THE AUTOMATIC INTERIOR ORIENTATION AND ITS DAILY USE

Wolfgang Schickler 1)
Zoltan Poth 2)
1) Analytical Surveys, Inc., Colorado Springs, Colorado, USA, wolfgang@anlt.com
2) Carl Zeiss, Oberkochen, Germany

Commision III, Working Group 3

KEY WORDS: Photogrammetry, Orientation, Automation, Softcopy, Matching, Digital

ABSTRACT

We present a fully automatic and operational procedure for the reconstruction of the interior orientation of digitized aerial images. The main task of the Automatic Interior Orientation (AIO) is the robust localization of the fiducial marks in the digital image and the reliable estimation of the transformation between image and pixel system. We paid particular attention to robustness of the procedure. Results of extensive tests confirm the high reliabilty of the approach and also the internal decesive measure for selfdiagnosis.

1 PERFORMANCE FEATURES

The presented approach is a generic solution for the AIO. Images from different types of cameras can be processed. The following performance features are implemented:

- The procedure is fully automatic. Except for the approximate image resolution and the camera type, no additional information or approximate values are needed.
- Gray level and color images can be processed.
- The orientation of the image is automatically recognized.
- Whether the image is positive or negative is automatically recognized.
- Robust algorithms guarantee correct results even in low contrast images.
- Reliable self-diagnosis enables automatic recognition of unsolvable situations.

The fully automatic process of the reconstruction of the interior orientation expects the following:

A The image in digital form, including
  - an image pyramid and
  - the approximate resolution of the image. The pixel size is usually known from the scanning process with an accuracy of $\pm 1 \mu m$ which is more than sufficient.

B The camera type and the usual camera calibration information, including
  - Fiducial mark patterns and
  - a pattern of an unsymmetric feature for the recognition of the orientation.

Except for the calibration data all the information specific to one camera type is stored in a so-called *camera description file*, which is available for all the conventional camera types.

2 CONCEPTUAL ASPECTS

There are eight different possibilities (orientations) for how the images could have been placed in the scanner for digitizing: wrong reading or right reading with four different 90^o rotation respectively. For a generic solution to recognize the orientation of the image we expect a unique *asymmetric feature* in the image. With a template and the coordinates in plate system of this asymmetric feature, it is possible to match the template with all eight possible positions where it could appear depending on how the image was scanned. A classification of all the matching results leads to the orientation of the image.

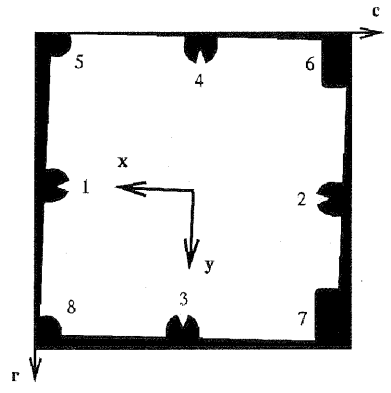

Figure 1: Schematic illustration of a digitized image

All conventional cameras have at least 4 symmetrically placed fiducial marks, so that at a certain position a fiducial mark can be found whose shape is known, independent of whether the image is right reading or wrong reading and from 90^o rotations of the image. We call those *orientation invariant*. The fiducial marks 1 to 4 in figure 1 fulfill this criterion, while the shape of the fiducial marks in the corners in this case is dependent on the orientation in the image.

Our approach is based on the location of at least four *orientation invariant* fiducial marks in the image, without any prior information. After this, the transformation between the pixel and the plate system can aproximately be estimated, except for an unkown factor of 90^o rotations, and possibly a mirror reversing. Therefore we locate an asymmetric feature to derive how the image was placed in the scanner.

As fiducial marks are 2-D objects with well defined geometric and radiometric characteristics and usually the scanned images have no more rotation than 10^o, the cross correlation is an appropriate matching strategy for locating fiducials. Fiducial marks are usually synthetically faded into the image on an unexposured, and therefore dark and homogeneous background. This model information has been integrated in the localization process by applying a binarization the image using both criteria, the low intensity and the homogeneity. An additional advantage of the binarization is the possibility of using a very efficient binary correlation.

On different pyramid levels we use different representations for the fiducials, i.e. different templates for the correlation. Fig. 2 illustrates these different representations for a RMK-TOP camera. On the highest levels we use the whole fiducial mark including its surroundings (Fig. 2 left). On the lower levels the fiducial figure (Fig. 2 middle), and only on the lowest level where the final measurement is done, the fiducial mark itself is used (Fig. 2 right). To exclude areas with undefined radiometric characteristics like the contents of the image or yet unsolved asymmetric features we use so-called *don't care* regions. These regions, shown in grey in Fig. 2 left, are not taken into consideration when correlating.

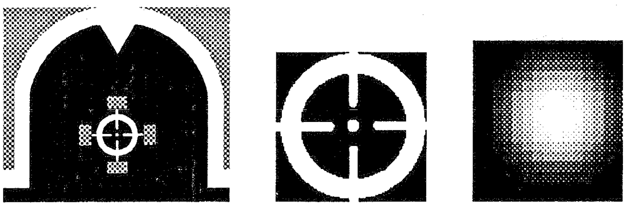

Figure 2: Left: Pattern of the fiducial mark surrounding, with grey *don't care* regions. Middle: Pattern of the fiducial mark figure. Right: Pattern of the fiducial mark

3 OVERVIEW

The whole procedure of the AIO consists of following steps:

- **Resampling of the templates.**
  According to the resolution of the image.
- **Image pyramid derivation.**
  If not yet present.
- **Robust localization of at least four orientation invariant fiducials.**
  In this main part of the hierarchical pattern recognition process an outlier detection algorithm is implemented to make the localization process more robust. It results in very good approximate positions for the orientation invariant fiducials of less than ± 5 pixels. Within this procedure the system also determines whether the image is positive or negative. The robust localization process is described in chapter 4.
- **Detection of the orientation of the image.**
  There are eight different possible orientations how to scan an image, i.e. wrong reading or right reading with four different multiple 90^o rotations respectively. To detect the correct orientation, one asymmetric feature in the image is located.
- **Fine measurement of all fiducial marks.**
  This is done using a grey level correlation with subpixel estimation. The accuracy of the individual location is about 1/10 of a pixel.
- **Estimation of the transformation parameters.**
  In this step the transformation between the plate system and the image system is estimated, and different kinds of transformation types are possible.
- **Self-Diagnosis.**
  This is done by analyzing the final result with respect to precision and sensitivity. The self-diagnosis is described in chapter 5.

4 ROBUST LOCALIZATION OF ORIENTATION INVARIANT FIDUCIAL MARKS

The principle here is to locate at least four orientation invariant fiducial marks individually using a binary correlation. As the interior orientation should be performed automatically without any approximate values, we use a hierarchical search strategy through the image pyramid from coarse to fine for the location of the rotation invariant fiducial marks.

The procedure of the robust localization which is performed hierarchically consists of the following four steps:

1. Definition of the search space
2. Binarization
3. Binary correlation
4. Consistency check

Steps 2 and 3 are replaced by a grey level correlation and a positive/negative recognition, in the case that the information on whether the image is positive or negative is not available, or as long as this recognition task has not been solved significantly.

Each of these steps is described in the following subsections.

4.1 Positive - Negative Recognition

The task here is to detect whether the image is positive or negative, which can easily be solved by analyzing the grey levels in the surrounding of the fiducial marks. The idea is to use the definition of the fiducial mark surrounding which is given by the templates. After a fiducial mark is located all the corresponding pixels in the image which are black in the template are used to calculate a mean grey level, from which the information whether the image is positive or negative can be derived.

This approach needs a localization technique which is independent from the information whether the image is positive or negative. Against previous assumptions the approach to use a binary correlation and only the homogeneity as the criterion for the binarization to first locate the fiducial marks and then do the grey level analysis, has been proven to not work reliably enough.

Therefore we now use a grey level correlation on the highest pyramid level, which will result in a negative correlation coefficient in the case the image is negative and the template positive. For all further steps the much more efficient binary correlation is used, if the grey level analysis is significantly solving the positive/negative problem.

4.2 Definition of the search space

It makes sense to adapt the search space for the fiducial marks in the i-th pyramid level to the quality of the localization on the previous $i+1$-th pyramid level. We use the σ_0 of the transformation estimation between pixel and plate system as a quality measure and the result of the consistency check to define the search space. On the highest level ($i = i_{max}$), if there are no approximate values available, the search space is set to a multiple of the size of the fiducial template $M^{(i)}$. The search space $A^{(i)}$ at the i-th pyramid level is defined as:

$$A^{(i)} = \begin{cases} M^{(i)} \cdot 4 & : \; for \quad i = i_{max} \\ M^{(i)} \cdot f \cdot \sigma_0^{(i+1)} & : \; otherwise \end{cases} \tag{1}$$

where f is set to 3, which corresponds to a 99.7% confidence region.

The consistency check (ref. 4.5) may not be solvable, if all the found positions are inconsistent. This indicates that at least 50% of the fiducials are not correctly located. In this case the search space is opened again to a multiple of the size of the fiducial template. Thus a very efficient definition of the search space is possible. If the quality of the localization on the previous level is high, the search space is small, normally only about 30 by 30 pixels.

4.3 Binarization

Corresponding to the fiducial marks i.e. in a positive image where the fiducials are black, the binarization divides the image in dark and homogeneous versus non-dark and inhomogeneous areas. The binary image B is derived from:

$$B(x,y) = \begin{cases} 1 & : \; I \leq T_1 \; ; \; \|\nabla I\|^2 \leq T_2 \\ 0 & : \; otherwise \end{cases} \tag{2}$$

with the absolute value of the squared gradient of the image $I_{(i,j)}$

$$\|\nabla I\|^2_{(i,j)} = \left[I_{(i,j+1)} - I_{(i+1,j)}\right]^2 + \left[I_{(i,j)} - I_{(i+1,j+1)}\right]^2 \tag{3}$$

The thresholds for the binarization are adaptively derived from the corresponding subsections of the image. T_1 is found using a histogram analysis and T_2 can be derived using an estimation of the noise variance σ_n [BRÜGELMANN, FÖRSTNER]. From the expectation value $E(\|\nabla I\|^2) = 4\sigma_n^2$ $T_2 = 9 \cdot \left(4\sigma_n^2\right) = 36\ \sigma_n^2$ can be found.

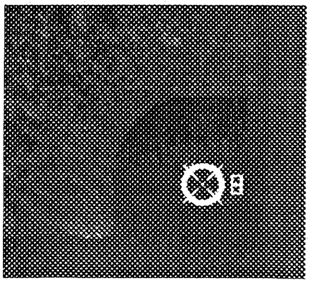

Figure 3: Subsection of an image with fiducial mark

Figure 4: Binarization using only $T_1 = 8\ [gr]$

Figures 3 to 6 demonstrate the efficiency of using both criteria for the binarization. The over- and under-segmentation in Fig. 4 and Fig. 5 respectively show how sensitive the binarization is if only the grey level is used. The thresholds only differ in 4 gray levels. Whereas Fig. 6 shows a clear segmentation when both criteria are used, independent of small changes in the thresholds.

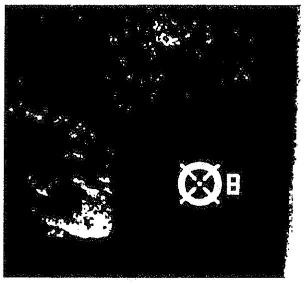

Figure 5: Binarization using only $T_1 = 12\ [gr]$

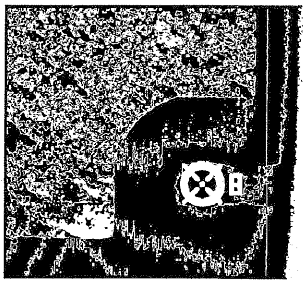

Figure 6: Binarization using $T_1 = 20\ [gr]$ and $T_2 = 5\ [gr^2]$

4.4 Binary correlation

The cross correlation [HARALICK/SHAPIRO 1993], following called *grey level correlation*, is based on the following model: Two corresponding images only differ

1. In geometry by a simple translation $T(u,v)$, and
2. In radiometry by a linear transformation in contrast and brightness.

As the binary correlation only deals with binary images, only point one is valid here.

Using a binary correlation to locate a fiducial mark, the template t is translated by $T(u,v)$. A first estimation for the position $P_{(0)}(\hat{u},\hat{v})$ of the template in the image b can be found by:

$$max_{(u,v)}\ \rho_{bt} \rightarrow (\hat{u},\hat{v})^{(0)} \tag{4}$$

with

$$\rho_{bt}(u,v) = \frac{\sigma_{bt}(u,v)}{\sigma_b(u,v)\sigma_t} \tag{5}$$

$$\sigma_{bt}(u,v) = \frac{1}{m-1}\left[\#(b \cap t) - \frac{\#b \cdot \#t}{m}\right] \tag{6}$$

$$\sigma_b(u,v) = \frac{1}{m-1}\ \#b\left[1 - \frac{\#b}{m}\right] \tag{7}$$

$$\sigma_t = \frac{1}{m-1}\ \#t\left[1 - \frac{\#t}{m}\right] \tag{8}$$

where m is the number of pixels in the template t, $\#b$ and $\#t$ the sum of all black pixels in the corresponding area of b and the template t respectively. $\#(b \cap t)$ is the size of the intersection of all the black pixels in b and t.

The estimated position $P_{(0)}(\hat{u},\hat{v})$ is integer valued, with a rounding error of 1/3 of a pixel. A sub-pixel estimation can be achieved by approximating the surface of the two dimensional correlation function ρ_{bt} by a second order polynomial in a neighborhood of $P_{(0)}(\hat{u},\hat{v})$. The final sub-pixel position is defined as the local maximum of the second order polynomial which leads to

$$(\hat{u},\hat{v})^T = (\hat{u},\hat{v})^{(0)T} - \left[H\rho\,|_{(\hat{u},\hat{v})^{(0)}}\right]^{-1} \nabla\rho\,|_{(\hat{u},\hat{v})^{(0)}} \tag{9}$$

with the Hesse matrix H and the gradient ∇ von $\rho(u,v)$

$$H\rho\,|_{(\hat{u},\hat{v})^{(0)}} = \begin{pmatrix} \rho_{uu} & \rho_{uv} \\ \rho_{vu} & \rho_{vv} \end{pmatrix} |_{(\hat{u},\hat{v})^{(0)}} \tag{10}$$

$$\nabla\rho\,|_{(\hat{u},\hat{v})^{(0)}} = \begin{pmatrix}\rho_u\\ \rho_v\end{pmatrix}|_{(\hat{u},\hat{v})^{(0)}}$$

The accuracy of this estimation is given by the covariance matrix

$$D\begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} = \frac{1}{m}\cdot\frac{1-\rho_{bt}}{\rho_{bt}}\left[-H\rho\,|_{(\hat{u},\hat{v})^{(0)}}\right]^{-1}\cdot\Delta x^2 \qquad (11)$$

where

m	is the number of pixels of the template t,
ρ_{bt}	is the similarity of the image and the template at $(\hat{u},\hat{v})^{(0)}$,
$H\rho$	is the roughness of the texture of the signal (ref. equ. (10)) and
Δx	the size of a pixel assumed to be identical in row and column.

The sub-pixel and the accuracy estimation are only based on the analysis of the correlation function, and can therefore be used for the binary as well as for the grey level correlation.

4.5 Consistency check

The result of the individual localizations on each pyramid level is checked using a consensus criterion to detect outliers and, if necessary, to predict a more likely position for the outliers. The outlier detection is similar to the RANSAC technique proposed by [BOLLES R. C. / FISHLER M. A. 81]. With a minimal set of observations a similarity transformation between pixel and plate system is estimated. This transformation is used to check the remaining observations based on remaining errors. In contrast to the RANSAC we do a complete search for the 'best solution' because the number of observations is small. The best solution is defined as the transformation having the smallest remaining errors. This transformation is used to detect outliers and eventually to predict a more likely position for the fiducial mark in the image.

5 SELF-DIAGNOSIS

It is important for each automatic system to be able to make a selfdecision on the acceptability of the result. Automation needs predictable results.

The principle of *Traffic Light Programs* proposed by FÖRSTNER 1994 classifies the result in three different states:

- **red:** The system was not able to solve the problem, or the found solution has been classified as incorrect. The system gives reasons for the failing.
- **yellow:** The correctness of the solution is doubtful, the system gives a warning including a certainty of the correctness and a diagnosis of possibly correct and incorrect parts.
- **green:** The found solution is verified as being correct.

For the control of the traffic light an objective quality control measure is necessary. The next section introduces the control measure we use to classify the result in these three stages.

5.1 Sensitivity Analysis

Gross errors can hide behind small residuals or excellent fitting of data and model, therefore they do not necessarily produce large variances in the estimated parameters. Consequently, an additional sensitivity analysis for self diagnosis is used, to classify the result.

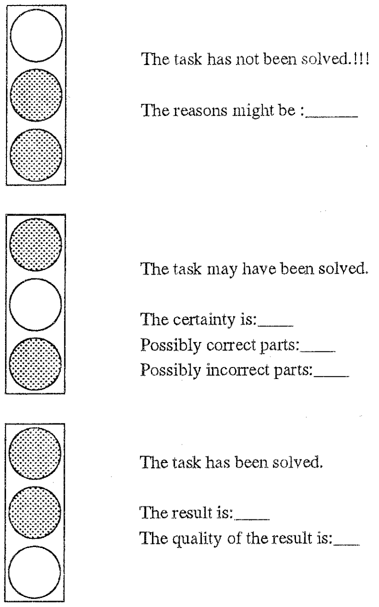

Figure 7: The principle of Traffic Light Programs

The concept of sensitivity analysis developed by Baarda [BAARDA W. 67 ,68] is based on the measures for the internal and external reliability. The elementary theory has been expanded and specified for our purpose [cf. FÖRSTNER W. 83, 92].

Here the sensitivity analysis is used to investigate the influence of the observed position of each fiducial or each combination of two fiducials onto the estimated transformation parameters. A single fiducial is represented by two coordinates, a combination of two fiducials is represented by four coordinates. Therefore the sensitivity analysis is applied to groups of observations. The sensitivity measures are as follows:

The **empirical sensitivity**

$$\bar{\delta}_i = T_i\cdot\mu_i \qquad (12)$$

(internal reliability according to Baarda) measures the maximum influence of the observation group i to the estimated parameters. If this group is omitted an arbitary function $\mathbf{f} = \mathbf{a}^T\cdot\mathbf{y}$ with variance $\sigma_f^2 = \mathbf{a}^T\Sigma_{yy}\mathbf{a}$ of the estimated parameters y does not change more than

$$\nabla_i f \quad \leq \quad \bar{\delta}_i\cdot\sigma_f \qquad (13)$$

The **theoretical sensitivity**

$$\bar{\delta}_{i0} = \delta_0\cdot\mu_i \qquad (14)$$

(external reliability according to Baarda) gives the maximum influence of undetected errors in observation group i onto the estimated parameters. The influence of an undetected error in the observation group i is bounded by

$$\nabla_{0i} f \quad \leq \quad \bar{\delta}_{0i}\cdot\sigma_f \qquad (15)$$

where

the **influence factor** is given by

$$\mu_i^2 = \lambda_{max}^2 \left[\left(\boldsymbol{\Sigma}_{yy}^{(i)} - \boldsymbol{\Sigma}_{yy} \right) \boldsymbol{\Sigma}_{yy}^{-1} \right] \qquad (16)$$

$$(17)$$

This measures the maximum relative increase in uncertainty of the estimates y if observation group i is omitted. $\boldsymbol{\Sigma}_{yy}^{i}$ is the covariance matrix of the estimates from an estimation without observation group i.

and the **test statistic**

$$T_i = \sqrt{\mathbf{e}^{(i)T} {\boldsymbol{\Sigma}_{ee}^{(i)}}^{-1} \mathbf{e}^{(i)}} \qquad (18)$$

with $\mathbf{e}^{(i)}$ being the residual vector of estimation group i and $\boldsymbol{\Sigma}_{ee}^{(i)}$ the corresponding covariance matrix. A gross error in the observation group i can be detected with a significance level α by checking

$$T_i > \chi^2(\alpha, n)$$

with

$n =$ Number of fiducial marks in observation group $i \cdot 2$
$\alpha =$ significance level e.g 99.9%

The goal of the sensitivity analysis is to get **one** overall objective quality measure which evaluates the success of the automatic fiducial mark location.

Equation (13) shows the influence of the observed positions of the fiducials in the image to an arbitrary function of the unknowns, namely the transformation parameters from the plate system to the pixel system. Thus the maximum influence (worst case) of an observation group to a transformed point using the estimated transformation from image system to the pixel system in pixel is

$$\nabla_{max} = max\left(\bar{\delta}_i\right) \cdot max\left(\sigma_{\hat{x}}\right) \qquad (19)$$

where $\sigma_{\hat{x}}$ is the variance of the adjusted observations.

5.2 AIO: A Traffic Light Program

There are two critical matching tasks within the AIO: these tasks are critical in the sense that they may fail for some reason. The first is the detection of the orientation of the image, and the second is the localization of the fiducial marks.

The result of both matching tasks is evaluated and classified into one of the three categories of the *traffic light*.

For the evaluation of the orientation detection we use a hypothesis test to check whether the maximum correlation coefficient ρ_{max1} is significantly different from the second maximum ρ_{max2}. The normalized test statistic is

$$T = \frac{\rho_{max1} - \rho_{max2}}{\sqrt{\frac{\sigma^2_{\rho max1}}{n} + \frac{\sigma^2_{\rho_{max2}}}{n}}} \sim N(0,1) \qquad (20)$$

Using this test statistic, the controlling of the traffic light can be achieved by applying the following different significance levels α:

$$\begin{array}{rcl} green & : & T \leq N(\alpha = 0.05\%) \\ yellow & : & N(\alpha = 0.05\%) < T < N(\alpha = 0.1\%) \\ red & : & T \geq N(\alpha = 0.1\%) \end{array} \qquad (21)$$

For the final evaluation of the the localization of the fiducial marks we use the sensitivity analysis. The maximum influence of an observation group on a transformed point using the estimated transformation from the image system to the pixel system (equation 19) is used to control the traffic light:

$$\begin{array}{rcl} green & : & \nabla_{max} \leq 0.5\, pixel \\ yellow & : & 0.5\, pixel < \nabla_{max} < 1\, pixel \\ red & : & \nabla_{max} \geq 1\, pixel \end{array} \qquad (22)$$

The test statistic shown in equation 18 is used to inform the user of possible incorrect fiducial mark localizations.

Example:

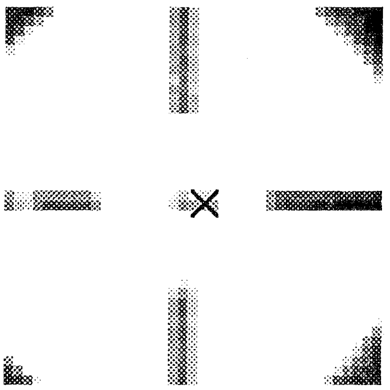

Figure 8: A failed localization of a fiducial mark

Figure 8 shows a negative image of a fiducial mark. The black cross marks the incorrect localization of the fiducial mark.

i	T_i	μ_i	$\bar{\delta}_i$	$\bar{\delta}_{i0}$
1	1.11	0.70	2.91	2.81
2	0.16	0.65	0.38	2.59
3	1.36	0.70	3.55	2.81
4	1.01	0.65	2.44	2.59
5	2.20	0.70	5.74	2.81
6	3.73	0.65	8.99	2.59
7	1.71	0.70	4.47	2.81
8	1.20	0.65	2.88	2.59

Table 1: Sensitivity values; number 6 is the fiducial mark shown in figure 8

Table 1 shows the corresponding result of the sensitivity analysis and the normalized test statistics for each single fiducial mark. In this case the maximum influence of an observation group ∇_{max} was 0.9 $[pixel]$, which indicated an error. Fiducial number 6, having the largest test statistics, is the one shown in figure 8.

6 AIO IN A PRODUCTION ENVIRONMENT

The development was a joint cooperation between the Institute of Photogrammetry, Bonn University and Carl Zeiss, Oberkochen, Germany. The presented AIO is part of the Digital Photogrammetric Workstation (DPW) PHODIS (Photogrammetric Digital Image Processing System).
The product family PHODIS from Carl Zeiss company uses several automatic processes to support the daily work with a (DPW). Besides the AIO there are several other automatic

or autonomous processes like ARO (Automatic Relative Orientation) and PHODIS AT (Automatic Aerotriangulation), which are based on the results from the AIO. Therefore it is very important that the process works robust and reliable.

The AIO has been in use for quite a while and been tested on different cameras (RMK TOP, RMK A, LMK and RC) respectively images with different resolutions and quality. The results confirm the high reliabilty of the approach and also the internal decesive measure for selfdiagnosis.

The succcess rate with respect to the selfdiagnosis is summarized in the following table:

Sensitivity analysis Reality	correct	incorrect
correct	49	(I) 0
incorrect	(II) 0	3

Table 2

94% of the interior orientations have been correctly found to be good (green cases). In 6% the system has correctly detected a failed orientation (red or yellow cases). Erroneously detected as bad (type I error) and the most expensive type II error, a failed orientation classified as beeing correct, has not occured.

The computational time for an automatic interior orientation including the orientation determination on an aerial image with 8 fiducial marks is about 20 sec on a SGI Indigo2. The accuracy of a single fiducial mark measurement is about one tenth of a pixel. The sigma nought of the transformation estimation, which is a quality measure of the estimation, is obout 0.2 pixel. This is $3 \mu m$ at pixel resolution of 15 μm, and therefore comparable with a high precission manually measured interior orientation.

Finaly we could say, the AIO is an autonomous process, ready for use in production.

REFERENCES

[BAARDA W. 67] Statistical Concepts in Geodesy, Netherlands Geod. Commision new series, Vol. 2, no. 4 Delft, 1967

[BAARDA W. 68] A testing procedure fore use in geodetic networks, Netherlands Geod. Commision new series, Vol. 2, no. 5 Delft, 1968

[BOLLES R. C., FISHLER M. A. 81] Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Comm. of ACM, Vol. 24, no. 6, 1981 pp 381 - 395

[BRÜGELMANN R., FÖRSTNER W. 92] Noise Estimation for Color Edge Extraction, in FÖRSTNER/RUWIEDEL 1992, pp. 90 - 106

[HARALICK R. M., SHAPIRO L. G. 93] Computer and Robot Vision, Vol II, Addison-Wesley, USA 1993, p 316 ff

[FÖRSTNER W. 83] Reliability and Discernability of Extended Gauss-Markov Models, DGK A98, Munich, 1983, pp 79 - 103

[FÖRSTNER W. 92] Quality of Vision Algorithms, tutorial notes, Bonn, Institute of Photogrammetry, 1992, pp 30 - 40

[FÖRSTNER W., RUWIEDEL ST. (Ed. 1992)] Robust Computer Vision – Quality of Vision Algorithms, Wichmann, Karlsruhe, 1992