INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE CENTRATOM TRIESTE

H4.SMR/383 - 18

WORKSHOP ON REMOTE SENSING TECHNIQUES WITH APPLICATIONS TO AGRICULTURE, WATER AND WEATHER RESOURCES

(27 February - 21 March 1989)

DIGITAL PROCESSING OF STEREO IMAGES AND 3-D RECONSTRUCTION TECHNIQUES

V. CAPPELLINI
University of Florence
and
IROE - CNR
Florence
ITALY

Digital processing of stereo images and 3-D reconstruction techniques

V. Cappellini
Department of Electrical Engineering - University of Florence - Via S. Marta, 3 - 50139 Florence - and IROE-C.N.R. Via Panciatichi, 64 - Florence - Italy
L. Alparone, G. Galli, P. Langé, L. Menichetti
Dpt. of Electr. Engin. - Florence
G. Capanni
Galileo Syscam S.P.A. - Florence
R. Carlà
IROE-C.N.R. - Florence

ABSTRACT

The main problems involved in the reconstruction of 3-D profiles are related to the definition and resolution of a suitable equation set for the calibration of the whole system and to the selection of proper criteria to discriminate whether matched points are reliable or not.
This paper present a complete system able to extract data and measurement from digital stereo images obtained from aircraft or space sensor.
Both area-based and feature-based techniques have been investigated. In order to perform a good point matching, the system makes use of criteria based on a modified test, taking into account the shape of the correlation function in a neighbourhood of its maximum.
Some experimental results of automatic point matching are reported for different values of the characteristic parameters.

INTRODUCTION

The automatic extraction of three-dimensional measurements from a stereo pair of two-dimensional images is a very pursued aim, due to its many applications in several interest fields, such as robotics, geology, cartography, biomedicine and other ones.
The construction of a 3-D model of a scene starting from a digital stereo pair is a task involving the resolution of three main problems:

- definition of a parametric model relative to the acquisition geometry;
- matching of corresponding points in either of the acquired images;

- determination of the 3-D coordinates of each point, whose corresponding pixels are known, in a world reference system.

The first step, known as camera calibration, needs the knowledge for each camera of such parameters as: focal lenght, absolute position and direction of the view axis. The problem is solved by using the knowledge of the 3-D absolute position of some points, at least four (Ground Reference Points). The second problem is solved by means of a suitable mathematical function giving an estimate of the match quality and confidence. Finally, the 3-D reconstruction problem is much easier to solve than the previous ones, but its results are very sensitive to the accuracy of both calibration parameters and correlation measurements.

1. CALIBRATION

Once a pair of digital images, relative to the same scene taken from two different points of view, has been obtained, the first processing step to be performed is the definition of a suitable parametric model describing the acquisition geometry. In this work we used three coordinate systems, the first is a global reference one relative to which all the other measurements are expressed. The other two are orthonormal reference systems bounded to each camera point of view and having their z axes passing through the focal center and pointing towards the scene, while the x and y axes are coplanar to two acquisition planes (see Figure 1).

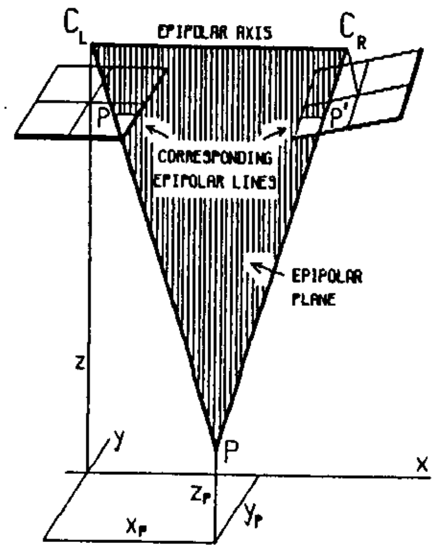

Fig. 1 - Acquisition Model of a Stereo Pair.

Using the collinearity constraint relative to a point in the 3-D space, the corresponding point projection on the image plane and the focal center of the camera, it is possible to obtain the following equations expressing the coordinates x_i, y_i on the image plane, relative to a point $\underline{P}$:

$$(1)\quad x_i = -f[(\underline{P} - \underline{C})\cdot\underline{H}]/[(\underline{P} - \underline{C})\cdot\underline{A}]$$
$$y_i = -f[(\underline{P} - \underline{C})\cdot\underline{V}]/[(\underline{P} - \underline{C})\cdot\underline{A}]$$

where $\underline{P}$ is the vector relative to the point under consideration, $\underline{C}$ is the vector relative to the focal center, $\underline{H}$, $\underline{V}$ and $\underline{A}$ are the vector defining the coordinate system bounded to the camera, while f is the focal lenght of the camera itself.
Equations (1) are not in a form suitable for computation because they assume that the intersection between the camera line of sight and the image plane lies in the center of the digitalized image and that the image itself is not scaled.
In practice these two hypotheses are not verified, so a preliminary manipulation is required to implicitly conglobate possible scale variations into the calibration phase and to obtain an arbitrary displacement of the coordinate system origin.
It can be proven (Yakimovsky, 1978) that the result of this preliminary manipulation is the definition of a new orthogonal coordinate system relative to which the collinearity equations are to be formulated again.
Using a set of at least seven points whose real world coordinates are known (theoretically four should be enough: this fact is due to the ill conditioned nature of such problems) the following equation set can be obtained:

$$(2)\quad P_{mx}\cdot i_m\cdot A_1+P_{my}\cdot i_m A_2+P_{mz}\cdot i_m\cdot A_3-P_{mx}\cdot H_1+P_{my}H_2+P_{mz}\cdot H_3-i_m\cdot C_A+C_H=0$$
$$P_{mx}\cdot j_m\cdot A_1+P_{my}\cdot j_m A_2+P_{mz}\cdot j_m\cdot A_3-P_{mx}\cdot V_1+P_{my}V_2+P_{mz}\cdot V_3-j_m\cdot C_A+C_V=0$$

where m = 1,2,...,n, while: $\underline{A}=(A_1,A_2,A_3)$ $\underline{V}=(V_1,V_2,V_3)$ $\underline{H}=(H_1,H_2,H_3)$ are the vectors defining the new orthogonal reference system, i_m and j_m are the image coordinates of the point whose real world coordinates are P_{mx}, P_{my}, P_{mz} and finally $C_A=\underline{C}\cdot\underline{A}$, $C_H=\underline{C}\cdot\underline{H}$, $C_V=\underline{C}\cdot\underline{V}$ where · denotes scalar vector product.
The system given by (2) can be solved grouping equations five by five and multiplying each group for suitable constants with the aim to eliminate the coefficients relative to H_1, H_2, H_3 and C_H:

	k_{11}	equation 1	group 1 (equations 1–5)	
k_{21}	k_{12}	equation 2		group 2 (equations 2–6)
k_{22}	k_{13}	equation 3		
k_{23}	k_{14}	equation 4		
k_{24}	k_{15}	equation 5		
k_{25}	.....	equation 6		

(3)

$k_{n-4,1}$	.....	equation n-4	group n-4 (equations n-4 – n)
$k_{n-4,2}$	.....	equation n-3	
$k_{n-4,3}$	.....	equation n-2	
$k_{n-4,4}$	.....	equation n-1	
$k_{n-4,5}$	.....	equation n	

For each group the following constraints are imposed:

$$(4)\quad k_{l1}\cdot P_{l,x}+k_{l2}\cdot P_{l+1,x}+k_{l3}\cdot P_{l+2,x}+k_{l4}\cdot P_{l+3,x}+k_{l5}\cdot P_{l+4,x}=0$$
$$k_{l1}\cdot P_{l,y}+k_{l2}\cdot P_{l+1,y}+k_{l3}\cdot P_{l+2,y}+k_{l4}\cdot P_{l+3,y}+k_{l5}\cdot P_{l+4,y}=0$$
$$k_{l1}\cdot P_{l,z}+k_{l2}\cdot P_{l+1,z}+k_{l3}\cdot P_{l+2,z}+k_{l4}\cdot P_{l+3,z}+k_{l5}\cdot P_{l+4,z}=0$$
$$k_{l1}+k_{l2}+k_{l3}+k_{l4}+k_{l5}=0$$

where l = 1,2,....,n-4.
Choosing the points from the original set so as to be not coplanar four by four, system (3) of n-4 groups of equations can be solved so as to obtain a new linear system of equations containing only A_1, A_2, A_3 and C_A as unknowns; finally, by solving this system, it is possible to obtain the vectors $\underline{A}$, $\underline{H}$, $\underline{V}$ and $\underline{C}$.
The previous steps can be repeated for the second reference system, bounded to the second point of view, recovering in such a way the complete knowledge about the acquisition geometry.
Using this knowledge and the coordinates of corresponding points in both digitalized images, the 3-D coordinates of the related points can be obtained as we will describe below.

2. DIGITAL CORRELATION OF CORRESPONDING POINTS

In order to determine couples of corresponding points in both images two approaches have been developed: an area-based matching, where the corrispondence of different points is determined by evaluating the similarity of areas around each pixel; a feature-based matching, in which structural informations are preliminary extracted from each image separately and then a matching criterion is applied to certain suitable characteristics (features) peculiar of each structure. In the following subsection we present a feasible procedure for feature-based stereo matching and a more detailed analysis of the algorithms and the problems involved in the area-based approach, which is by itself simpler and more suitable for point matching with a high degree of accuracy.

2.1 Feature-based Matching

The principal steps of a feature-based matching may be summarized in the block diagram of Figure 2, where the calibration phase has been omitted because it is not specific of this kind of approach.
On both digitized images (left and right) a processing operation may be eventually performed to reduce noise and disturbs and, at the same time, to enhance gray level discontinuities (edges).
Such edges can be easily detected by means of circular operators (i.e. Sobel operator), able to overcome the limitation due to the discrete nature of image.
Subsequent thresholding and thinning steps permit to follow the contours of the scene with thick lines. Eventual gaps due to the above steps are furtherly filled by linking edges and approximating then with segments of solid lines (Davies, 1986).
A proper choice of significant features such as contrast and orientation allows to extract from images a synthetic and adequately complete representation of their information content.
Now it is possible to match the elements of a scene both single and grouped in more complex structures, represented through polygonals.

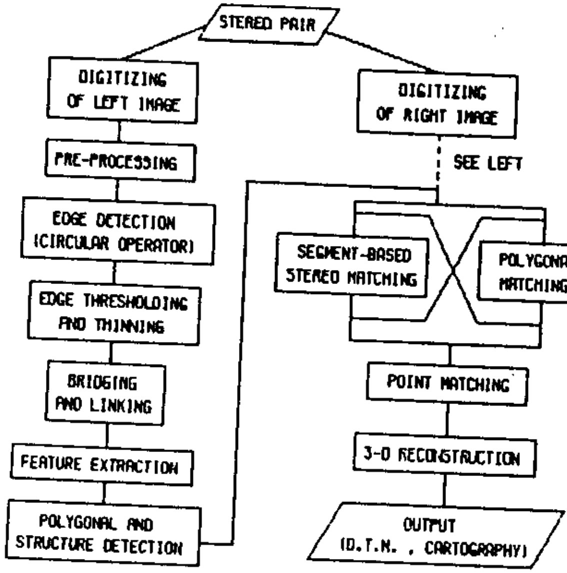

Fig. 2 - Block Diagram for Feature-based Stereo Matching.

A search in both images of polygonal structure with specified characteristics can provide an useful feedback for the stereo matching of single segments (Medioni, 1985). On the other hand the correspondence of certain couples of segments can influence the polygonal matching itself.
After this complex and articulated process of segment-polygonal matching, single point correspondances are very easy to find, although with less precision than by employing area-based algorithms.
The subsequent phases of 3-D reconstruction and data output in different forms are not specific of feature-based approach and will be seen in the following.

2.2 Area-based Matching

This second approach will be developed in more details, and also experimental results will be presented in Section 4.
In order to evaluate the similiarity of small areas around points of interest, two kinds of techniques are presented: the former is based on the use of a proper correlation function to determine a coefficient from which to decide whether a points in the left image may be corresponded to another in the right one; the latter makes use of an iterative least square resolution algorithm.
Before discussing in details the above methods, let us introduce some definitions, useful to understand the following. With reference to Figure 3, we denote as target area a rectangular window of size MxN centered around the pixel P of the left image of which we search for the correspondent in the right image.

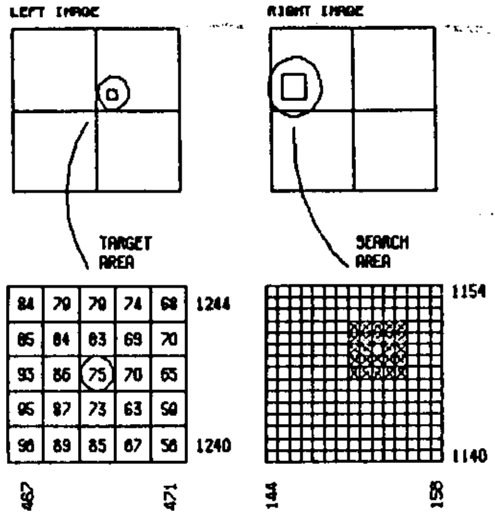

Fig. 3 - Definitions for Area-based Stereo Matching.

Such a search is brought about on a suitable rectangular window in the right image called search area. The size of this window is chosen so as to guarantee that it does include the corresponding point we are looking for. For every pixel of the search area of coordinates (m,n) we measure the similarity between the target area and a subwindow of the search area, centered around the pixel (m,n), of size MxN. The corresponding of pixel P in the search area is the point (m',n') which maximize such a similarity, according to the adopted criterion to evaluate it.

2.2.1 Matching based on the Correlation Coefficient

In order to evaluate the similarity between corresponding areas a coefficient may be computed for each pixel (m,n) of the search area. Such a coefficient is determined by applying a suitable mathematical function between points of target areas and of corresponding subwindows of the search area.
The choice of the function depends on the required accuracy as well as on computation requirements.
The most widely used function is the classical cross-correlation function (Rosenholm, 1985); with reference to target and search areas we define:

$$(5) \quad r(m,n) = \frac{\Sigma_i \Sigma_j (g_t - \mu_t)(g_s - \mu_s)}{([\Sigma_i \Sigma_j (g_t - \mu_t)^2]\cdot[\Sigma_i \Sigma_j (g_s - \mu_s)^2])^{1/2}}$$

where the subscripts i and j span window sizes, g_t and μ_t denote the gray level at (i,j) and its mean in the target area, g_s and μ_s the gray level and its mean in the corresponding subwindow of

the search area centered in (m,n). Such a function yields values normalized between 1 and -1: 1 means complete similarity, 0 means no similarity and -1 opposite similarity.
Another suitable function is the correlation intensity function (Ehlers, 1982), derived from considerations of coherent optics. It is defined as:

$$(6) \quad I(m,n) = \frac{(\Sigma_i \Sigma_j \cos[p(g_1-g_2)])^2 + (\Sigma_i \Sigma_j \sin[p(g_1-g_2)])^2}{(M \cdot N)^2}$$

where p, normalization parameter, is given by:

$$(7) \quad p = \frac{0.5\,\pi\,(M \cdot N)^{1/2}}{(\Sigma_i \Sigma_j (g_1 - \mu_1)^2 + \Sigma_i \Sigma_j (g_2 - \mu_2)^2)^{1/2}}$$

Values of (6) span over (1,0) with meaning as above.

A simple function to measure the similarity of areas may also be given by the sum of the absolute differences between gray levels. In order to obtain normalization between 1 and 0 for an easy comparison with the above ones, it is defined as:

$$(8) \quad d(m,n) = 1 - \frac{\Sigma_i \Sigma_j \,|\, g_1 - g_2 \,|}{MNL}$$

where L is the maximum number of gray levels of both digital images, usually L=256.
Function (8) requires a smaller number of operations but, unfortunately it has proven itself less efficient than the other ones.
After defining formally three proper mathematical functions, we need an operative criterion to detect point correspondences and avoid mismatch errors. The simple choice of the maximum over (m,n) of the correlation coefficient, although adequately thresholded, is not a suitable criterion because leads to a high percentage of mismatches, that causes gross errors.
As suggested by a recent paper (Rosenholm, 1985) we choose a criterion based not on the maximum value but on the shape of the two dimensional discrete function given by the correlation coefficient, quantitatively evaluated by means of its second order derivatives (approximated by the second order differences).
By thresholding the values of such derivatives around the maximum point, we decide whether to accept the point as well matched, or to discard it because mismatched.
Up to now we have implicitly considered a correspondence between integral values of the coordinates (pixel correspondence).
Now we want to extend the concept of correspondence also to non-integer values (subpixel correspondence) because in many cases, due to the discrete nature of the image data, the corresponding point in the left image of one pixel may be spread over four neighbour pixel in the left one. In this case, the position of the maximum values can be given by a suitable interpolation (we used quadratic interpolation) in the neighbourhood of the maximum. The match criterion remains unchanged and now the second derivatives can provide information about the subpixel displacement of the corresponding point (Rosenholm, 1985).

2.2.2 Least Square Matching

Let $g_t(x,y)$ be the gray level matrix denoting the target area and $g_s(x,y)$ the search area. Differences between corresponding areas are both geometric, because images have been taken from different points of view, and radiometric, because of eventual different conditions of light during takes. Either area will be exactly corresponded, once it is known a particular geometric and radiometric transformation that minimizes the sum of the squares of the differences between the respective gray levels.
The algorithm is based upon a mathematical model regarding target area as obtained through a geometric and radiometric transformation of the search area: a theoretical justification lies in the fact that both $g_t(x,y)$ and $g_s(x,y)$ represent approximately the same object.
If we express analytically the above consideration, we obtain the following relation:

$$(9) \quad g_t(x_t,y_t) + n(x_t,y_t) = g_s(x_s,y_s)$$

and

$$(10) \quad \begin{aligned} x_t &= f_1\,(x_s,y_s) \\ y_t &= f_2\,(x_s,y_s) \\ g_t &= f_3\,(x_s,y_s) \end{aligned}$$

where x_s and y_s are coordinates of pixel of the search area; x_t, y_t those of the target area; $g_s(x_s,y_s)$ and $g_t(x_t,y_t)$ represent the gray levels of the search and target area respectively; $n(x_t,y_t)$ is noise, supposed additive. Geometrical transformations are denoted by x_t and y_t, while the radiometric one by g_t; they are given by functions f_1, f_2 and f_3 respectively.
The simplest transformation model, without radiometric parameters, is given by a geometrical translation, given by

$$(11) \quad \begin{aligned} x_t &= x_s + p_1 \\ y_t &= y_s + p_2 \\ g_t &= g_s \end{aligned}$$

where p_1 and p_2 represent the shift factors which, added to each pixel (x_s,y_s) of the search area, give the best match with the target area.
Such parameters can be evaluated by means of an iterative least square procedure (Rosenholm, 1987). At each iteration step the search area is resampled starting from the original values quadratically interpolated, regarding as new coordinates of each pixel the following values:

$$(12) \quad \begin{aligned} (x_s)_{i+1} &= (x_s)_i + (dp_1)_i \\ (y_s)_{i+1} &= (y_s)_i + (dp_2)_i \end{aligned}$$

where subscripts i and i+1 denote iteration steps and $(dp_1)_i$ and $(dp_2)_i$ are the increments to give to $(x_s)_i$ and $(y_s)_i$ at the i-th step.

In this way the method is intrinsically subpixel and may provide very accurate results, depending on the choice of the transformation functions.

3. 3-D RECONSTRUCTION

The calibration parameters previously evaluated can be used to obtain the position of any point, when the coordinates of its projection on both images are known. The transformation from real world coordinates to the image plane is a projection, so that it is possible to write:

$$(13) \qquad \underline{P} - \underline{C} = t\underline{R} \quad \Rightarrow \quad \underline{P} = \underline{C} + t\underline{R}$$

where $\underline{R}$ is a vector parallel to the direction of the straight line connecting the focal center $\underline{C}$ with the generic point $\underline{P}$.
By using one of the above equations for either point of the corresponding couple we obtain two straight lines crossing the point $\underline{P}$ we are looking for, at least in the ideal case.
Practically the intersection between such lines does not exist because of the numerical errors in the calibration phase and of the approssimate coordinates of each pixel; for such reason, $\underline{P}$ will be approximately given by the point in the 3-D space which minimizes the sum of its distances from both lines.
Two main techniques exist to determine the coordinates of the point P. The former makes use of an analytical approach (Yakimovsky, 1978) computing the minimum of a non-linear function. The latter employs a geometrical approach: P is determined as the middle point of the segment orthogonal to both lines simultaneously.

4. EXPERIMENTAL RESULTS AND CONCLUSIONS

Performance evaluations of the different automatic matching techniques presented in the above sections have been divided according to two different purposes. First we deal with the algorithms based on the correlation coefficent method, evaluating their pixel precision for different values of the characteristic parameters. Then we evaluate subpixel precision for both the correlation coefficient and the least square methods.
Test images are subparts of an aerial stereo pair digitized with a resolution of 50 µm and are shown in Figures 4 and 5 for left and right image respectively.
For our purposes we need the knowledge of the exact positions of a suitable subset of corresponding points (50 points), chosen as distinguishable points.
Pixel (integer) values for reference have been obtained by comparing interactively suitably zoomed windows of the digitized images. High precision subpixel values have been provided by GALILEO SYSCAM S.P.A., working with a DIGICART analytic stereoplotter, from the original analog images.
For the pixel approach such parameters have been determined:

- percentage of matched points for different dimensions of the target area and different functions (See Table 1, where CR denotes the cross-correlation function defined in (5), CI the correlation intensity function (6) and DL the absolute difference (8));

Fig. 4 - Original Left Image.

Fig. 5 - Original Right Image.

Table 1 - Percentages of exactly matched points for different correlation functions and different sizes of target area.

	3x3	5x5	7x7	9x9	11x11	13x13	15x15	17x17
CR	50	84	92	78	72	60	52	52
CI	52	82	88	70	60	50	36	36
DL	42	74	78	66	62	60	48	48

- percentage of mismatch errors evaluating the second derivatives of the correlation coefficient for different threshold values and for different dimensions of the target area (Table 2);
- computation times for the different correlation functions (Figure 6).

The use of the maximum of the correlation coefficient in order to determine the matching of a couple of point provides considerably large errors and threfore the results are not presented here.
As one can note from the numerical values reported, the cross-correlation coefficient provides the best results and is acceptably economical in terms of computation time. Furthermore, the choice of the matching criterion based on the second order derivatives gives a very low percentage of mismatch errors and is therefore suitable for an automatic matching.
As to subpixel evaluations, for what regards the correlation coefficient we discarded as gross errors those above a value of one half of a pixel in either (x or y) direction; in this case the percentage of matched points %Npm results considerably lowered. Reported results (Table 3) include also the total standard deviation of the error, given by

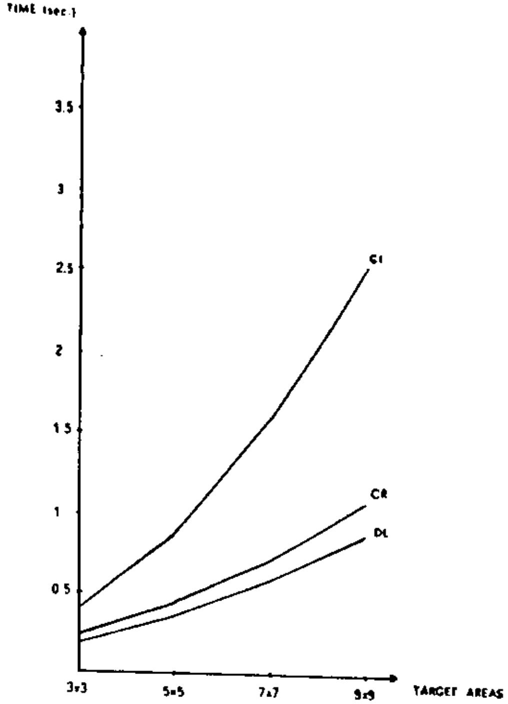

Fig. 6 - Unitary Computation Times.

Table 2 - Percentages of mismatch errors for different threshold values and different sizes of target area.

	3x3	5x5	7x7	9x9	11x11	13x13	15x15	17x17
0.2	30	10	6	16	22	28	26	28
0.3	28	12	14	22	32	26	30	32
0.4	26	24	18	32	44	38	38	32

Table 3 - Subpixel accuracy evaluations for cross-correlation coefficient (CR) with different sizes of target area.

	3x3	5x5	7x7	9x9	11x11	13x13	15x15	17x17
%Npm	52	64	52	54	56	46	44	36
Sxy	0.35	0.33	0.35	0.31	0.32	0.35	0.33	0.33

(14) $$S_{xy} = (S_x^2 + S_y^2)^{1/2}$$

where S_x and S_y are the std. deviations evaluated in x and y directions respectively. Once more results are given for different types of target areas.

In a similar way also the results of the least square method are presented. In this case a further parameter is the maximum number of iterations allowed. Once more the results refer to different target sizes (Table 4).

Least square method, which is intrinsically subpixel, provides slightly better results, although the computation time is generally greater. Further inprovements in subpixel precision could be achieved by employing a more sophisticated model of pixel trasformation, such as complete geometrical affine transform as well as a radiometric trasformation of the gray level (Fertl, 1985).

Table 4 - Subpixel accuracy evaluations for least square matching with different sizes of target area.

Max. number of iterations n=8

	3x3	5x5	7x7	9x9	11x11	13x13	15x15
%Npm	28	44	38	48	42	34	34
Sxy	0.30	0.35	0.38	0.37	0.33	0.34	0.30

Max. number of iterations n=16

	3x3	5x5	7x7	9x9	11x11	13x13	15x15
%Npm	42	46	40	50	44	34	34
Sxy	0.33	0.35	0.38	0.38	0.33	0.34	0.30

REFERENCES

Journals: DAVIES E. R., 1986, Image Space Transforms for Detecting Straight Edge in Industrial Images, Pattern Recognition Letter N.4, Elsevier Science Publishers B. V. (North-Holland) pp. 185-192

EHLERS M., 1982, Increase in Correlation Accuracy of Remote Sensing Imagery by Digital Filtering, Phot. Eng. Rem. Sens. N.3, pp. 357-365

MEDIONI G. NEVATIA R., 1985, Segment-Based Stereo Matching, Comp. Graph. Im. Proc. N.31, pp. 2-18

PERTL A., 1985, Digital Image Correlation with an Analytical Plotter, Photogrammetria N. 40, pp. 9-19

ROSENHOLM D., 1987, Least Square Matching Method: Some Experimental Results, Photogrammetric Record N. 10, pp. 493-512

YAKIMOVSKY Y. CUNNINGHAM R., 1978, A System for Extracting Three-Dimensional Measurements From a Stereo Pair of TV Cameras, Comp. Graph. Im. Pr. N. 7, pp. 195-210

Reports: ROSENHOLM D., 1985, Digital Matching of Simulated SPOT Images, Fotogrammetriska Meddelanden, 2:50, Department of Photogrammetry Roy. Ins. Tec. Stocholm