RESULTS OF TEST ON IMAGE MATCHING OF ISPRS WG III/4

Eberhard GUELCH
Institute for Photogrammetry
University of Stuttgart, Kepler Str. 11, D-7000 Stuttgart 1
Federal Republic of Germany
COMMISSION III/4

Abstract

In 1986 an empirical test on image matching was started by ISPRS WG III/4. The main purpose of the test was to find out the state of art of image matching procedures applied in the fields of computer vision, pattern recognition and photogrammetry.

The paper presents results of the test analysis with respect to different topics. A comparative study of approaches and strategies is performed. The matching results are compared with manual or computational checks, leading to precision and reliability studies of the procedures.

1. INTRODUCTION

The aim of the test on image matching, initiated by ISPRS working group III/4 in 1986 was to provide information on the state of art of algorithms for finding the correspondence between two digital images. This problem is of increasing importance in computer vision as well as in photogrammetry, where the step to digital image processing is taken by more and more institutes. Several techniques and sensors are used for data acquisition. If two or more images are taken of the object, the problem is to find homologous points in the images to derive e.g. a 3D model of the object. The variety of applications is great, from mapping or production control to biological examinations. On this matching problem many institutions are working. Approaches from different disciplines solving similar problems should be compared to provide information for all sides.

In chapter 2 is given a description of the test on image matching which has been designed to compare the performance of different algorithms and strategies used for 2D matching.

The test material was sent to more than 60 institutions all over the world. From 15 institutions results came back. Involved disciplines are computer vision, computer science, electronics, physics, robotics and photogrammetry. A list of the active institutions is given in table 1.1, ordered by country. One aim of the test, the interdisciplinary aspect is reached, by almost 50% photogrammetrists and 50% non photogrammetrists performing the test.

Table 1.1 LIST OF PARTICIPATING INSTITUTIONS (BY Oct. 1987)

Wuhan Technical University of Surveying and Mapping (2)	CHINA
University of Copenhagen, Institute of Datalogy (2)	DENMARK
Institut National de Recherche en Informatique et Automatique	FRANCE
TH Darmstadt, Institut fuer Photogrammetrie	FRG
TU Berlin, Fachgebiet Photogrammetrie und Karthographie	FRG
Universitaet Bonn, Institut fuer Photogrammetrie	FRG
Universitaet Hannover, Inst. f. Photogrammetrie u. Ingenieurverm.	FRG
Universitaet Karlsruhe, Inst. f. Photogrammetrie und Fernerkundung	FRG
Universitaet Stuttgart, Institut fuer Physikalische Elektronik	FRG
Romanian Committee for Photogrammetry and Remote Sensing	ROMANIA
University College London, Dept. of Photogrammetry	UK
University of Sheffield, AI Vision Research Unit	UK
Defense Mapping Agency, Aero Space Center	USA
SRI-International (2)	USA
University of Southern California, I. f. Robotics a. Intell. Systems	USA

(2) two participants

The amount of results given by the active participants is big enough to get an overview on quite different algorithms available for image matching. All image pairs could be matched by at least three or four participants. With four image pairs done by more than 50% of the participants intensive comparisons were possible. Table 1.2 provides the distribution of results ordered by image pair.

Table 1.2 DISTRIBUTION OF RESULTS

image pair		number of participants	
1	CAR I	▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒	17
2	QUARRY	▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒	9
3	OLYMPIA I	▒ ▒ ▒ ▒ ▒ ▒ ▒	7
4	SOUTH AMERICA	▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒	12
5	BRIDGE	▒ ▒ ▒ ▒ ▒	5
6	TREE	▒ ▒ ▒ ▒	4
7	ISLAND	▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒	14
8	SWITZERLAND	▒ ▒ ▒ ▒	4
9	CAR II	▒ ▒ ▒ ▒	4
10	WALL	▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒ ▒	13
11	OLYMPIA II	▒ ▒ ▒ ▒	4
12	HOUSE	▒ ▒ ▒ ▒ ▒ ▒	6

In chapter 3 is given a classification of applied algorithms and strategies. A detailed description of the algorithms can be derived from the literature list in the appendix. This list compiled out of the participant's descriptions provides references on the applied algorithms.

In chapter 4 the strategy of analysis and the results are presented. The results can't be described very detailed of course. The aim is to give a summarized report on the results, under consideration of some main aspects of the analysis. The complete and detailed analysis of each participant's result will be given in the final report of the working group. Beside the comparison of approaches the main aspect of analysis was the manual or computational check of matching results to derive information on precision and reliability of the applied procedures, causing e.g. more than 23,000 check measurements by human operator. These comparative results are presented as well as informations on the success of selfdiagnosis of participants, respectively their algorithms. The results presented here are from 16 out of 18 active participants. Due to problems with the mailed data, the results of two participants couldn't be finished completely in time for this paper, therefore they are excluded, but they will be presented in the detailed final report.

In chapter 5 some conclusions out of the experience with the analysed data are provided.

2. TEST DESIGN

The addressed problems by the test are briefly presented in 2.1. According to them several tasks were formulated which were given to the participants. In 2.2 the tasks and questionnaires are described in detail. In 2.3 information on the distributed data is given. The image material was choosen to cover a wide range of applications opening the possibility for very different approaches to run the test.

2.1 Aim of Test

The aim of the test on image matching was to provide information on the precision and reliability performance and on the flexibility of currently applied matching algorithms in computer vision and photogrammetry.

Specifically the following problems were addressed:

- How do the procedures react on images of different complexity?
- How much exterior or a priori knowledge do the procedures require to be able to yield a good result?
- In how far do the procedures assess the quality of their result?
- How precise, in terms of a standard deviation in pixels, are the estimated parallaxes/disparities?

One special question was if all images could be managed by participants without special adaption of their program. As a common theory for comparing the different approaches was not likely to be available within the runtime of this test, it was expected at least to derive some objective measures indicating which procedure one should use under specified conditions. Out of the test results one should be able to get a ranking of the procedures with respect to flexibility, reliability and precision.

2.2 Distributed Tasks

QUESTIONNAIRE FOR MATCHING PROCEDURE

To get a detailed information on the used procedures each participant was asked to give information about his matching algorithm concerning several topics, which might be useful for characterization and classification. These topics are listed below.

Main field of application
Type of feature selection
- motivation for the selection and the degree of invariance with respect to geometric and radiometric differences between the images.

Similarity measure
- invariant/noninvariant to geometric/radiometric changes

Mathematical model for object surface
- parametric, stochastic, syntactic, edges, occlusions

Matching algorithm
- consistency achievement
- topdown/bidirectional procedure
- integration of mathematical model for the object surface

Treatment of edges, occlusions and multiple surfaces
Tuning parameters
- theoretical justification
- sensitivity of the result to small changes of the parameters

Monocular clues
- use of shading or other monocular clues to support the stereo matching

MATCHING TASKS

For each image pair was given a set of possible tasks from which the participants could choose one or more. There were two standard tasks, repeated in the task description of most of the image pairs.

standard task A

"Determine the parallaxes at a predefined grid and indicate how you determined them. The grid is given in the left image. If available give quality measures for the parallaxes. For each grid point give an indicator I, which notes if no parallax could be determined, if the parallax is measured directly or the parallax is interpolated or others."

standard task B

"Determine the parallaxes at selected points. Please specify the selection scheme. If available give quality measures for the determined parallaxes."

To some images were given special tasks, like segmentation, determination of object planes, area mensuration etc.

QUESTIONNAIRE FOR PERFORMANCE ON TEST IMAGES

Besides the general information, the participants were further asked to describe the performance of their algorithm(s) on the test images. The following problems were addressed:

Initialization
- use of (a priori) knowledge given together with the data
- approximate values for parallaxes, or other transformation parameters
- choice of tuning parameters
- use/need of epipolar geometry

Matching
- change of procedure to meet conditions of the test images
- change of tuning parameters

Analysis of actual matching
- criteria for acceptance of result
- evaluation of precision/reliability
- use of graphical display
- check on completeness and consistency

Interpolation scheme
- used model of surface
- used model for interpolation

Selection of image pairs and tasks, if some were not processed

- Why did you select certain images or tasks?
- Have trials been necessary to decide on the selection or could you decide just from the information given in the image pair or task description?
- Can you generalize your decision? Is there an objective and simple criterium to decide, which images are managable?

2.3 Distributed Data

12 image pairs with 240x240 pixels have been scanned from transparencies of metric and amateur cameras, covering scales between 1:20 and 1:30 000.

For image pair 1 the true correspondences are known in order to be able to estimate the absolute precision of the matching procedures without relying on human stereopsy.

Table 2 Distributed Image Pairs
(also see fig. 1)

1 C A R I Ia epipolar digital surface model input to CAD inspection	5 B R I D G E IIa rel. orient.* digital surface model navigation	9 C A R I I IIIa rel. orient. digital surface model input to CAD inspection
2 Q U A R R Y Ib rel. orient. digital surface model individual points volume determination monitoring movements	6 T R E E IIb rel. orient.* digital line model individual points biological application monitoring growth	10 W A L L IIIb epipolar digital surface model wire model of edges volume determination geological analysis
3 O L Y M P I A I Ic rel. orient. digital surface model individual points mapping contours monitoring movements	7 I S L A N D IIc epipolar digital elevation model topographic mapping	11 O L Y M P I A I I IIIc rel. orient. digital elevation model individual points mapping monitoring movements
4 SOUTH AMERICA Id epipolar digital elevation model topographic mapping	8 SWITZERLAND IId rel. orient. digital elevation model height of trees topographic mapping tree age determination	12 H O U S E IIId rel. orient. digital surface model position of house terrain contour lines large scale mapping

Explanations for table 2

number	name
type	orientation
x)	
xx)	

x): possible results of the matching

xx): possible applications

types: complexity & scale

complexity:

(I) low : the object surface is smooth over the entire window with max. one edge

(II)medium: the object surface shows large differences in depth, several edges and possibly some occlusions.

(III)high : the object is highly distributed in depth, the perception of occlusions is essential for a description of the object.

scale:

(a) app. 1:20 (c) app. 1:2 000

(b) app. 1:200 (d) app. 1:20 000

orientation: epipolar - normal image window, no y-parallaxes

rel. orient. - parameters of relative orientation are given (* approximately)

Fig. 1 IMAGE PAIRS

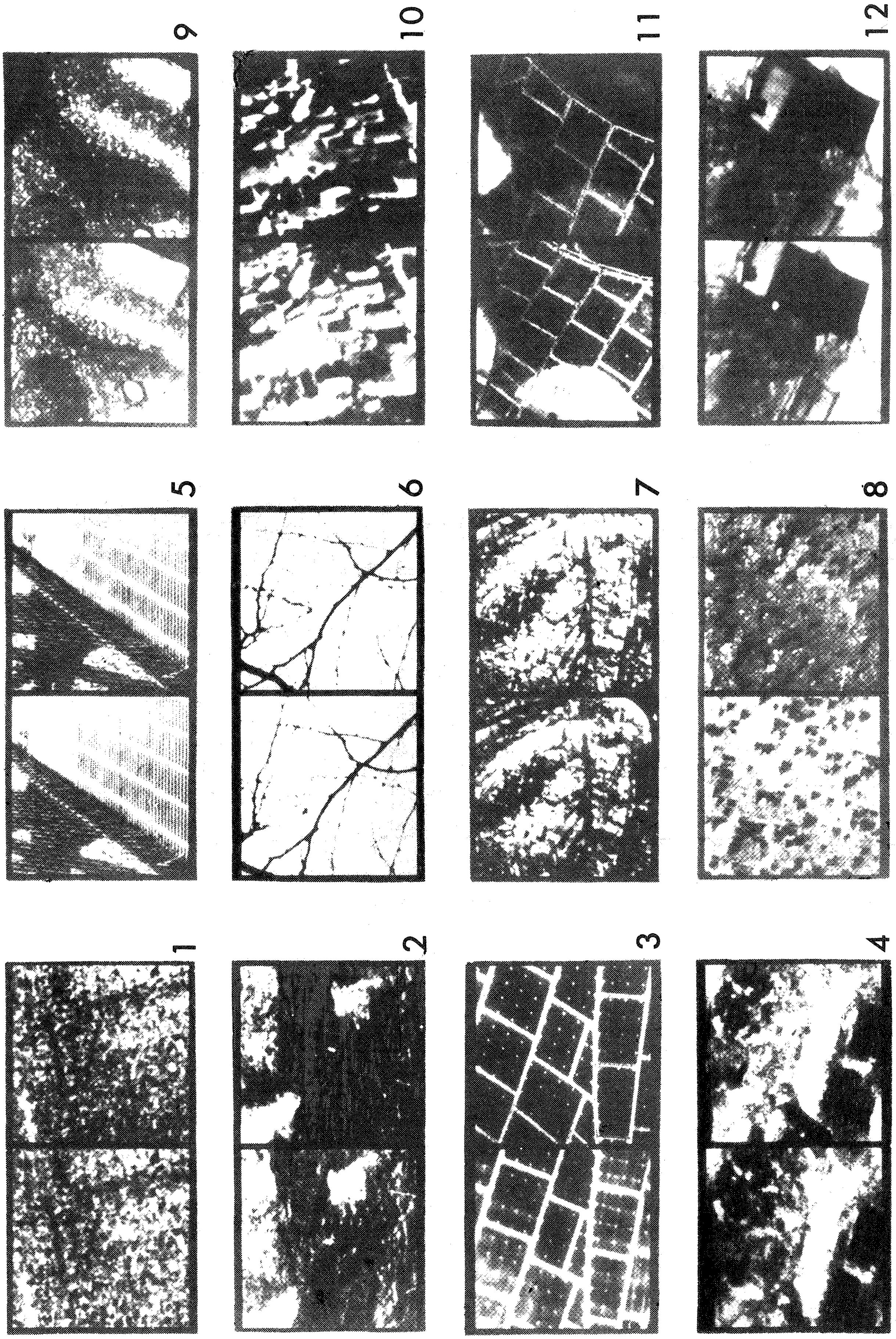


Fig. 1 shows the half tone pictures of all image pairs. They are ordered according to table 2. The vertical order is done by the image scale in four classes. The horizontal order is done by divison into three complexity classes provided by the initiators.

Four image pairs are windows of "normal images", i.e. they lie in one plane parallel to the basis between the two projection centres. Only horizontal or x-parallaxes had to be established in these cases. For the other eight images the transformations from the distributed images into normal images were given. Thus all image pairs could be used by procedures which are based on the "normal case", exploiting the advantage of the epipolar geometry.

To each image pair possible products of the matching and applications are given.

3. ALGORITHMS AND STRATEGIES

Several algorithms have been applied to the image data. Out of the received descriptions of algorithms and answers to the questionnaires it was tried to classify and order the applied techniques on several subjects. The main idea was to classify into type of matching and into strategy used by the algorithms. The classification is described in the following and visualized in table 3.

Type of matching

The algorithms are divided into area based and feature based techniques with possible subdivisions. The Cross Correlation, Least Squares Matching and Simulated Annealing techniques form the group of area based matching procedures. Cross Correlation is divided in pixel and subpixel results gained by parabolic interpolation or least squares fitting. The Least Squares Matching has local and global aspects if the consistency area is taken into account. The other group are the feature based matching algorithms using point like features (blobs, corners) or edges, straight edge segments, or zero crossings, as well as regions.

Some of the area based matching methods used interestoperators to get approximate values. This was often done in one image only and no feature based matching is performed.

Strategy

The strategy used is an important subject if the reliability of the algorithm has to be checked. Some of the algorithms need interactive starting point(s). One strategy is to use the same algorithm on different hierarchy levels, whereas others combine several algorithms to different hierarchies or within one resolution level. To get approximate values for the next match point, not only the hierarchy in resolution or algorithms is used, but also prediction in one resolution level using the previously matched point(s). Some images were handled by combination of results of several algorithms, without direct connection between each other, in these cases the human interaction is quite high, because the decision which algorithm to take for what subproblem is not made by machine. The human interaction concerning starting points, or final check was described very seldom in detail.

Interpolation

The way how to get a dense disparity map, heavily depends on the resolution of results. Most applied algorithms provide sparse results, i.e. an interpolation has to be done external. Some algorithms provide pixelwise result, where the interpolation is part of the procedure.

Relative orientation/epipolar necessary

For some algorithms the a priori knowledge of the relative orientation of the images, respectively the epipolar geometry is necessary. Some use it only for final check, e.g. check for remaining y-parallaxes in epipolar image. Others use epipolar constraints too, but derive the orientation parameters out of the images without any a priori knowledge.

Table 3 ALGORITHMS AND STRATEGIES

APPROACH NR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MATCHING																				
AREA BASED																				
* CC pixel				X				X				X				X				
subpixel	X	X	X		X		X		X	X	X									
* LSM local									X	X	X	X	X	X		X				
global														X						
* Simulated Annealing															X					
FEATURE BASED *points			I	I	X					I			X							X
*edges - straight edge segm.																	X			
- zero cross.																X		X	X	X
*regions						X														
STRATEGY																				
INTERACTIVE (start point)		X					X	X			X	X		X						
HIERARCHY * none	X	X				X	X	X									X	X		
* same alg.			X	X					X	X					X				X	X
* diff. alg.					X						M	M	X	M		X				
PREDICTION (in level)		X						X	X	X		X	X							
RESOLUTION * pixel														X	X			X	X	
* sparse	X	X	X	X	X	X	X	X	X	X	X	X	X			X	X			X
REL.ORIENT. NECESSARY	X						X	X	X	X	X	X	X	X	X	X	X	X	X	X
(final check)		*		*	*		*				*									
SURFACE MODEL (local)	—						/							/\	_⌐		\/\			?

CC Cross Correlation
LSM Least Squares Matching
I Interestoperator (no matching)
M Included manual interaction (start point(s))

Surface model: (local)

—	horiz. plane	_⌐	step + horiz. planes
/	tilted plane	\/\	step + tilted planes
/\	breaks		
?	no information		

Surface model

Out of the received information it was tried to classify the algorithms to the used local surface model. Three main models can be derived, the horizontal plane, a tilted plane or a surface with breaks and steps. Due to the sparse information delivered there might be some misunderstanding.

The table 3 shows the applied algorithms and strategies classified to the above mentioned themes. There are not exactly 18, because some of the participants used totally different algorithms to different images. On the other hand some participants used a similar or even the same approach.

The horizontal order is mainly done by the local surface model applied. A further order is reached by type of matching and applied strategy.

4. ANALYSIS

In the analysis the received data was checked computationally or by human operator. In table 4.1 an overview on the checked results is given. They can be devided in standard tasks and the special tasks.

Table 4.1 DISTRIBUTION OF CHECKED RESULTS ORDERED BY TASK

Image	TASK A (FILES)	TASK B (FILES)	SUM AB	SPECIAL TASKS
1	9 *********	16 ****************	25	
2	6 ******	6 ******	12	
3	2 **	3 ***	5	2 **
4	5 *****	6 ******	11	
5	1 *	1 *	2	2 **
6	0	3 ***	3	
7	6 ******	9 *********	15	
8	1 *	2 **	3	
9	1 *	2 **	3	
10	6 ******	8 ********	14	
11	1 *	1 *	2	2 **
12	2 **	4 ****	6	1 *

In a first part of this chapter there is given the information on the performed special tasks, the solutions and the results.

Image pair 1 is different from the others, because this image was created artificially. In the following the way of creating the image pair is described together with a theoretical evaluation of the parallax precision. For image pairs 2-12 which are checked by operator the accuracy of the operator's measurement is given.

The results of all received task A and task B files were evaluated by computational or stereo measurement check. The precision and reliability performance is described in 4.1. Section 4.2 provides the applied methods for selfdiagnosis and information on their success.

Special tasks

Only three participants tried to solve one of the proposed special tasks. For some of the tasks they used manual interaction to a great extent, i.e. not only starting point, but approximate values for all points. Those results were not checked, because this is out of the scope of this test. The special tasks show interesting single results, but they are not included in the investigations concerning precision and reliability. This is only done by task A and task B results.

Table 4.2 SPECIAL TASKS (TASK,SOLUTIONS,CHECKS)

Img.	Special task	Solved by	Check
3	* location of buttons * edge at the surface (segmentation) * borderlines of the plexiglas patches mensuration of area of two patches (segmentation)	* edge correlation (DP) * interactive +LSM * edge correlation (DP) * edge correlation (DP) + LSM	operator visual ok a) not checked operator visual ok a) checked by operator b) measurement of area Operator Particip. 1) 2.95 m^2 2.83 m^2 2) 6.12 m^2 6.11 m^2
5	* segmentation into planes, parameters of the intersection line in the images	* edge correlation (DP) * segmentation with gray val. features, relax. estim. prec. 4-5 [pel]	not checked graphical check ok c)
11	* location of buttons * surface border (segmentation) * borderlines of the plexiglas patches mensuration of area of two patches (segmentation)	* edge correlation (DP) * interactive +LSM * interactive + edge correlation (DP) * edge correlation (DP) + LSM	operator visual ok a) not checked not checked checked by operator b) measurement of area Operator Particip. 1) 5.45 m^2 5.35 m^2 2) 2.79 m^2 2.68 m^2
12	* location of house corner points, roof	* edge correlation (DP) + interactive meas. of corner points	not checked

DP: Dynamic Programming LSM: Least Squares Matching

ad a): the location of buttons and the location of the surface edge on the olympia roof were checked by human operator. The results are within the range of one pixel.
ad b): The location of the borderlines of two plexiglas patches and the mensuration of area was checked by independent human operator stereomeasurement and computation of area. The results suit quite well.
ad c): The parameters for the intersection line were graphically checked in the left image, confirming the participants estimated precision of 4-5 pixels reached by a region segmentation and relaxation technique.

Image pair 1

For image pair 1 (Car I) the true correspondences are known in order to be able to estimate the absolute precision of the matching procedures without relying on human stereopsy.

Image pair 1 was derived from one image only using the epipolar constraint and a model of x-parallaxes. The left image was created directly from a scanned image by only adding gaussian distributed noise (SigmaGL=2 gray values). The right image was computed by interpolating the parallaxes from the parallax model, adding a linear radiometric transformation and noise to the grey values (SigmaGR=5.0 gray values).
The LSM Algorithm applied by several participants directly provides for an estimation of the noise and the accuracy of parallax determination.

The introduced noise is $SigmaN=SQRT(SigmaGL^2 + SigmaGR^2)= 5.4$ [GV].

Three participants evaluated the following results:
SigmaN(estim.) = 4.8 [GV] (mean of 437 correlations)
SigmaN(estim.) = 4.4 [GV] (mean of 10 correlations)
SigmaN(estim.) = 4-7 [GV] (depending on window size)
which suit quite well to the introduced noise.

The theoretical precision of parallax measurement (SigmaX) can be estimated from the image material.
The gradient in x-direction (G_x) is given by

$$G_x = \frac{G(x_{i+1}) - G(x_{i-1})}{2} \quad \text{with} \quad G = \text{gray value},\ G_x = \text{gray value gradient in x}$$

and the variance of the gray value gradients by

$$(SigmaG_x)^2 = [G_x * G_x] / N \quad \text{with} \quad [\ ] = \text{sum},\ N=239*239$$

The theoretical precision of parallax (SigmaX) for a window of w pixels can be written as follows (ref. Foerstner 1986):

$$(SigmaX)^2 = \frac{1}{w} \frac{(SigmaN)^2}{(SigmaG_x)^2}$$

According to this formula a theoretical precision for parallax can be derived out of e.q. the left image under consideration of the introduced noise. To compare it with participants results the following table 4.3 shows the derived theoretical precision, the participants expected precision and the rms between the parallax taken from the parallaxmodel and the participants result. All results are in [pel]. The number (n) of matched points by the participants is appended.

Table 4.3 ESTIMATED PRECISION OF PARALLAX

	SigmaX (estimated) [pel] theoretical	participant	RMS [pel]	n
Window w=11x11	.05	.04	.16	437
Window w=11x11	.05	.06	.20	10
Window w=19*19	.03	.03	.21	512

Two conclusions can be drawn. The theoretical precision is reached by the participants estimation because they used selected points with a higher variance of gradients than the mean value derived from all image points. Secondly the estimated precision is too optimistic by a factor of 3 to 6 compared to the rms value, which includes further only the parallax interpolation. This performance of the estimation is already reported by several researchers too. Nevertheless the ability to estimate the parallax precision was used quite often with great success by the participants as a useful criteria for selfdiagnosis to eliminate outliers already during the matching.

Image Pairs 2-12

For these image pairs the transparencies were used to check the reported correspondences manually, using an Analytical Plotter (Planicomp C100), a photogrammetric measuring device of high precision of about 2-4 micrometer in image scale, which refers to about 1/10-1/5 of a pixel in the test images. The operator kept the point in the left image fixed and measured the parallax by human stereopsy. This parallax is then compared with the participant's result. All together more than 23,000 parallaxes have been compared in that way. If the delivered grid in task A files was to dense, a 10 pixel grid for check was choosen to reduce the amount of data. From task B files all points were measured, except when they exceeded 500, then about 500 points were randomly choosen out of the data.

To get an impression of the accuracy of the operator, for image pairs 2/3/4/5/7/8/9/10/11/12 randomly choosen participant files were measured twice by operator. There were between 92 and 414 points per file. From the both operator results the following accuracies for a single parallax measurement (not differences) can derived (see table 4.4). The RMS is computed without blunders and given in pixel and micrometer.

Table 4.4 PRECISION OF PARALLAX MEASUREMENT BY OPERATOR

	RMS (single parallax) [pel]	[10^{-3}mm]	Blunders [%]
min	0.15	2.2	0
mean	0.20	3.6	2.2
max	0.28	5.4	5.2

The results differ from image to image with worst results in image pair 4 due to the low contrast.

4.1 Precision and Reliability

To get an overview of the reached accuracy, the matching results were compared by computer with the parallax model, or stereoscopically by human operator in all other image pairs. The accuracy can be expressed in terms of precision and reliability. The following measures are derived out of the parallax differences.

The median (MED)* of all parallax differences in a file was computed to get a robust estimation of the mean value.

To estimate the standard deviation, the median absolute difference (MAD) was evaluated.

$$MAD = MED(|d_i - MED^*|) \qquad d_i = \text{parallax difference}$$

The standard deviation can be estimated by Sigmad= 1.5*MAD. The parallax differences were expected to be normally distributed. The thresholds (MED*-4.5*MAD) and (MED*+4.5*MAD) were applied to the parallax differences. If the MAD was below 0.1 [pel] the thresholds were set to +/-0.5 [pel].

All differences below, resp. above the thresholds were taken as blunders and the percentage was computed, which gave information on the reliability of the result. If the percentage of blunders is low, the result is more reliable compared to a high percentage of blunders.

From remaining differences lying between the above mentioned thresholds, the Root Mean Square (RMS) was computed, giving a measure for the reached precision.

To get a summarized overview two types of classifications were performed. Each file was classified in one of three precision classes by the algorithm used for the final match (cf. table 3).

Precision classes (PRE)				marked by
highly precise	: -	Least Squares Matching	(LSM)	
		Cross Correlation (CC)	subpixel	**blank**
precise	: -	Cross Correlation (CC)	pixel	
		Simulated Annealing		+
		Feature Based Matching	points/edges	
less precise	: -	Feature Based Matching	regions	o

A similar classification was done for the comparison of the algorithms with the percentage of blunders. Three classes of strategy were introduced.

Strategy classes (STR)		marked by
	- no hierarchy	**blank**
strategy	- hierarchy with same algorithm	=
	- hierarchy with different algorithms	#
	+ possible human interaction (start point)	**M/#M**

In table 4.5 the minima, median and maxima values of the rms of the parallax differences of all files in one class are listed according to the image pair. The same is done in table 4.6 for the percentage of blunders. To the numerical values of the results the number N of files per class is added. In table 4.5 also the number of checked points is given. It is extremely high for image pair 1. There all received results could be compared because no operator measurements were needed. Each table also provides the minimum, median and maximum value of rms, resp. blunders of all files per image (under 'SUM'). These summarized results of tables 4.5 and 4.6 are visualized in fig. 2 to get an impression of the performance on different images. The range and median value of rms and blunders is given for each image.

Table 4.5 PRECISION(RMS) AND PRECISION CLASSES

PRE	IMAGE PAIR 1	2	3	4	5	6	7	8	9	10	11	12
	.16	.18	.38	.73			.21			.26		.32
	.21	.28	.49	.81	.34	.14	.40	.32	.15	.47	.64	.43
1	.59	1.08	.59	1.64			.62			.73		.89
	13	8	2	3	1	1	7	1	1	3	1	3
	.24	.70	.90	.76			.20	1.01	1.35	.52		.63
2+	.77	.77	1.07	5.74	5.44	.75	.45	1.80	3.88	.80	.69	1.01
	1.79	1.61	1.23	29.51			2.22	2.58	6.41	7.23		1.39
	12	3	2	7	1	1	7	2	2	10	1	2
3o	-	26.54	1.67	12.82	-	6.10	2.78	-	-	.66	-	5.42
		1	1	1		1	1			1		1
S	.16	.18	.38	.73	.34	.14	.20	.32	.15	.26	.64	.32
U	.30	.55	.90	5.30	2.89	.75	.42	1.01	1.35	.70	.67	.76
M	1.79	26.54	1.67	29.51	5.44	6.10	2.78	2.58	6.41	7.23	.69	5.42
	25	12	5	11	2	3	15	3	3	14	2	6
P	27125	3628	1167	3571	686	453	4587	1020	973	5422	287	993

Table 4.6 RELIABILITY (BLUNDERS) AND STRATEGY CLASSES

STR	IMAGE PAIR 1	2	3	4	5	6	7	8	9	10	11	12
	4.3	.0	13.5	.0		.0	14.2			9.4		6.7
A	8.2	4.5	16.8	0.0	-	3.5	18.9	21.1	30.1	16.4	21.4	9.4
	19.7	18.4	20.0	27.9		7.0	27.1			40.0		12.1
	6	3	2	5		2	5	1	1	5	1	2
	2.1	.2	3.6	5.4	4.4		6.2	20.9	19.1	4.1		5.4
B=	3.2	7.2	15.9	9.5	5.0	10.1	10.2	24.7	20.9	5.7	-	13.4
	14.2	33.0	28.2	24.4	5.5		27.4	28.5	22.7	10.4		21.3
	9	6	2	4	2	1	6	2	2	7		2
	.4	1.8		7.7			4.6			4.4		11.4
C#	2.4	1.9	8.5	12.7	-	-	7.3	-	-	5.5	3.4	14.5
M	7.9	11.6		17.6			16.8			6.6		17.5
	10	3	1	2			4			2	1	2
	M	M		M			M			M		M
S	.4	.0	3.6	.0	4.4	.0	4.6	20.9	19.1	4.1	3.4	5.4
U	3.2	5.6	13.5	10.8	5.0	7.0	14.2	21.1	22.7	8.4	12.4	11.8
M	19.7	33.0	28.2	27.9	5.5	10.1	27.4	28.5	30.1	40.0	21.4	21.3

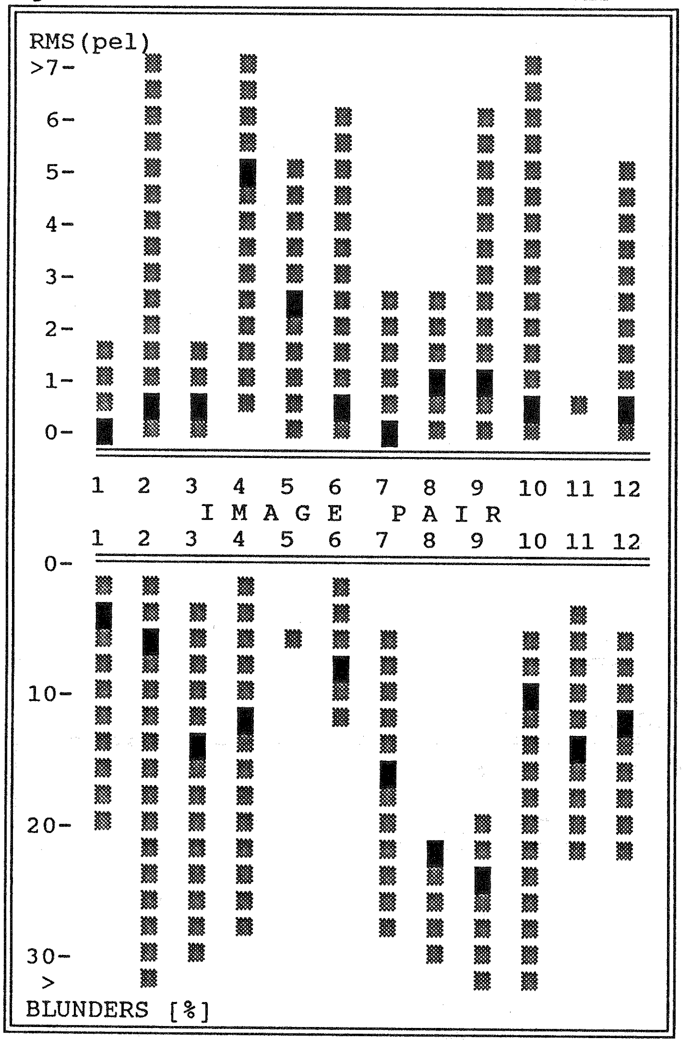
Explanations for table 4.5 and 4.6 :

Numerical Values			CLASS					
RMS		BLUNDERS	PRE(cision)				STR(ategy)	
[pel]	min	[%]	high precise	BLANK	1	A	BLANK	no hierarchy
[pel]	<u>median</u>	[%]	precise	+	2	B	=	same algorithm
[pel]	max	[%]	less precise	o	3	C	#/M	dif. algorithms [man. interact.]

N: number of files

SUM: minimum, median, maximum values P: number of checked points

Fig. 2 RMS AND BLUNDERS FOR ALL IMAGES



■ = median ▒ = range

TABLE 4.7 PRECISION CLASSES

PRE	N	RMS [pel]	BLUNDERS [%]
		.14	.2
1	44	**.35**	**6.1**
		1.64	17.6
		.20	.0
2 +	50	**1.71**	**13.1**
		29.51	33.0
		.66	.0
3 o	7	**8.00**	**11.6**
		26.54	40.0

TABLE 4.8 STRATEGY CLASSES

STR	N	RMS [pel]	BLUNDERS [%]
		.20	.0
A	33	**3.51**	**10.0**
		29.51	40.0
		.14	.2
B =	43	**.65**	**8.9**
		5.44	33.0
		.16	.4
C #	25	**.88**	**5.6**
M		7.23	17.6

for table 4.7 and 4.8 :
[pel] min (median) [%]
[pel] **mean**(median) [%]
[pel] max (median) [%]
N number of files

Fig. 3 PRECISION CLASSES

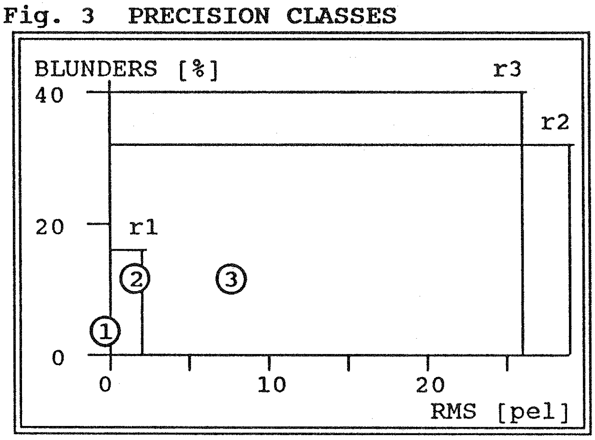
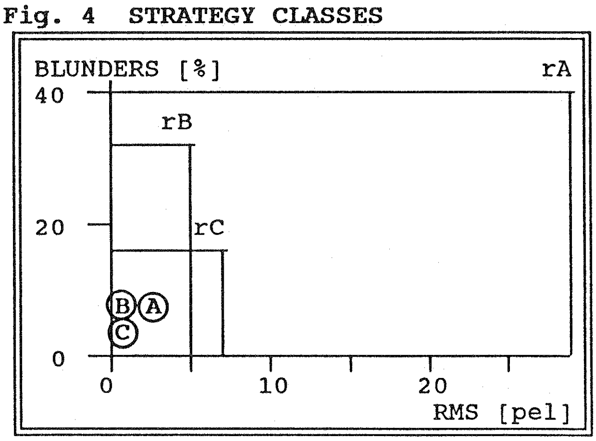
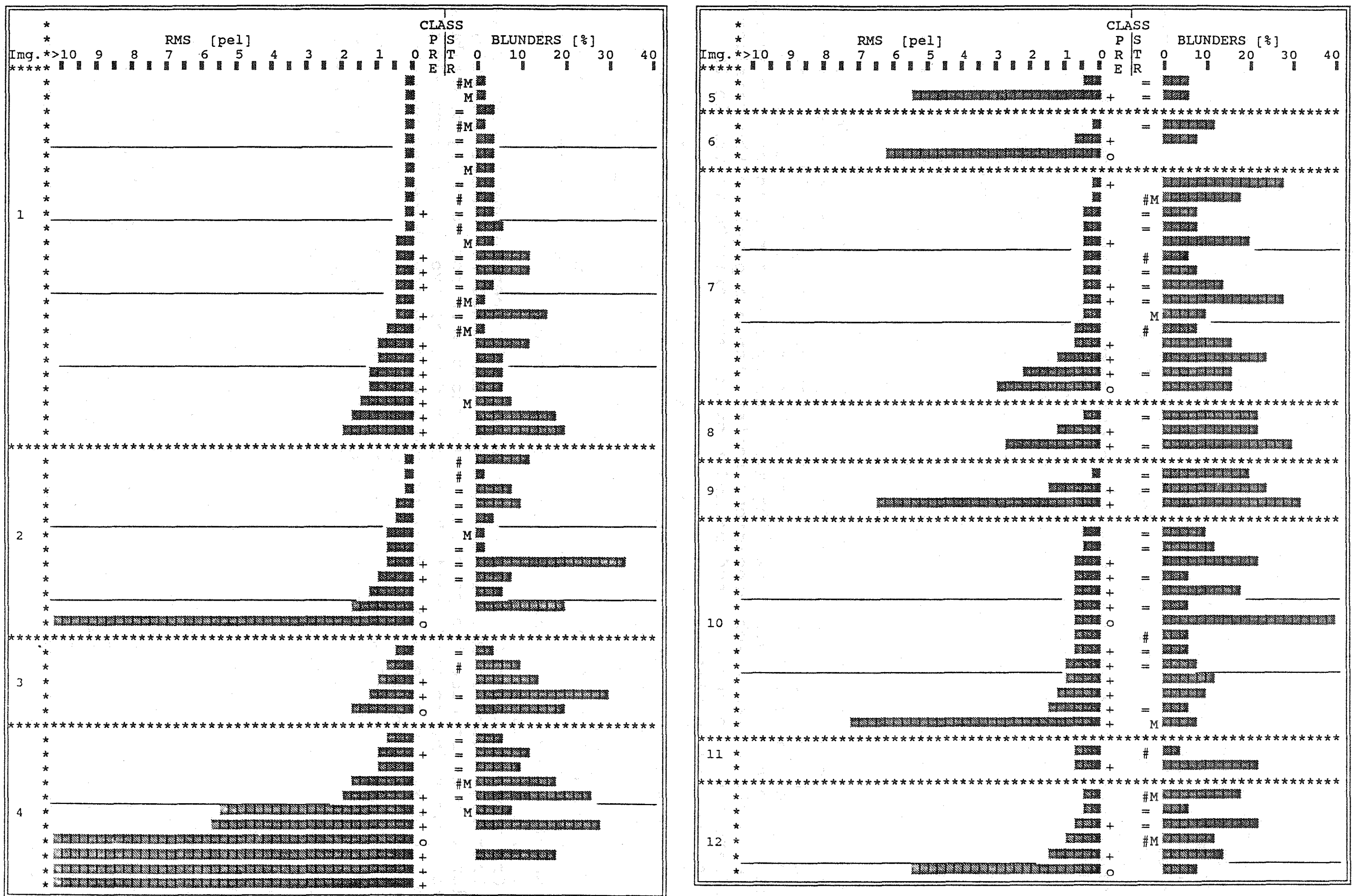


Fig. 4 STRATEGY CLASSES



r range of rms/blunders ◯123/ABC meanpoint in precision/strategy class

Fig. 5 RESULTS PER IMAGE (FILES ORDERED BY INCREASING RMS)



Another subject is the performance on all images of the algorithms combined in one class. Table 4.7 shows the performance of the precision classes. The minimum, weighted mean and maximum value of the median values on the different images is given per class. This is done for rms and blunders. This table is visualized in fig. 3, giving the range and mean of the different classes. In similar form Table 4.8 shows the performance of the strategy classes. This table is visualized in fig. 4.

To get a more detailed information on the results all files are listed with their according rms and blunder value for each image pair. Fig. 5 shows the single results ordered per image by increasing rms. The marks for precision and strategy class are added to each file.

COMMENTS

The different image pairs show different results. This agrees with the expectation and with the human operator's measurements. The medianvalue of the rms of all files per image ranges between 0.3 and 5.3 pixels, and the medianvalue of the blunders between 3.2 and 22.7 percent. The best analysis can be drawn from images 1,4,7 and 10 which were taken most often, due to the already performed transformation into normal images. The low texture in image pair 4 provided some difficulties to several algorithms with greater problems for the edge based methods.

Image pair 1 shows in general the best results. The main reason is the low complexity combined with good texture for matching. Secondly no operator measurement is included in the rms of the parallax differences. Only the linear interpolation of parallaxes out of the parallaxmodel used for creating the right image and the participants result contribute to the rms.

The precision of the algorithms on the other image pairs can be computed from the rms of the parallax differences given in the figures and the mean value of the precision of the operator's measurement given in table 4.4.

The precision results of the expected high precise algorithms are in general on subpixel level reaching also the operators precision. The feature based methods also reach precisions better than one pixel in the parallax differences. The region based method provides approximate values in the order of 1 to 20 pixels.

One approach kept the continuity constraint too fix, therefore especially on image 4 and 7 there was a systematic difference on the part of the image where the object surface decreases rapidly. The participants result was adapted to the larger horizontal part of the image, taking the model of a horizontal plane for the rest of the image too.

The reliability of results is better if there is applied hierarchy either in form of different algorithms or in form of same algorithm on different resolution levels. The manual start in some cases not always provides for better reliability.

One approach without hierarchy has a quite low percentage of blunders but rather big values in RMS. In the analysis the model of blunder detection didn't work, because the parallax differences were not normally distributed. From grid point to grid point the delivered parallaxes differ from -30 pixel to +30 pixel, single points showing subpixel accuracy. Here the neighborhood should have been taken into account to check for consistency.

Local and global LSM show differences in precision and reliability. Local LSM shows significantly better precision results than the global approach, using the local LSM as approximate values. The reliability of the global approach on the other hand was slightly better than in the local case.

4.2 Selfdiagnosis

The quality assessment of the results was performed by several participants. It was done either automatically or manually or in combination of both. For some files no information was given. Only the measures received are described. The detailed information of internal tuning parameters and checks can be derived from the literature.

The participants choose different ways to perform the selfdiagnosis. Some used manual check of the results. Quite often was used graphical display to check for gross errors in the result. This can be done e.g. with raster monitor and overlay display in red and green (anaglyph). Others look 3-D perspective views, overlay of matched points on the gray value images, plots of matched and unmatched primitives or other visual descriptions provided by their algorithms.

Some approaches performed an automatic check, by comparing certain values to a given threshold. The threshold was either kept fix for all image pairs out of the experience with other data, or it was adapted to the problem.

The following measures for quality of match were used:

- normalized cross correlation coefficient with minimum threshold 0.35-0.70
- square of normalized cross correlation with preservation of sign with minimum threshold 0.3
- ratio of correlation and autocorrelation with minimum threshold of 0.8
- estimated precision of parallax with maximum thresholds of some 1/10's of a pixel
- edge stength

But also checks after the matching were performed making use of the relative orientation parameters, to test for remaining y-parallaxes in the normal image, which would indicate problematic correspondencies.

The result of the selfdiagnosis was either used to distribute 'good' parallaxes only, or it was given verbal or in form of numerical values or in combination.

Two typical verbal answers were:

' Good results everywhere exept along the lower right border'

' Results poor or disastrous'

supported also by photographs, 3-D descriptions, plots of matched and unmatched points or other features.

The numerical results were connected to each parallax, either transformed to a good/bad indicator or given in the original form or in combination.

Out of the descriptions and the data files it was tried to derive summarized information on the ratio of manual, automatic and no diagnosis. For each image is also given a quite rough decision if the provided verbal or numerical results were too optimistic, consistent or too pessimistic concerning the comparison of the results with the operator measurement. Table 4.9 shows in very condensed form the result of this analysis. The numerical values were in general a good indicator for the quality of a matched point. Some of the measures were sometimes slightly too optimistic, one measure didn't work in any case.

The threshold for cross correlation coefficient lower than 0.5 often resulted in poor results. But also a threshold of more than 0.7 couldn't prevent false matches in the data. Good performance showed the application of estimated parallax precision.

Table 4.9 SELFDIAGNOSIS AND RESULT

IMAGE	SELFDIAGNOSIS auto Comb. man ? None	RESULT VERBAL	RESULT NUMERICAL
1	aaaaaaaaaaaaaa CCC m ?? NNNNN	OOOOOOOOOOO	-- OOOOOOOOOOO + ####
2	aaaaa CCC m ?? N	OO +++ #	- OOOOO +
3	a m ?? N	- OO #	O
4	aaaaa CC m ?? N	OOOOOO #	OOO + ###
5	a ?	O +	O
6	a m ?	- OO	O
7	aaaaaaa CCCC m ?? N	OOOOOOO	- OOOOO ###
8	a m ?	OO +	O
9	aa m	OOO	O +
10	aaaaaaa CC m ?? NN	OOOOOO +	OOOO + ###
11	C m	+	
12	aa CC m ?	- O +	- O ++

a	automatic	=	too pessimistic
m	manual	-	slightly pessimistic
C	combination (a,m)	o	o.k.
?	not clear	+	slightly optimistic
N	none applied	#	too optimistic

5. CONCLUSIONS

The test on image matching was successful. A wide range of different approaches ran the test images, even images which are out of the main field of their usual application. This shows the high flexibility of the algorithms to react on images of different type.

The accuracy of the applied algorithms is high. Out of the test results it can be derived that precision and reliability performance have to be seen together.

Except the region matching which by definition just provides very coarse approximate values, all other algorithms were more or less able to reach a precision of 1 pixel or less. Some outliers resulted from different reasons, e.g. from unproper choose of tuning parameters or missing internal consistency checks. It has also been noticed that interestoperators should work on subpixel level, i.e. not only subpixel parallax should be submitted, but also subpixel position in left image, if this is not done, then interesting points e.g. in image 6(tree) are located in the sky and not on the branches.

The reliability performance was different for different strategies. In general can be derived that hierarchical performance is advantageous, also for this small area of 240 by 240 pixels. If the same algorithm is used in several levels, backtracking, matching left right, right left was used with good success.

The manual start is not unproblematic, it is no guarantee for good results. The manual interaction was used to solve the problem of approximate values. Some methods applied in the test are very suitable under certain conditions to solve this problem.

In strategy two extrema occured, one very sophisticated algorithm, where no human interaction is needed, where internal consistency checks are performed in level and between levels, and another one where the operator just put together several suitable algorithms, which for themself worked really well. In the later case still work has to be done to perform the human decision automatically.

In general the methods of selfdiagnosis work but there is still a great amount of manual checks which should be replaced by automatic means.

Each of the strategies 3,17 and 18 were applied to more than 10 out of 12 image pairs by only changing some tuning parameters. The results ranged from excellent to disastrous, but one algorithm solved 11 image pairs with reasonable results in all images, using hierarchical performance.

The a priori knowledge of relative orientation parameters is quite often needed for the algorithms, but the parameters can also with good success be automatically derived from the image data.

The amount of measured points, necessary to represent the object, totally depends on the object form. Sometimes a clustering of matched points, distributed only on a part of the image was noticed. To solve this problem one idea is to work with subimages or set a fixed number of matched points per epipolar line for example, both were applied with success.

In image 10 often there are some good points measured, which might not be enough for detailed description. Advantageous is here an algorithm with more dense results, because interpolation would be quite difficult. One approach yielded information about the disparity discontinuities used for modelling the object in the matching process.

The high complexity of images 8,9,10,11 and 12 caused quite a few problems which were circumvented or only partly solved. Occlusions or the transparency surface in image 11 e.g. couldn't be handled up to now.

Further research has to be done to solve these complex problems. Approaches to their solution are already available. The tendency is going to hierarchical solutions with combination of several algorithms and automatic decision of their use. The connections between the different disciplines, their algorithms and their solution techniques should be combined to get new ideas to solve the 2D-matching problem for general surfaces.

Acknowledgements:

The author likes to thank first of all the participants for their results which were the basis for the success of this test and secondly several members of the Insitute for Photogrammetry in Stuttgart for their support in the analysis.

The test on image matching has been initiated and assisted by Dr. W. Foerstner. Ir. G. Vosselman contributed the reading and transformation of the received tapes. Dipl.-Ing.(FH) M. Englich did all stereo measurements at the Analytical Plotter necessary for the analysis. U. Hirth helped on the evaluation of the results and the preparation of plots and lists which will be presented in the final report.

LITERATURE

ACKERMANN, F. : **High Precision Digital Image Correlation;** Proceedings 39th Photogrammetric Week, Institut fuer Photogrammetrie, Universitaet Stuttgart, Stuttgart, FRG, Heft 9, 1983

AYACHE, N., FAVERJON, B. : **Efficient Registration Of Stereo Images By Matching Graph Description Of Edge Segments;** International Journal Of Computer Vision, 107-131 (1987)

BARNARD, S.T. : **Stereo Matching By Hierarchical Microcanonical Annealing;** SRI-International, Technical Note No. 414, Menlo Park, USA, 1987

BENARD, M., BOUTALEB, A.K., KOELBL, O., PENIS, C. : **Automatic Stereophotogrammetry: Implementation And Comparison Of Classical Correlation Methods And Dynamic Programmming Based Techniques;** Proceedings ISPRS COMM. III Symposium, Rovaniemi, Finland, 1986

CLAUS, M. : **Korrelationsrechnung in Stereobildpaaren zur automatischen Gewinnung von digitalen Gelaendemodellen, Orthophotos und Hoehenlinienplaenen;** DGK, Heft 283, Muenchen, FRG, 1983

FOERSTNER, W. : **A Feature Based Correspondence Algorithm For Image Matching** Proceedings ISPRS Commission III Symposium, Rovaniemi, Finland, 1986

HANNAH, M.J. : **Test Results From SRI's Stereo System;** SRI International, Menlo Park, USA, 1987

HARTFIEL, P. : **Zur Leistungsfaehigkeit von Bildzuordnungsverfahren bei der Erzeugung Digitaler Gelaendemodelle;** Proceedings ISPRS Commission III Symposium, Rovaniemi,Finland, 1986

HENRIKSEN, K., BAJCSY, R. : **Feature Based Stereo Matching;** Technical Report DIKU-86-7, Computer Science Department, University of Copenhagen, Copenhagen, Denmark, 1986

NORVELLE, R. : **Interactive Digital Correlation Techniques For Automatic Compilation Of Elevation Data;** U.S. Army Engineer Topographic Laboratories, ETL-0272, 1981

OLSEN, S.I. : **Concurrent Solution Of The Stereo Correspondence Problem And The Surface Reconstruction Problem;** Presented at the Eighth International Conference On Pattern Recognition, Paris, France, 1986

POLLARD, S.B., MAYHEW, J.E.W., FRISBY, J.P. : **PMF: A Stereo Correspondence Algorithm Using A Disparity Gradient Limit;** Perception, 1985, Volume 14, Pages 449-470

PRICE, K. : **Relaxation Matching Techniques - A Comparison** IEEE, PAMI-7, No. 5, 1985

STRAUB, B. : **IPE-Report On Test On Image Matching;** Institut fuer Physikalische Elektronik, Universitaet Stuttgart, Stuttgart, FRG, 1987

WROBEL, B.P. : **Facets Stereo Vision (Fast Vision) - A New Approach To Computer Stereo Vision And To Digital Photogrammetry;** Proceedings ISPRS Intercommission Workshop, Interlaken,