ROBUST PROCEDURES FOR DATA PREPROCESSING, TESTING AND ARCHIVING

Tamara Bellone* Bruno Crippa** Luigi Mussio**

* DIGET - Politecnico di Torino (Italy)
Corso Duca degli Abruzzi, 24 – 10129 Torino, Italy
** DIIAR - Politecnico di Milano (Italy)
Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy
bruno@ipmtf4.topo.polimi.it
Commission I, Working Group 6

Keywords: Accuracy, reliability, robustness.

Abstract

Robust estimation techniques are essentially downweighting methods and, among them, redescending estimators are the most promising ones, because their breakdown point is often very high. A method, recently proposed by Rousseeuw and Leroy, is here presented and applications to outlier identification in photogrammetry are discussed.

1. The problem

Outlier identification and solution methods insensitive to outliers are a main topic in the photogrammetric and generally survey and mapping community and many significant results have been established. The fundamental concepts of internal and external reliability introduced by Baarda (Baarda '67 et '68) received a widespread acknowledgment and provide guidelines in network design as well as in outlier identification. Many testing strategies have been suggested to improve the efficiency of data snooping and reduce masking effects: some are based on still unidimensional test statistics and look for a satisfactory backward and/or forward elimination procedure. In the last decade also robust estimation procedures became part of the mathematical background of photogrammetric and generally survey and mapping community; further achievements are coming out in robust testing. This might lead in the future to a decline of the fortune of the least squares principle; at present, nevertheless, robust estimation methods heavy rely on least squares since, as outlined above, their computational scheme is based on iterative least squares adjustments.

"Robustness is insensitivity against small deviations from assumptions" (Huber '81): it is looked for an estimator being perhaps less efficient when all model hypothesis are satisfied, but which is still capable, to the contrary, of identifying the kernel of consistent observation. Among model assumption violations, the more understood is perhaps the shape of the true underlying distribution deviating slightly from the assumed (usually the gaussian distribution). According to (Hampel et al. '86), "robust statistics are the statistics of the approximate parametric models"; this means robust estimators are derived under a distributional model more flexible than the maximum likelihood estimators: more precisely they provide an infinite dimensional neighbourhood of a specified parametric model. Contaminations of the basic distribution are explicitly accounted for. The estimation procedure is designed to provide a screening among the observations, taking a priori into account that not all of them should be given the same role in determining the solution. This does not happen to least squares estimates, where all observations equally contribute, on the basis of their a priori variance, to the solution.

Most robust estimation techniques are basically downweighting methods where in an iterative least squares scheme suspicious observations undergo to a decrease of their role in determining the solution, through the modification of their weights according to some specified criterion. The amount of the weight change is generally determined on the basis of the (standardized) residual of the observation and may involve from a theoretical point of view all observations. Following a more pratical approach (Bucciarelli et al. '92), changes to the weights will be assumed to be significant, only for the observations directly affected by outliers and for a small other group around (roughly speaking, all the observations closely connected by the functional model to erroneous ones). This means that, apart from pathological situations, only a small percentage of the weights will change from two successive iterations. In this frame, sequential updating becomes again an attractive proposal for outlier removal. A weight change will be obtained by removing from the equation system the same (normal) equation, with weight equal to the given weight change.

Sequentially building an equation system is a widespread technique in many areas of scientific computing: this is for instance the case in all dynamic measurement processes, where on-line data acquisition is often requires to control in real (or near real time) the process evolution. In regression analysis (Draper, Smith '61), when testing for the significance of the parameters involved in determining the observed quantity, the obvious way of modifying the functional model is by using sequential algorithms.

In photogrammetry this approach became interesting with the advent of on-line triangulation, where the possibility of direct data acquisition on the computer and the opportunity of having a quick check and repair of measurement and identification errors strongly suggested the use of such tool. Many algorithms have been presented and investigated to this aim in the last decade, among which Givens transformations (Golub, van Loan '86) are perhaps the most popular, in order to meet the specific requirements of on-line triangulation.

Robust procedures and sequential strategies are very useful when data collection was sequential (even in a kinematics way) too; moreover, the same procedures and strategies may be profitable for adjustment (that means, for densification of an already existent network or in optimization methods) and for interpolation and approximation (that may be, for progressive or selective sampling).

Digital photogrammetry and image processing offer more interesting occasions to this approach; in fact many steps (e.g. image quality control and assessment, features extraction and parsing, image/map/object matching, surface reconstruction, form descriptors) foresee robust procedures and sequential strategies as important tools, useful in the whole process from data acquisition to data representation, taking into account data processing (including pre-processing and post-processing), testing and archiving too. There are many reasons which confirm the actual trend; the power of electronics and computer sciences improves the use of soft images (remotely sensed or acquired by digital scanners, as well as obtained by hardcopy scanning), emphasizing the mathematical treatments instead of some analog methodologies.

Finally a more refined and conservative procedure has been recently presented by statisticians (Rousseeuw, Leroy '87); it goes over the capacity of classical robust estimators, because it has a very high breakdown point. This means that outliers of bigger size and in a large number may be considered: therefore because, as already said, blunders, leverages and small outliers occur often in the observations and they must be identified and eliminated, in order to get the expected results, the applications of robust estimators with a very high breakdownpoint to photogrammetry and, in generally, survey and mapping in its many fields are welcome.

2. The method

The most promising robust estimators are, among the downweighting methods, the redescending estimators, specially when their breakdown point is very high. In fact outliers of bigger size and/or in a large number may be considered; moreover different explanation can be set up, when the anomalous data, after rejection, show a homogeneous behaviour.

The basic idea follows some suggestions of Hampel for introducing a rejection point in the loss function, so that the data outside the interval get, automatically, weights equal to zero. On the contrary, the data inside get weights equal to one, if they belong to the inner core of data, or ranging from one to zero, if they stay in intermediate region of doubt.

There are many ways to concretize Hampel suggestions. The easiest is represented by the Generalized M-estimators, where some suitable weight functions correct the behaviour of the M-estimators, as defined by Huber. Unfortunately this strategy (called by some authoritative authors: minimax), although it increases the breakdown point, is unable to raise it substantially.

The best modality is represented by the least median of squares, where the median of the squares of the residuals is minimized in order to obtain the expected results:

$$\phi = min \underset{k=1,m}{med} (v_k^2)$$

being m the number of observations. Unfortunately this strategy is, at present time, computationally too expensive, because no efficient algorithm is known to solve $\binom{m}{n}$ systems, selecting n observation among the m ones, forming the sample, when n is the number of unknown parameters.

An advantageous alternative is represented by the least trimmed squares, where the average of the squares of the residuals, belonging to the inner core of data, is minimized in order to obtain the expected results:

$$\phi = min \sum_{k=1}^{h} v_k^2.$$

In other words, only a part of the observations are processed in each step of linear adjustment.

The use of a sequential updating of a preliminary computed solution is a possible alternative to repeating many times the whole adjustment (Lawson, Hanson '74). The formulae for updating, in terms of parameters as well as in term of observations, of the Cholesky factor and the inverse matrix are shown in Tables 1 and 2.

Sequential updating of the Cholesky factor[1]

observation:

$$\begin{aligned}
t'_{ii} &= \sqrt{t_{ii}^2 \pm (w_i^{(i)})^2} \\
w_i^{(i+1)} &= 0 \\
t'_{ij} &= (t_{ii}t_{ij} \pm w_i^{(i)} w_j^{(i)})/t'_{ii} && (j > i) \\
w_j^{(i+1)} &= (w_j^{(i)} t_{ii} - w_i^{(i)} t_{ij})/t'_{ii} && (j > i) \\
w_j^{(1)} &= a_j && \forall j
\end{aligned}$$

parameter:

$$t'_{ij} = t_{ij} \qquad (i < h,\ i \le j < h,\ j > h)$$

"in" only

$$\begin{aligned}
t_{ih} &= (c_{ih} - \sum_{k=1}^{i-1} t_{kh}t_{ki})/t_{ii} = t'_{ih} && (i < h) \\
t_{hh} &= \sqrt{c_{hh} - \sum_{k=1}^{h-1} t_{kh}^2} = t'_{hh} \\
t_{hj} &= (c_{hj} - \sum_{k=1}^{h-1} t_{kh}t_{kj})/t_{hh} = t'_{hj} && (j > h)
\end{aligned}$$

$$\begin{aligned}
t'_{ii} &= \sqrt{t_{ii}^2 \pm (w_i^{(i-h)})^2} && (i > h) \\
w_i^{(i-h+1)} &= 0 \\
t'_{ij} &= (t_{ii}t_{ij} \pm w_i^{(i-h)} w_j^{(i-h)})/t'_{ii} && (i > h,\ j > i) \\
w_j^{(i-h+1)} &= (w_j^{(i-h)} t_{ii} - w_i^{(i-h)} t_{ij})/t'_{ii} && (j > i) \\
w_j^{(1)} &= t_{hj} && (j > h)
\end{aligned}$$

Table 1

Sequential updating of inverse matrix[2]

observation:

$$\begin{aligned}
(C \pm a^t pa)^{-1} &= C^{-1} + \\
&\mp C^{-1}a^t(p^{-1} \pm aC^{-1}a^t)^{-1}aC^{-1}
\end{aligned}$$

parameter:

"in"

$$\begin{aligned}
\gamma &= C^{-1} + C^{-1}rs^{-1}r^tC^{-1} + \\
&+ C^{-1}r(s + r^tC^{-1}r)^{-1}r^tC^{-1} \\
\rho &= -s^{-1}\gamma r \\
\sigma &= s^{-1}(1 - r^t\rho)
\end{aligned}$$

"out"

$$\begin{aligned}
C^{-1} &= \gamma - C^{-1}rs^{-1}r^tC^{-1} + \\
&+ C^{-1}r(s + r^tC^{-1}r)^{-1}r^tC^{-1}
\end{aligned}$$

Table 2

The weighted least trimmed squares could be minimized, avoiding a rough partition between inliers and outliers, where the weighted average of the squares of the residuals takes into account the inner core of the data with weights equal to one, an intermediate doubt region with weights ranging from one to zero, whilst the data in the tails get weights equal to zero:

$$\phi = \sum_{k=1}^{h} w_k v_k^2.$$

Least median of squares and least trimmed squares (or weighted least trimmed squares) have the same breakdown point near to 0.5, when the number h is around $m/2$, i.e. only the best half part of the observations are processed in each step of linear adjustment.

[1]In the following formulae, the symbol a_j indicates a generic element of the row of the design matrix A, to be added to or dropped by the system, the symbols c_{ii}, c_{ij} indicate, respectively, a generic main–diagonal element and a generic off–diagonal element of the normal matrix C, the symbols t_{ii}, t_{ij} indicate the same elements of the Cholesky's factor T.

[2]In the following formulae, the symbol a indicates a generic row of the design matrix A, to be added to or dropped by the system, the symbol p indicates the weight of the corresponding observation; furthermore the normal matrix is split in four parts, being their sub–blocks C, r, r^t and s, and their inverse matrix is split again in four parts, being their sub–blocks γ, ρ, ρ^t and σ.

The number h could be increased until $2m/3$, preserving the reliability of the observations, globally and locally, according to geodesists community suggestions (Benciolini et al. '82), if the amount of suspected outliers isn't too large. The breakdown point decreases, obviously, but not too much, so that the procedure continues to be effective.

These methods are grouped together and generalized by means of the definition of the S–estimators.

3. Examples

The presented methods are already tested and discussed in the scientific literature by statisticians. Unfortunately whilst downweighting methods have been broadly studied by photogrammetrists too, since the last fifteen years (Kubik '80, Förstner '86), the redescending estimators seems to be not popular, but for the simple Hampel estimator.

On the other hand, redescending estimators with a very high breakdown point have been recently introduced in the survey and mapping disciplines (see Carosio's research team: Wicki '92 a et b). Therefore some examples of photogrammetry and cartography are welcome, with the aim to spread out information.

The most interesting examples in photogrammetry and cartography involve S–transformation, fitting and matching. The first two classes of examples are common between photogrammetry and cartography, whilst the last class of examples is central for photogrammetry and it constitutes the experimental conclusion of this paper.

Image matching can be done in image space, as well as in object space, by using complanarity condition or collinearity equations, respectively, adding a grey level model and, eventually, an object model. As well known, the complanarity conditions is one of the most critical examples, concerning well–conditioning, reliability and robustness.

For these reasons, the relative orientation of a couple of images is adjusted, by using redescending estimators with a high breakdown point, where the amount of outliers ranges until $m/3$, being m the number of observations. The data collect three series of observations, according to Ackermann suggestions (Ackermann '79), in the canonical points, with different combinations of outliers.

The outlier location shows 2, 4 and 6 outliers in a series of observations in the canonical points (see Figure 1), preserving the global and local reliability. Least squares and downweighting methods fail the adjustment, because their breakdown point is zero or too low. On the contrary, redescending estimators with a very high breakdown point catch all outliers, in all combinations of them.

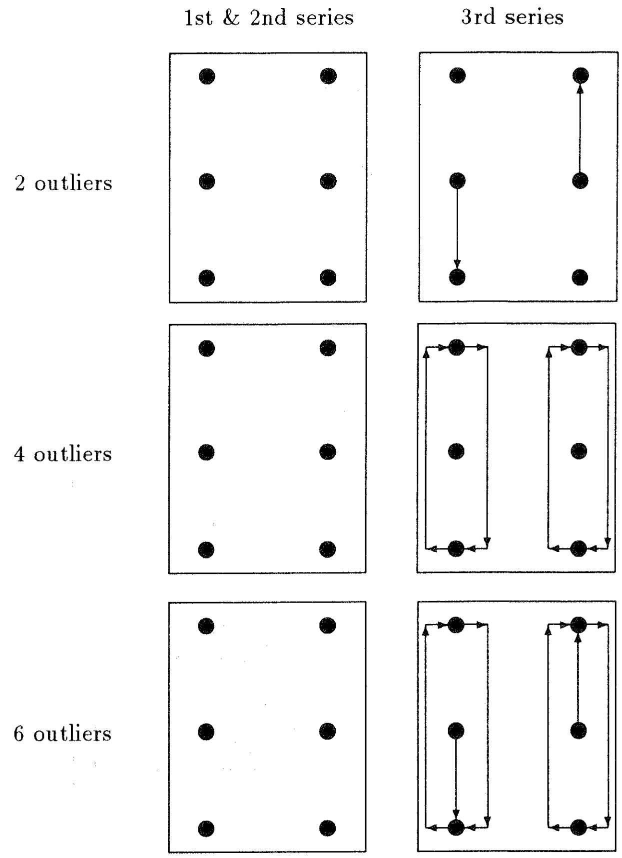

Fig. 1

The strategy of application of the presented procedure (the same of Barbarella, Mussio '85) is an adjustment of the best observations, after a preliminar least squares adjustment. Successively the suspected outliers, which don't show blunders, leverages or small outliers, are forward accepted by using the Hawkins test:

$$H_0 : P(H_\nu^{(-)}(\alpha/2) \leq \hat{H}_e \leq H_\nu^{(+)}(1+\alpha/2)) = 1-\alpha$$

being $\nu = l - n$ the degrees of freedom, where $l \leq m$ the number of observations actually processed at the present step of adjustment (remember, m is the number of observations), and n the number of unknowns parameters.

The expected value $\hat{H}_e$ is computed, as follows:

$$\hat{H}_e = \max_{i=1,l}(\hat{r}_i^2/(\nu\hat{\sigma}_0^2)) = \max_{i=1,l}(\hat{v}_i^2/(\nu\hat{\sigma}_{v_i}^2))$$

being $\hat{v}_i$ the residuals, r_i the recursive residuals, $\hat{\sigma}_{v_i}^2$ the variances of the residuals, $\hat{\sigma}_0^2$ the squares sigma zero and ν the degrees of freedom.

The critical values, for a parametric test on two sides, are derived from the Hawkins probability distribution (Hawkins '80), defined as follows:

$$H_\nu = \max_{i=1,l}((\chi_1)_i)/\chi_\nu^2$$

being $\nu = l - n$ the degrees of freedom.

Figure 2 shows the flow chart of the above explained strategy.

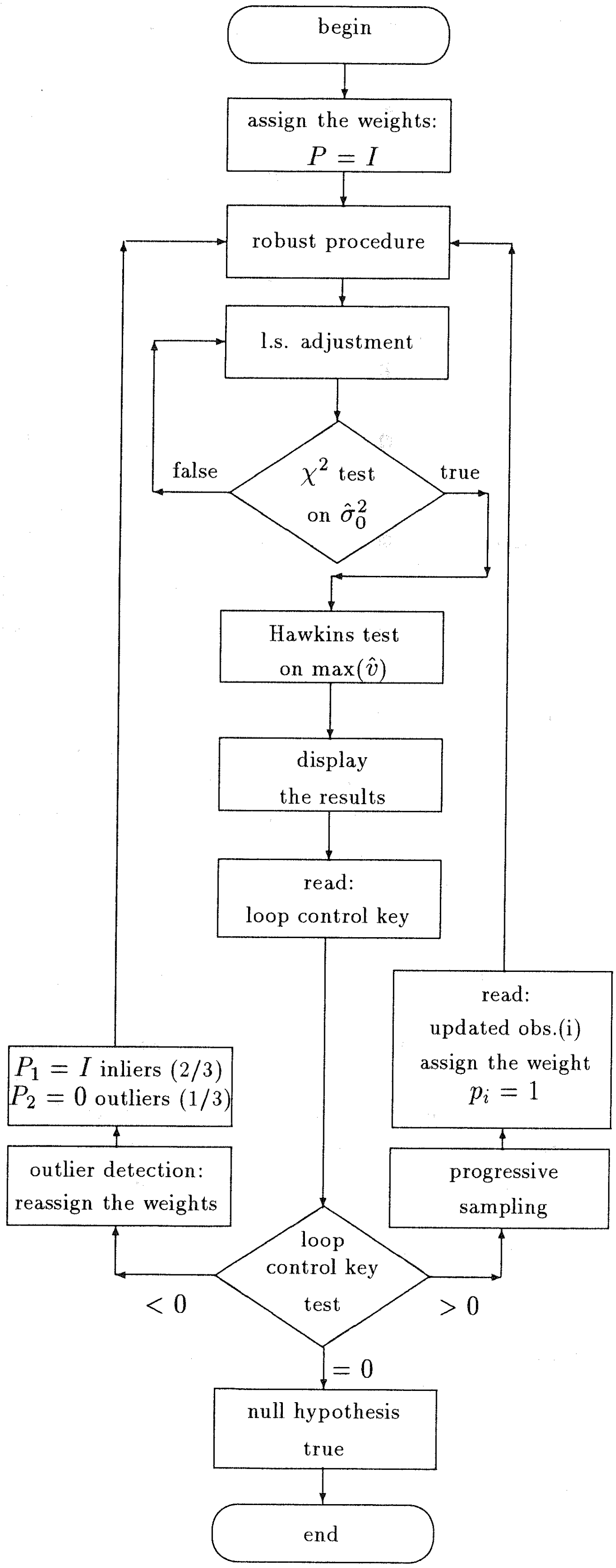

Fig. 2

4. Fields of application

Least squares methods and some other related procedures (e.g. cluster analysis, multiple regression, variance analysis, robust estimators) are usually appropriate for two types of problems:

- network adjustment;
- interpolation and approximation.

In the first one, the observables are functions of point position differences, whilst in the second are functions of point positions. These point positions (or the point position differences) could depend on time.

Moreover the observables, depending from point positions and time, are influenced by physical fields, according to the data collection procedure.

Morphological factors and/or eventual kinematics parameters are functions of the point positions, since they are supposed to have a similar behaviour in the neighbouring points.

In the case of network adjustment, the geometrical model is quite familiar. On the contrary, in the second one two main sub–cases may occur:

- a deterministic law for the behaviour of the phenomenon under study has been previously checked, by a variety of causes, that may be physical, geometrical, or others;
- no deterministic law is previously known for the phenomenon behaviour.

The theory of models has a proper classification for both sub–cases as a "grey box" model and a "black box" model, respectively.

In the "grey box" model, the aim is the estimation of model coefficients, followed by proper significance tests for estimated parameters.

In the "black box" model, the main deterministic and stochastic approaches are preferred:

- in the first case, aside from further details, one has a number of steps, as in the choice of an interpolation strategy (finite elements, Fourier analysis, wavelet interpolation, ...), the estimation of coefficients for the chosen models, the variance analysis (in order to estimate altogether significance of parameters and quality of model);
- the second one employs covariance estimation, covariance function modelling and collocation (linear filtering and prediction).

Least squares methods and some other related procedures may give solution to an important group of problems of photogrammetry and, generally, of survey and mapping, as for example:

- network/block/joint adjustment;
- surface reconstruction, form descriptors;
- feature extraction and parsing;
- image/map/object matching.

Going further with the above said division of adjustment and interpolation/approximation problems of photogrammetry and, generally, of survey and mapping, first area collects:

- on-line triangulation of images: spaceborne, airborne and terrestrial;
- GPS data processing, automatic surveying (robotics).

Prior to processing, it should be pointed out pre-processing of data collected by space photogrammetry techniques (SPOT, MOMS, SAR, ...), with due care, as well as in geodesy (e.g. GPS) and related sciences.

As far as interpolation and approximation are concerned, one should remind a class of problems of photogrammetry and cartography, related to:

- measurement devices (camera calibration and other systems and sensors) and secondary effects;
- morphological features extraction, image/map/object matching;
- DEM generation, orthoimage production and superimposition;
- image processing, analysis (classification) and understanding (semantic interpretation).

Note that these problems usually involve a large number of observations and parameters and, consequently, require the solution of large systems, i.e. systems with a large number of equations and unknowns. For these reasons, because blunders, leverages and small outliers occur often in large sets of data, robust procedures (robust estimators, reliability analysis, ...) suitably provide for data preprocessing, testing and archiving.

References

Ackermann F. (1979): *The concept of reliability in aerial triangulation.* Ricerche di Geodesia Topografia e Fotogrammetria, n. 1, CLUP, Milano.

Baarda W. (1967): *Statistical concepts in geodesy.* Netherlands Geodetic Commission, new series, vol. 2, n. 4, Delft.

Baarda W. (1968): *A testing procedure for use in geodetic networks.* Netherlands Geodetic Commission, new series, vol. 2, n. 5, Delft.

Barbarella M., Mussio L. (1985): *A strategy for a robust identification of the outliers in the geodetic sciences.* Statistics and Decisions, supplement issue n. 2, pp. 397–405, R. Oldenbourg Verlag, Munich.

Benciolini B., Mussio L., Sansò F. (1982): *An approach to gross error detection more conservative than Baarda snooping.* Int. Archives of Photogrammetry, vol. XXIV, part 3, pp. 41–59, Helsinki.

Bucciarelli B., Forlani G., Mussio L. (1992): *Robust estimation by using linear sequential algebra.* Int. Archives of Photogrammetry and Remote Sensing, vol. XXIX, part B2, pp. 328–333, Washington D.C.

Draper N.R., Smith H. (1961): *Applied regression analysis.* John Wiley & Sons, New York.

Golub G.H., van Loan C.F. (1986): *Matrix computations.* North Oxford Academic, London.

Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (1986): *Robust statistics.* John Wiley & Sons, New York.

Huber P.J. (1981): *Robust statistics.* John Wiley & Sons, New York.

Hawkins D.M. (1980): *Identification of outliers.* Chapman and Hall, London.

Krarup T., Juhl J., Kubik K. (1980): *Götterdämmerung over least squares adjustment.* Int. Archives of Photogrammetry, vol. XXIII, part B3, pp. 369–378, Hamburg.

Förstner W. (1986): *Final report on the joint test on gross error detection of OEEPE and ISP WG III/1.* OEEPE, official publication n. 18, Frankfurt am Main.

Lawson C.L., Hanson R.J. (1974): *Solving least squares Problems.* Prentice Hall, Englewood Cliffs (New Jersey).

Rousseeuw P.J., Leroy A.M. (1987): *Robust regression and outlier detection.* John Wiley & Sons, New York.

Wicki F. (1992): *Robuste ausgleichung geodätischer netze.* ETH-ZH, Inst. für Geodäsie und Photogrammetrie, Bericht 189, Zurich.

Wicki F. (1992): *Robuste M - schätzer und zuverlässigkeit.* ETH-ZH, Inst. für Geodäsie und Photogrammetrie, Bericht 190, Zurich.