scaling, histogram clipping, Wallis filtering) can be applied prior to or during matching and are not treated as model parameters. This is tolerable because the most important radiometric correction parameters are largely independent of the matching parameters in case of good approximate values for the transformation parameters.

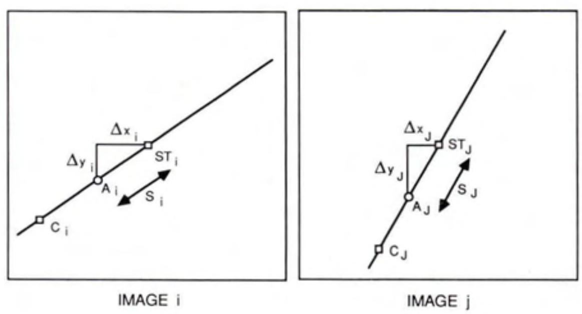


FIG. 3. Movement of match points along epipolar lines. ST_i, ST_j. . .starting wrong match point locations.
A_i, A_j. . .intermediate match point locations.
C_i, C_j. . .correct match point locations

$$\frac{\Delta x_i}{\Delta y_i} = \frac{Cx_i}{Cy_i}; \frac{\Delta x_j}{\Delta y_j} = \frac{Cx_j}{Cy_j}$$

- Approximate values for the location of the patches and XYZ object coordinates can be obtained by either selecting
  - gray level patches in the image space, whose centers, apart from the template, do not have to fulfill the collinearity condition

  or
  - by starting with approximations for X, Y, Z-coordinates of the object point and deriving the image patches by locating them by means of the collinearity condition.
- Deformation of the perspective sensor model can be compensated by applying discrete image coordinate corrections Δx_s, Δy_s in the collinearity conditions. Because the image corrections Δx_s, Δy_s follow a global image function and because of the relatively small pull-in range (compared to the whole image format), they are kept constant during the iteration steps.
- The complete covariance matrix for the estimated parameters is computed to allow for a thorough analysis of the results.
- Besides the global variance factor $\hat{\sigma}_o^2$, the partial variance factors $\hat{\sigma}_{oi}^2$ for the individual patches (i) are computed. This information can be used for mismatch and occlusion detection.
- For the approximate values of the transformation and object point coordinates parameters, information from previously processed points can be used.

PRACTICAL TESTS

These practical tests are published in order to highlight some interesting features of this multiphoto matching method. They are only part of a more extensive research effort and as such do not cover the problems in a very systematic manner.

Issues to be addressed in the following are

- approximate values and pull-in range,
- multiple solutions (mismatch), and
- occlusions (partial, multi-occlusions, template occlusion).

In order to work in a controlled test environment, a close-range test plate (format 50 by 50 cm) has been used. Figure 4 shows the arrangement. The plate consists of 25 precisely surveyed control points. Between these points a number of markings have been fixed, most of them showing a repetitive pattern. These markings are simple photo copies of quilt and basket patterns. Four convergent scenes ($|\phi| \approx 30^g$, $|\omega| \approx 20^g$) taken with an AQUA-TV HR 600 CCD frame transfer camera cover the plate.

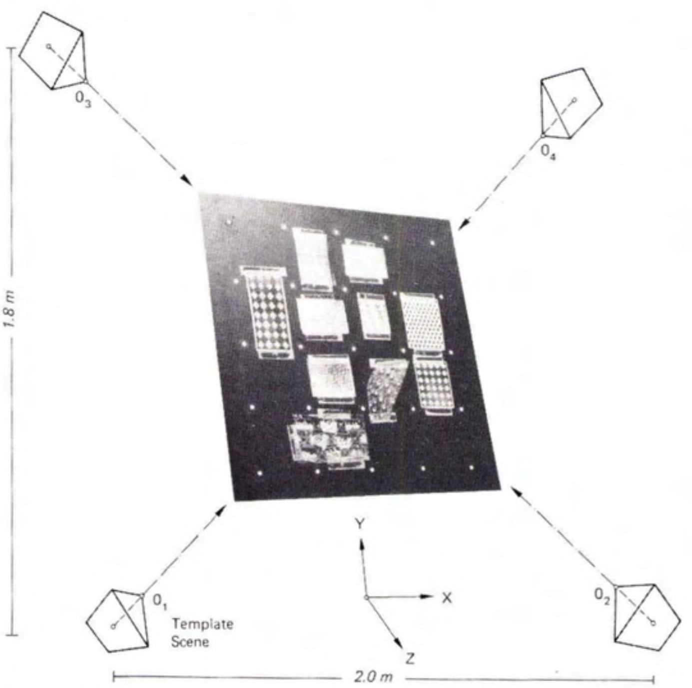


FIG. 4. CCD camera arrangement and test object.

At DIPS each scene covers 512 by 512 pixels, with a pixel spacing of 7.3 μm in x and 7.8 μm in y. The image scale was approximately 1:160 and the average object depth from the perspective centers was 2 m.

The interior and exterior orientation of the scenes and the X, Y, Z object coordinates of the control points were determined by self-calibrating bundle adjustment using all 25 control points. Image coordinate measurement was performed digitally with the unconstrained least-squares matching technique using two shift and three shaping parameters and a synthetically designed circular template. The image corrections for the deviation from the perspective model, as obtained from self-calibration, were applied in all subsequent computations. For the calibration problems and accuracy potential of these CCD cameras, see Beyer (1987) and Dähler (1987).

In general, the data processing procedure is set up such that a template is selected in reference scene no. 1. For the computation of the approximate values two options are used:

(a) Selection of object point coordinate $Z^{(o)}$; computation of the related $X^{(o)}$, $Y^{(o)}$-coordinates using the imaging ray of scene no. 1; computation of image patches for scenes nos. 2,3,4 by perspective backprojection of object point $X^{(o)}$, $Y^{(o)}$, $Z^{(o)}$

(b) Free selection of image patches 2,3,4 (in this project: interactive selection). $X^{(o)}$, $Y^{(o)}$, $Z^{(o)}$ computation according to (a)

Method (a) delivers geometrically consistent approximate values, while this consistency is not obtained with method (b) because the imaging rays of the point do not intersect in object space. However, with method (b) consistent values are obtained right after the first iteration step of the matching/point positioning procedure, if the weight of the collinearity conditions is set to $\mathbf{P}_{ti} \rightarrow \infty$ (the patches "jump" right away at the corresponding epipolar lines).

Some preliminary tests have shown that, in case of bad approximations, it is advisable to run a coarse matching, i.e., the first few iterations without shaping parameters. Thus, shaping parameters are used only at the fine matching level (final two to three iterations). In all computations the patches were prefiltered with a three by three convolution mask (local average) to speed up the iteration process. The derivatives of the gray levels were computed using the average of the patch and the template.

PULL-IN RANGE

In order to evaluate the pull-in range and thus the necessary goodness of approximate values, the circular targets of the single points were used (see Figure 4). The signal size is about seven pixels in diameter.

All 25 targets were tested, each with two different ΔZ errors (ΔZ is the deviation of the approximate value for the Z-object coordinate from its correct value):

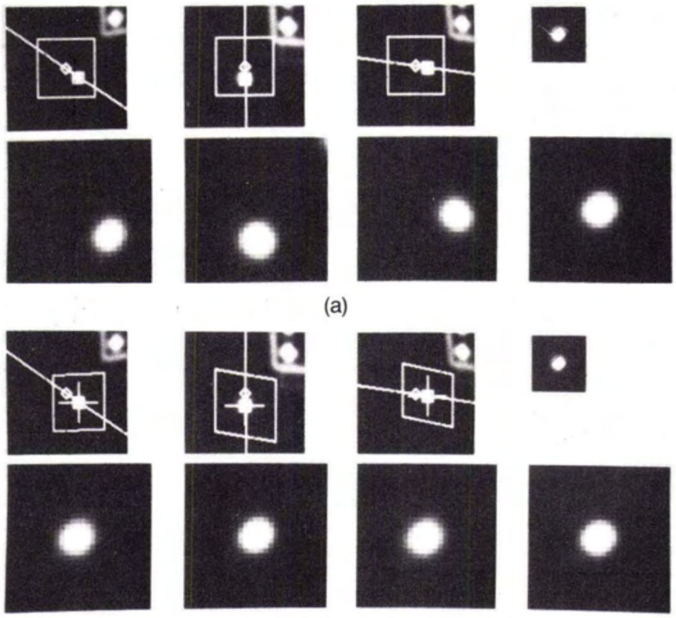
(a) ΔZ = 10 mm or 5 ‰ d_o (d_o. . .average depth of the object with respect to the perspective centers); patch size: 29 by 29 pixels

(b) ΔZ = 20 mm or 10 ‰ d_o; patch size: 41 by 41 pixels

ΔZ = 10 mm corresponds to a maximum pull-in range of six pixels. This pull-in range appears for all points in at least one image coordinate direction in all scenes. ΔZ = 20mm corresponds to 12 pixels pull-in range.

In the following figures the upper row shows the windows of images 4,3,2 from left to right and the template. The lower row shows the respective patches. In each window the epipolar lines and the transformation of the patch are shown. The cross represents the current position of the patch center, the small rhombus shows the starting position. In the figures showing the test results on pull-in range, the square represents the correct position of the patch center.

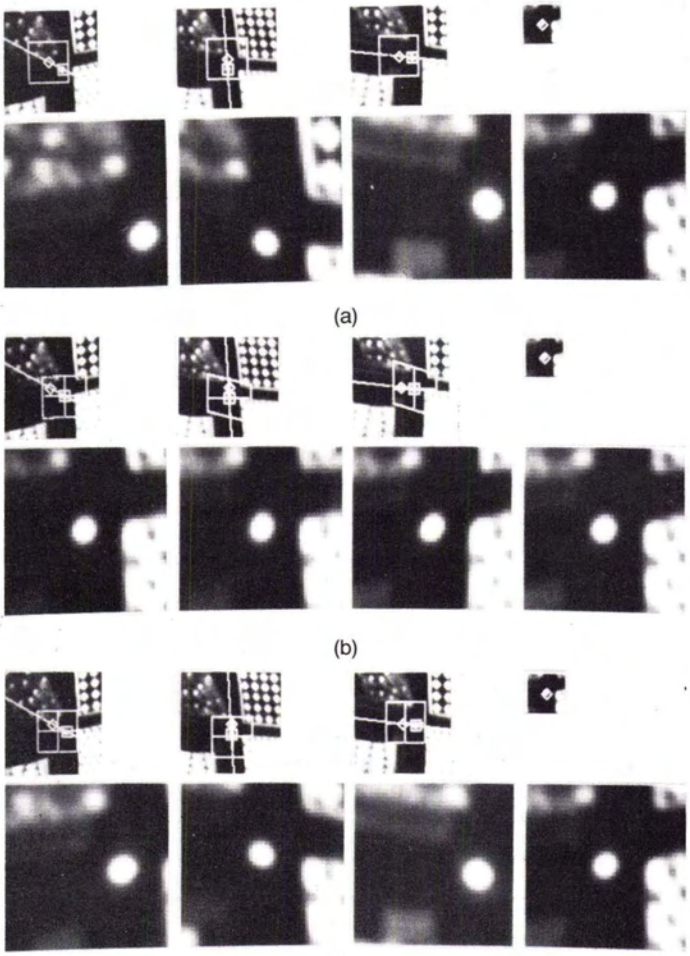
In case (a) all matches converged to the correct location (Figure 5), while in case (b) this was true only for 17 points (Figures 6a and 6b). The considerable improvements in convergence radius



(a)

(b)

FIG. 5. Six pixels pull-in range. (a) Starting, (b) End position of matching procedure.



(a)

(b)

(c)

FIG. 6. Twelve pixels pull-in range. (a) Starting, (b) End position of one case of convergence, (c) same iteration number as (b) without geometrical constraints.

and rate of the geometrically constrained multiphoto matching versus the unconstrained matching are clearly shown in Figures 6b and 6c. In Figure 6c the images 4 and 2, which have worse approximations than image 3, are still far away from the correct position, while in Figure 6b image 3 supports the convergence of the other two images.

Failure in convergence included two different cases:

- Patches with very homogeneous background and few edges. As a result the insufficient signal cannot pull in the patches (Figure 7a).
- Patches with many edges and enough texture but very dissimilar. Because of the dissimilarity, the shifts at the coarse matching stage do not converge and the subsequent transformation with all shaping parameters at the fine matching stage is wrong (Figure 7b).

Both problematic cases can be improved by increasing the patch sizes so that significant signal is included and the similarity between the patches is increased.

It must be noted that these results refer to points lying on the epipolar lines. If the image coordinate approximations are determined by selecting gray level patches in the image space, which do not fulfill the collinearity condition, then already in the first iteration the patches jump on the epipolar line, independently of how far they are initially away from it.

Figure 8 shows a special characteristic which was already observed in the DTM matching algorithm (Gruen and Baltsavias, 1986). If the approximate values are bad, the alterations of the estimated parameters are small in the first part of the iteration process. Then there is a point where the alterations take on bigger values and the iteration process speeds up considerably.

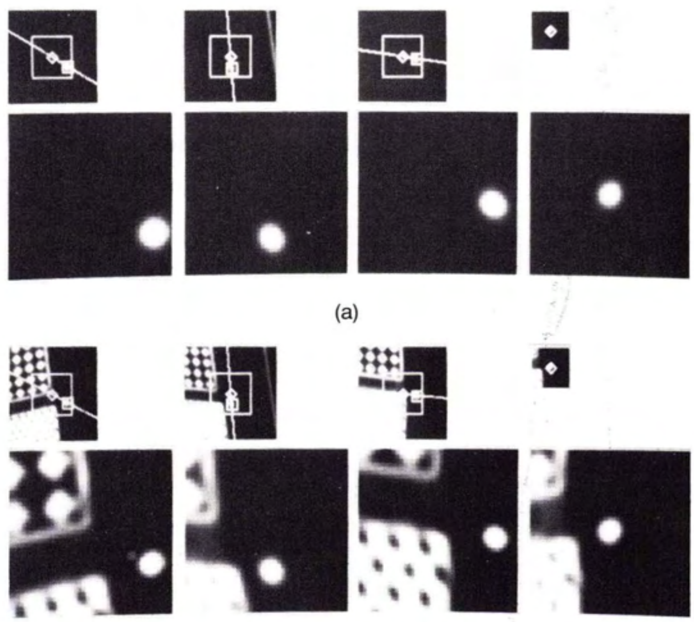
OCCLUSIONS

Because of the flatness of the test plate, occlusions were simulated by blackening a number of the pixels.

The occlusion test distinguishes three cases:

(a) Semi-occlusion of the patch in image 2 by 50 percent (see Figure 9);

(b) Multi-occlusions; patch 2 is occluded by 50 percent and, in addition, patch 3 is occluded by 25 percent (see Figure 10); and



(a)

(b)

FIG. 7. Twelve pixels pull-in range. (a), (b) two typical failure examples (starting position).

(c) Template occlusion; a 31- by 31-pixel template is occluded by a 12- by 12-pixel square in the center (see Figure 11).

The patch sizes were 31 by 31 pixels and a ΔZ-value of ΔZ = 6 mm was used. Correct convergence was obtained in all cases.

Table 1 shows the total and component standard deviations of unit weight of the matches that were used to detect occlusions. It shows clearly that occlusions can be very well detected as long as at least one image is non-occluded. If the template is occluded, other measures of goodness of fit than those of Table 1 have to be used.

Multiple Solutions

For this test, image patches of the repetitive patterns were chosen in different unfavorable positions and the convergence behavior was examined. All versions were without radiometric correction, with filtering as mentioned above, ΔZ = 6 mm, only shifts at the beginning and all other affine transformation parameters at the end stage. The dimensions of the patches were 9 by 9 pixels, which was about the size of the repetitive pattern.

Multiple solutions (side-maxima) are strongly reduced with this matching technique and mismatches are signalized in most cases. The collinearity condition forces the patches to move along the epipolar lines, which greatly reduces the danger of side-maxima. All ambiguites and mismatches outside the epipolar lines are geometrically inconsistent and are thus automatically raised, as shown in Figure 12.

Apart from that, because of the conditional one-dimensional search, the success rate is increased and the detection of false

Table 1. Standard Deviations of Unit Weight of the Matches (Total and Components) with Respect to the Template. The Dimensions Are Gray Levels

Case	σ_o Total	σ_{o2} Image 2	σ_{o3} Image 3	σ_{o4} Image 4
a	19.8	31.6	5.1	12.5
b	34.8	53.6	25.6	12.6
c	20.5	21.0	20.2	20.6

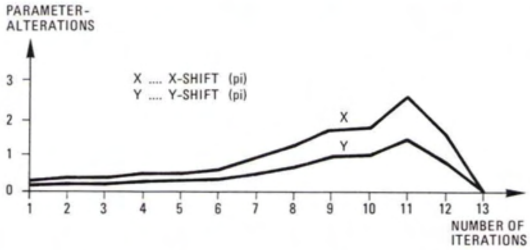


Fig. 8. Number of iterations versus parameter alterations, with ΔZ = 20 mm for one patch.

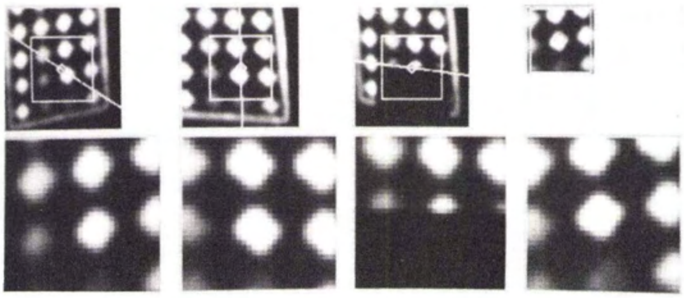


Fig. 9. Semi-occlusion in one image.

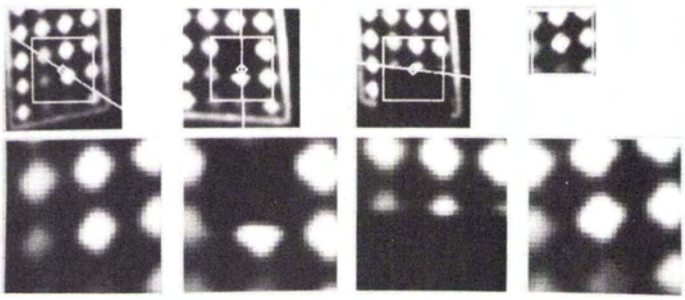


Fig. 10. Multi-occlusions.

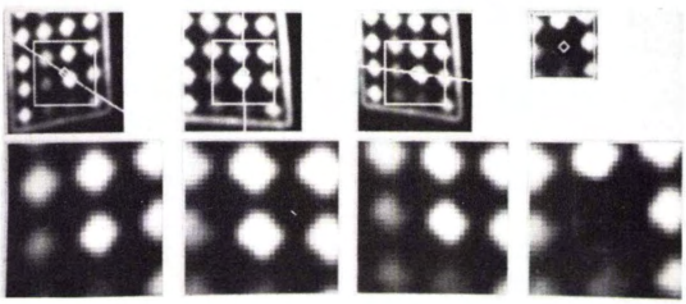


Fig. 11. Template occlusion.

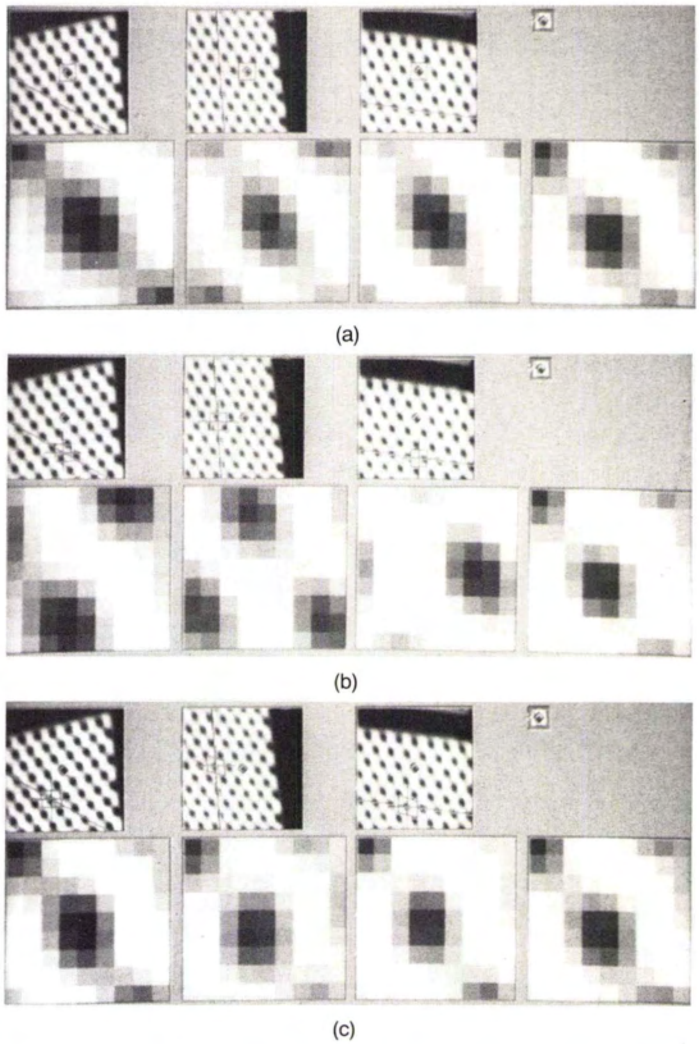


Fig. 12. Three mismatches outside the epipolar lines. Correct convergence. (a) Start position, (b) First iteration, (c) End position

convergence is in most cases possible. Success rate is increased because of the mutual relation and support among the image patches, especially when more than two images are used. As an example, a patch that is closer to a side-maximum will converge correctly if the other patches are closer to the correct match point. Thus, good patches can support and pull in candidate patches for mismatch (Figure 13). Additionally, as the number of image patches increases, the probability that all patches hit a side-maximum simultaneously decreases and, therefore, the more easier it is to detect a false matching.

False convergence can not be detected only if a side-maximum is selected on the epipolar line of each image simultaneously, which is highly improbable. If at least one patch (i) does not hit a side-maximum, this would be reflected in the size of some partial quality measure (e.g., $\hat{\sigma}_{o_i}$ and correlation coefficient of patch (i)). Also, in such a case the shaping parameters are going to be inflated. The resulting object coordinates are wrong and thus a comparison with the results of neighboring points can provide an additional control. An example is shown in Table 2 and Figure 14. Image 2 is a perfect mismatch but images 3 and 4 cannot reach a side-maximum because of the geometric constraints. Thus, the values of $\hat{\sigma}_o$, the correlation coefficient, and the shaping parameters for images 3 and 4 signalize the false convergence.

CONCLUSIONS

The geometrically constrained multiphoto matching technique offers considerable advantages with respect to pull-in range, reliability, and occlusions. Because of the use of all the geometric information available and the internal consistency of the algorithm, the success rate increases and many problematic situations are signalized.

- Approximate values with errors of about 5 to 10‰ d_o (6 to 12 pixels) are successfully handled. The use of the geometrical constraints increases the convergence radius and rate because the search is one dimensional. The use of multiple scenes has the same effect. In addition, the search is here constrained with respect to direction and step size, so that the less displaced image patches can support the further displaced ones.
- In many cases occlusions do not prevent correct convergence. Again, the less occluded patches beneficially influence the more distorted ones. The quality measures of the algorithm allow for the detection of occlusions.
- Multiple solutions are drastically reduced because of the conditional one-dimensional search. Mismatches can be detected unless all image patches hit false maxima along epipolar lines simultaneously, a very rare case.

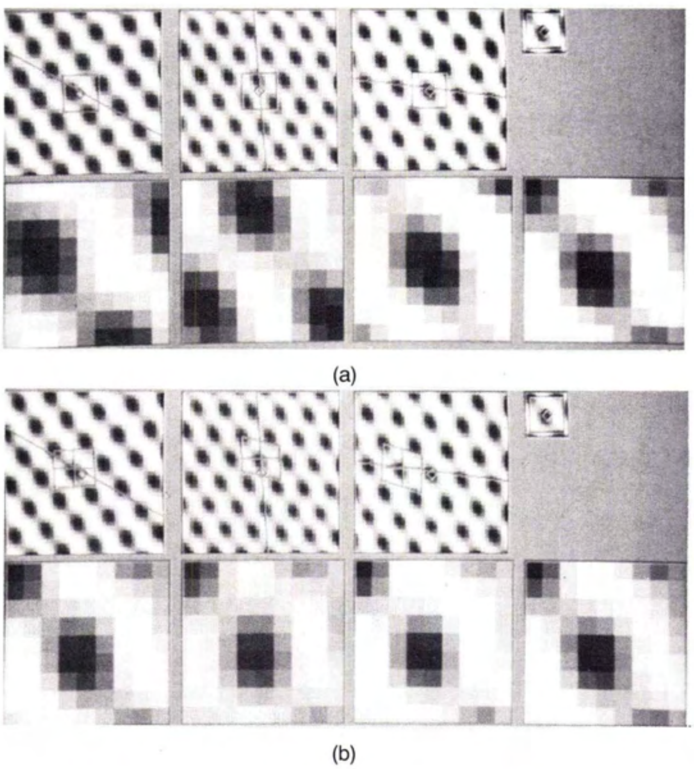


FIG. 13. One mismatch and two patches closer to correct match point along the epipolar line. Correct convergence. (a) Start, (b) End position

TABLE 2. DETECTION OF A FALSE CONVERGENCE FROM THE PARTIAL RESULTS OF IMAGE 3 AND 4 (EXAMPLE OF FIGURE 14)
ρ . . . CORRELATION COEFFICIENT Sx. . .X-SCALE, Sy. . .y-SCALE, Rx. . .x-SHEAR, Ry. . .y-SHEAR

	Total	Image 2	Image 3	Image 4
σ_o	34.9	16.6	41.2	44.1
ρ		0.92	0.13	0.07
S_x		1.17	2.14	0.03
S_y		0.99	0.78	0.40
R_x		−0.11	0.15	0.36
R_y		0.42	0.14	0.53

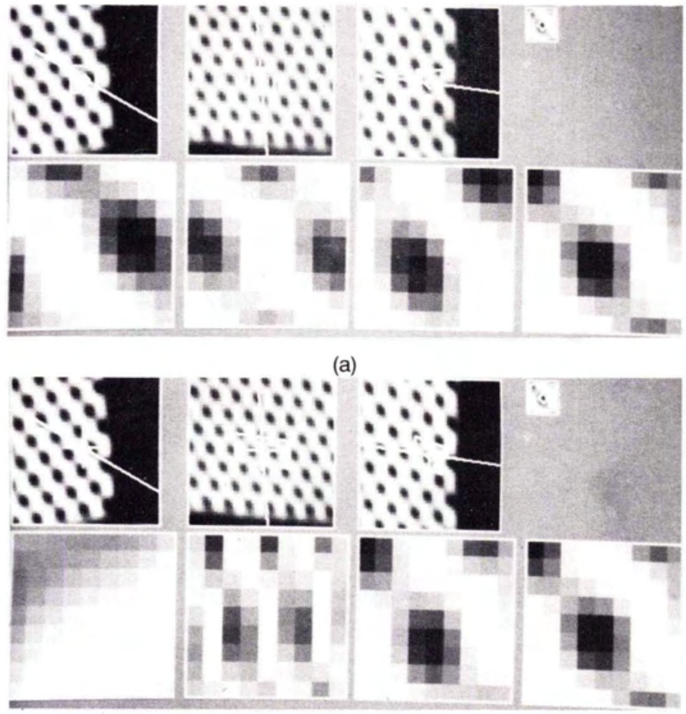


FIG. 14. Convergence to a wrong point, with at least one imperfect mismatch. (a) Start, (b) End position

Further studies will make use of more extensive data sets and will examine other aspects of the algorithm.

REFERENCES

Beyer, H., 1987. Some Aspects of the Geometric Calibration of CCD-Cameras. *Proceedings of Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, 2–4 June, pp. 68–81.

Daehler, J., 1987. Problems in Digital Image Acquisition with CCD-Cameras. *Proceedings of Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, 2–4 June, pp. 48–59.

Gruen, A., 1986. The Digital Photogrammetric Station at the ETH Zurich. *Intern. Arch. of Photogrammetry and Remote Sensing*, Vol. 26–2, pp. 76–84.

Gruen, A., and E. Baltsavias, 1985. Adaptive Least Squares Correlation with Geometrical Constraints. *Proc. of SPIE*, Vol. 595, pp. 72–82.

———, 1986. High Precision Image Matching for Digital Terrain Model Generation. *Intern. Arch. of Photogrammetry and Remote Sensing*, Vol. 26–3/1, pp. 284–296.

(Received and accepted 28 January 1988)

1988 Short Courses in Remote Sensing
South Dakota State University

The courses present basic to advanced techniques for remote sensing of natural resources. Data handling and processing, image interpretation, and field methods are among the tools presented. An applications orientation accompanies all courses. A seven-week program (courses 1 to 6) is also an option. For a complete presentation, see course listing number 13.

(1) 12–16 September 1988 - Basic Remote Sensing and Image Interpretation
(2) 19–23 September 1988 - The Basics of Digital Image Processing
(3) 26–30 September 1988 - Geographic Information Systems (GIS)
(4) 3–7 October 1988 - Remote Sensing in Land Use Studies
(5) 11–14 October 1988 - Remote Sensing: A Tool for Soil Surveys and Resource Inventories
(6) 17–23 October 1988 - Understanding Hydrologic Systems
(7) 14–18 November 1988 - Tools in Remote Sensing for Cultural Resources Survey and Management
(8) 1–5 August 1988 - Remote Sensing and GIS Applications in the Social Sciences
(9) 24–26 October 1988 - Photographic Theory and Production Methodology for the Image Interpreter
(10) 19–23 September 1988 - Reproduction of Colors and Tones in Remotely Sensed Imagery
(11) 26–28 September - Theory and Practice of Digital to Analog Conversion of Digital Data
(12) 7–9 September 1988 - How to Develop a Comprehensive Quality Assurance Program for Remotely Sensed Data
(13) 12 September–28 October 1988 - Tools and Techniques for Natural Resource Evaluations

For further information please contact

Mr. Kevin Dalsted
Office of Remote Sensing
South Dakota State University
Box 507
Brookings, SD 57007-0199
Tele. (605) 668-6894

Short Course
Introduction to Remote Sensing

Hyatt Regency Hotel. Kansas City, Missouri
17 October 1988

This Short Course – being held in conjunction with the 1988 AEG meeting and cosponsored by the American Society for Photogrammetry and Remote Sensing – is oriented toward the use of aerial and space photography, infrared, radar, and advanced sensors for engineering geology. The course is designed for those with some knowledge of aerial photointerpretation who wish to learn more about "remote sensing," photography and imagery, computer-based image processing, and recent developments in this dynamic field. The course will concentrate on applied aspects of the various sensors: when, where, and how to obtain the photography you need; and what to do with the photography/imagery when you get it. A consultant's forum will be held to discuss tactics in practical applications of remote sensing in the field and office. Participants will be introduced to

- Remote sensing terms
- Non-traditional sensors and imagery
- How to select, obtain, and use remote sensing imagery
- Applications for remote sensing in engineering geology and in regional and site studies
- Geographic information systems and their use in engineering and environmental investigations
- Microcomputer-based image processing and GIS

For further information please contact

Christopher J. Stohr
Illinois State Geological Survey
Natural Resources Building
615 East Peabody Drive
Champaign, IL 61820
Tele. (217) 244-2186