1 The Role of Robustness in Computer Vision

Wolfgang Förstner, Bonn University

Abstract — The paper discusses tools of diagnostics and robustness in the context of automating vision. Motivation is the building of so-called *traffic light programs* which contain a reliable selfdiagnosis enabling to chain vision modules. Special attention is payed to show the prerequisites for using tools for quality evaluation. The paper concludes with open questions.

1.1 Role of Vision Modules

The scope of this paper is to discuss means for increasing the acceptability of vision modules in applications. The bottle neck appears to be the lack of robustness of existing vision algorithms, using the term robustness in a broad sense, including the resistance to changes in the environment, the stability of the results over a wide range of imagery and the ease of adapting the algorithm to a new situation.

Vision modules always are part of larger systems which not necessarily aim at solving a vision task, such as inspection, navigation or positioning. On the other hand, vision modules themselves consist of vision submodules, which interact in a complex manner.

For designing such systems clear performance characteristics of all submodules are necessary. Following [15] the devolpment of an algorithms can be considered finished if:

1. the algorithm works in XX % of all cases,
2. the algorithm reports failure in the other 100 - XX % cases, and
3. one is satisfied with XX %.

This motivates the development of so-called *traffic light programs* which have three outcomes:

- **green**: The result is: ... . Its quality is: ... .
- **yellow**: The result may be: ... . I am not sure about: ... (Please, check)
- **red**: I did not succed. The reasons might be: ... . Partial results are: ... .

In order for a vision module to perform a selfdiagnosis of this kind several prerequisites have to be fullfilled, which will be discussed in the following. The algorithms need to be based on a clearly described model. This model also

needs to include also stochastic components to be able to handle the random fluctuations in the original measurements, e. g. the intensities (cf. section 1.2). For achieving a high enough robustness, a certain amount of modelling also needs to refer to expected larger deviations from the ideal model, resulting from changing lighting, occlusions, missing features, under- or oversegmentations, matching errors etc. Based on such models both, diagnostic tools and robust estimators (cf. section 1.3) must be provided.

1.2 Error Modelling

1.2.1 Modelling the observation process

Models are surrogates for reality representing that part which is necessary to achieve a certain goal. They are never complete thus never really correct but always approximations, described in some symbolic thus mathematical manner.

It is useful to split a mathematical model into a functional and a stoachstical model:

1. The functional model describes the generally nonlinear relations between the true or expected values. Many physical laws are of that type.
2. The stochastical model describes the expected random fluctuations from the functional model which can be expected under good conditions. The stochastical model could also be termed the noise model.

A typical example for a mathematical model is based on the first and second moments, thus the expectation and the dispersion of the n observed values $\boldsymbol{l}$:

$$E(\boldsymbol{l}) = \boldsymbol{f}(\boldsymbol{p}) \tag{1.1}$$

$$D(\boldsymbol{l}) = \boldsymbol{\Sigma}_{ll} \tag{1.2}$$

The observation process is made explicit by the function $\boldsymbol{f}$, depending nonlinearily on u unknown continuous parameters $\boldsymbol{p}$. If no other information is provided the principle of maximum entropy implies that a Gaussian distribution is assumed. The redundancy $r = n - u$ of the system must be ≥ 0 in order to allow for a solution, $r > 0$ is necessary for being able to check the result with respect to deviations from the assumtions.

Examples for such a model can be found in many low and mid-level vision processes, such as image restauration, feature extraction, curve fitting, coordinate transformations, image matching, optical flow, object location, relative orientation, surface reconstruction etc.

Remark: The model may be extended towards discrete parameters, which are necessary in case of classification tasks. The function $\boldsymbol{f}$ then is not continuous anymore and the uncertainty of the observations has to be additionally described by the probability of belonging to a certain class. Though this extension is necessary, e. g. for segmentation, grouping or object identification, the

theory for handling these type of problems has not been worked out as far as the above mentioned mathematical model is concerned. For this reason we do not discuss classification in more detail.

The strength of the above mentioned setup of the mathematical model is its ability to propagate uncertainty from the original observations to the final result, even in case of cascading several such analysis steps. E. g. one could proceed with the following sequence for deriving surface heights z from intensities g, including the determination of the orientation $\boldsymbol{o}$ of the images:

image intensities	g	σ_g
edge elements	$p = f_p(g)$	$\sigma_p = h_p(\sigma_g)$
edges	$e = f_e(p)$	$\sigma_e = h_e(\sigma_p)$
corners	$c = f_c(e)$	$\sigma_c = h_c(\sigma_e)$
orientation	$\boldsymbol{o} = \boldsymbol{f}_o(c, e)$	$\boldsymbol{\Sigma}_{oo} = \boldsymbol{H}_o(\sigma_c, \sigma_e)$
matching parallaxes	$p_x = f_{p_x}(\boldsymbol{o}, c, e)$	$\sigma_{p_x} = h_{p_x}(\boldsymbol{\Sigma}_{oo}, \sigma_c, \sigma_e)$
interpolation/z-values	$z = f_z(p_x, \boldsymbol{o})$	$\sigma_z = h_z(\sigma_{p_x}, \boldsymbol{\Sigma}_{oo})$

All steps obviously contain a geometrical and/or physical model represented in the functions f. The uncertainty is covered by the standard deviations σ or the covariance matrices $\boldsymbol{\Sigma}$. Due to the central limit theorem this appears to be an adequate approximation for the ideal observation process.

1.2.2 Modelling Deviations from the Observation Process

The ideal model will not always hold. This is the reason why it has become common practice to distinguish three types of error sources: random errors (cf. above), systematic errors and gross errors.

- *Random* errors actually cover several error sources, such as photon noise, electronic noise, discretisation, quantization or random athmospheric perturbations due to turbulences etc. These errors in most cases are covered in the stoachstical model as discussed before.

- *Systematic* errors cause a bias in the model. Examples are lens distortion, refraction, mis-calibration, neglected dependencies between the observations, wrong weighting of the observations, wrong smoothness constraints etc. Most of these effects may be taken care in the model, however, may not be recoverable under production conditions, thus require proper calibration. In many cases systemtic effects can be incorporated into the functional model. Then a quite complete theory is at hand for analysing their effect on estimation results, which is the basis for many diagnostic tools (cf. below).

- *Gross* errors in contrast to systematic errors, may be of any size and influence only single observations. Typical examples are classification errors, matching errors, labeling errors, shot noise, spurious data etc. Only in

simple cases can gross errors be represented as deviations from the functional model, e. g. assuming the mean of an observation to be different by a certain amount. This allows to use the same diagnostic tools as for systematic errors. If gross errors appear in a large percentage their properties need to be captured by the stochastical model, assuming long tailed distributions, e. g. the double exponential distribution $\exp(-|x/a|)/a$, leading to the L_1-norm as optimal estimate, the median being a special case. The theory for handling these cases is by far not worked out for being practical.

The next section will discuss how these error models are used for achieving robust and reliable vision algorithms.

1.3 Diagnostics and Robustness

Following the guidline by *Pregibon* [15], reaching a certain percentage of succes requires the algorithm to be *robust*, whereas being able to tell failure requires the algorithm to contain *diagnostic tools*. Accordingly the two terms are defined as follows (cf. [11]):

- *Robustness*: The purpose of robustness is to have safeguards against deviations from the assumtions.
- *Diagnostics*: The purpose of diagnostics is to find and identify deviations form the assumtions.

We want to discuss both notions in an informal manner (for details cf. [5], [7]). As diagnostic tools can also be used for planning purposes, guaranteeing a mensuration design which allows to achieve robust results, we will discuss diagnostic tools first.

1.3.1 Diagnostics

Diagnostic tools work on different levels, depending on the redundancy in the system. Moreover, diagnostics may refer to the design of the mensuration system without referring to data or without taking the actual measurements into account. Finally, we need to distinguish between analysing the data and analysing the estimated parameters.

The different notions are collected in the table (cf. also [5]) together with the equations related to single outliers Δl_i and estiamted parameters $\widehat{p}_j$. We want to discuss three representative measures.

Determinability.

The determinability of parameters characterizes their sensitivity with respect to random errors in the observations. It can be represented by the covariance

matrix of the estimated parameters $\widehat{\boldsymbol{p}}$:

$$\boldsymbol{\Sigma}_{\widehat{p}\widehat{p}} = (\sigma_{\widehat{p}_j\widehat{p}_k}) = \left[\left(\frac{\partial \boldsymbol{f}}{\partial p_j} \right)^T \boldsymbol{\Sigma}_{ll}^{-1} \left(\frac{\partial \boldsymbol{f}}{\partial p_k} \right) \right]^{-1} \tag{1.3}$$

depending on the quality of the given observations and the partial derivatives of the functional model. The matrix in [] is the Fisher-Information matrix. The covariance matrix (1.3) is the Cramer-Rao bound on the precision, i. e. a lower bound for the variance achievable with any type of estimator. Also the effect of the actual observations can be taken into account for determining the precision, e. g. $\widehat{\sigma}_{\widehat{p}_j}$, by using the estimated variance factor $\widehat{\sigma}_0^2 = \Omega/r = \widehat{\boldsymbol{\varepsilon}}^T \boldsymbol{\Sigma}_{ll}^{-1} \widehat{\boldsymbol{\varepsilon}}/r$ with the residuals $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{f}(\widehat{\boldsymbol{p}}) - \boldsymbol{l}$ (cf. the following table).

It is the task of *modelling* to actually reach this lower bound in order to fully exploit the information contained in the observations, and the task of *empirical testing* to show that this matrix actually represents the achieved precision in order to have a reliable diagnostic tool at hand.

Evaluation of $\boldsymbol{\Sigma}_{\widehat{p}\widehat{p}}$ can be achieved by comparing it to a specified criterion matrix $\boldsymbol{H}$, requiring the confidence region with $\boldsymbol{\Sigma}_{\widehat{p}\widehat{p}}$ lying completely within $\boldsymbol{H}$.

Testability

Whereas the determinability only refers to random perturbations in the observations and can be determined even if no redundant observations are available, checking for systematic or gross errors requires redundant observations.

Testability or checkability can be easily derived when analysing the used checks or test statistics. E. g. when testing the residual $\widehat{\varepsilon}_i = f_i(\widehat{\boldsymbol{p}}) - b_i$ of the observation l_i we obtain the optimal test statistic $z_i = -\widehat{\varepsilon}_i/(\sigma_{l_i}\sqrt{r_i})$ where $r_i = \sigma^2_{\widehat{\varepsilon}_i}/\sigma_{l_i}$ is the contribution of the ith observation to the total redundancy $r = u - n$ of the system, as $\sum r_i = r$. Requiring errors to be detectable with a certain minimum probability β_0 one obtains a lower bound $\Delta_0 l_i$ on the size of a detectable outlier:

$$\Delta_0 l_i = \delta_0(\alpha_0, \beta_0)\sigma_{l_i}\sqrt{\frac{1}{r_i}} \tag{1.4}$$

where the parameter δ_0 depends on the significance number α_0 and the required probability β_0 for detecting the error (e. g. $\alpha_0 = 0.05$, $\beta_0 = 0.8$ leads to $\delta_0 = 4.1$). Thus only if $r_i > 0$ the testing is possible. $r_i = 0$ means l_i not to be checkable, $r_i = 1$ corresponds to full testability. Data points with $r_i << 0$ are called *leverage points*, as they heavily influence the result (cf. eq. (1.5))

redundancy	theoretical diagnosis w. r. t. data	theoretical diagnosis w. r. t. result	empirical diagnosis w. r. t. data	empirical diagnosis w. r. t. result
$r \geq 0$	effect of *random errors*, precision: covariance matrices, variances, standard deviations			
	assumed variance factor σ_0^2		estimated variance factor $\widehat{\sigma}_0^2 = \frac{\Omega}{r}$, $\frac{\widehat{\sigma}_0^2}{\sigma_0^2} \sim \mathrm{F}_{r,\infty}$	
	given precision	theoretical precision	estimated precision	estimated precision
		determinability		
	$\sigma_{l_i}^2 = (\mathbf{\Sigma}_{ll})_{ii}$	$\sigma_{\widehat{p}_j}^2 = (\mathbf{\Sigma}_{\widehat{p}\widehat{p}})_{jj}$	$\widehat{\sigma}_{l_i} = \frac{\widehat{\sigma}_0}{\sigma_0}\sigma_{l_i}$	$\widehat{\sigma}_{\widehat{p}_j} = \frac{\widehat{\sigma}_0}{\sigma_0}\sigma_{\widehat{p}_j}$
		separability $\rho_{\widehat{p}_j\widehat{p}_k} = \frac{\sigma_{\widehat{p}_j\widehat{p}_k}}{\sigma_{\widehat{p}_j}\sigma_{\widehat{p}_k}}$		
$r \geq 1$	effect of *(non-) detectable gross or systematic errors*			
	noncentrality parameter $\delta_0(\alpha_0, \beta_0)$		test statistic $z_i = \frac{-\widehat{\varepsilon}_i}{\sigma_{\widehat{\varepsilon}_i}} \sim \mathcal{N}(0,1)$	
	detectability factor	theoretical sensitivity factor	normalized size of error	empirical sensitivity factor
	$\delta_{0i}' = \delta_0\sqrt{\frac{1}{r_i}}$	$\overline{\delta}_{0i} = \delta_0\sqrt{\frac{u_i}{r_i}}$	$\delta_i' = z_i\sqrt{\frac{1}{r_i}}$	$\overline{\delta}_i = z_i\sqrt{\frac{u_i}{r_i}}$
	detectability **testability**	theoretical sensitivity	estimated size of errors	empirical **sensitivity**
	$\Delta_0 l_i = \delta_{0i}'\sigma_{l_i}$	$\Delta_{0i}\widehat{p}_j \leq \overline{\delta}_{0i}\sigma_{\widehat{p}_j}$	$\widehat{\Delta l}_i = \delta_i'\sigma_{l_i}$	$\widehat{\Delta_i\widehat{p}_j} \leq \overline{\delta}_i\sigma_{\widehat{p}_j}$
$r \geq 2$	effect of *(non-) locatable gross or systematic errors*			

Sensitivity

As gross or systematic errors may stay undetectedin spite of testing, the result may be deteriorated by non-detectable or non-detected errors. This reflects the sensititvity of the result.

The effect $\Delta_{0i}\widehat{p}_j$ of just non-detectable errors $\Delta_0 l_i$ onto the estimated parameters is bounded and determines the theoretical sensitivity.

The *estimated* effect $\widehat{\Delta_i\widehat{p}_j}$ of the actual error in the observations on the the estimated parameter $\widehat{p}_j$ is bounded according to

$$\widehat{\Delta_i\widehat{p}_j} \leq z_i\sqrt{\frac{1-r_i}{r_i}}\sigma_{\widehat{p}_j} \tag{1.5}$$

and obviously depends on the size of the test statistic z_i, on the geometry via the ratio $(1-r_i)/r_i$ and the precision of the result, namely the standard deviation of the parameter in question. The value $\widehat{\Delta_i\widehat{p}_j}$ also measures the change of the estimated parameter $\widehat{p}_j$ when leaving out observation p_i from the estimation

process. Obviously data at leverage points with $r_i << 1$ heavily influence the result.

Discussion

The diagnostic values can be generalized as follows:

- systematic errors, e. g. lens distortions (cf. [1], [2]).
- groups of observational errors, which may show correlations within the group but independence between the groups, e. g. when dealing with image features in stereo which are represented by a set of observational values e. g. straight line segments (cf. [16]).
- subsets of unknown parameters, e. g. restricting to form parameters, neglecting the influence onto pose parameters (cf. [5]),
- errors caused by wrong decisions during testing, e. g. when analysing the effect of erroneously identifying a model error as gross error instead of systematic error. This requires a redundancy $r \geq 2$ and leads to notions as locatability or separability of gross or systematic errors (cf. [4]).

The diagnostic tools may serve the following purposes in Computer Vision:

1. *Planning* of vision systems by choosing the design via $\boldsymbol{f}$ and $\boldsymbol{\Sigma}_{ll}$ to reach acceptable quality, e. g. for inspection purposes.
2. controlling *focus of attention* by identifying weak, unstable, unreliable non-testable, sensitive etc. situations and choosing. additional information to optimize gain in quality, e. g. by identifying none or only weak determinable parameters and choosing appropriate observations.
3. *Checking a design* with respect to achievable quality for feeding a *traffic light program*, e. g. by comparing the reached quality with the specifications.
4. *Comparing the quality of algorithms* with the same input/output structure, e. g. matching techniques.
5. developing *specifications* for vision algorithms in order to be able to perform proper evaluation.

1.3.2 Robust Estimators

The diagnostic tools provide insight into the state of a vision module. They are based on certain estimation techniques, mainly least squares estimators, best linear unbiased estimators or – most generally – maximum likelihood (ML) estimators. These estimators only are optimal provided the given model actually holds. Otherwise they lead to wrong or at least to non-optimal results.

Robust estimators aim at yielding nearly optimal results in the presence of deviations from the model. While deviations from the model may have nearly any characteristics certain types of deviations have shown to be theoretically tractable in terms of performance prediction. Two classes will be discussed here due to their importance in vision algorithms:

1. Maximum likelihood estimators provide an estimation scheme similar to the classical ones.

2. High-Breakpoint Estimators have been developed for cases with severe deviations from the model, namely for data with a large percentage of outliers.

ML-type Estimators

Maximum Likelihood-type estimators minimize

$$\Omega = \sum_i \rho(l_i - f_i(\boldsymbol{p})) \tag{1.6}$$

instead of the classical least squares, where $\rho(x) = \frac{1}{2}x^2$, here assuming all weights to be 1. This is equivalent to maximizing the likelihood function

$$L(\boldsymbol{l}, \boldsymbol{p}) = \prod_i \mathrm{e}^{-\rho(l_i - f_i(\boldsymbol{p}))} \tag{1.7}$$

revealing the estimator to be ML if the probability density of the residuals is exponential. A classical example is the L_1-norm where $\sum_i |l_i - f_i(\boldsymbol{p})|$ is minimized corresponding to the double exponential distribution $\exp|x|$ (cf. above).

ML-type may be characterized by the so-called *influence curve* (cf. [10]) which is

$$IC(x) = \rho'(x) \tag{1.8}$$

indicating the dependency of the result on errors in *single* observations. The diagnostic tools for handling *single* gross errors in a least squares estiamtion rely on the influence curve $IC(x) = x$ being linear. Obviously the L_1-norm with $IC(x) = \text{sign } x$ is a more robust estiamtor as the influence of errors in the observations is bounded. If very large errors should have no influence on the result one needs to choose an optimization function with *redescending* influence curve.

The estimation can be realized by the *method of modified weights*, thus in each iteration the weight of each observation is changed depending on the residual in the previous iteration according to

$$w(x) = \frac{\rho'(x)}{x} \tag{1.9}$$

which is a constant for the least squares estimate.

The following theorem however motivates to strictly distinguish between convex and non-convex minimum function or, equivalently, between non-redescending and redescending influence curve:

Theorem (Huber 1976): If the minimum function $\rho(x)$ is convex and symmetric, the weight function $w(x) = \rho'(x)/x$ is decreasing for $x > 0$ and the model is linear, the the method of iterative reweighting converges to a unique solution.

Therefore, in practical implementations one would start iterating with a convex minimum function, e. g. using $w(x) = 1/\sqrt{1+x^2}$ in order to closely approach the optimum and continue with a non-convex minimum function, e. g. using $w(x) = \mathrm{e}^{-x^2}$, in order to eliminate the influence of large outliers. An alternative are scaling schemes as proposed by [3]. ML-type estimators are closely linked to the principle of minimum description length ([6]) and very common in Computer Vision due to their flexibility and their low algortihmic complexity being $O(u^3)$ in the worst case (cf. e. g. [13], [9]).

A closer analysis of the concept of ML-estimators reveals some weaknesses:

- The doubly exponential distribution is the distribution with the slowest descent in probability (density), thus the distribution with the heaviest tails while at the same time showing convexity in the corresponding minimum function $\rho(x)$ and having finite moments (mean, variance, etc.). Minimum functions with redescending influence function do not actually correspond to distributions with finite moments, explaining the name ML-*type* estimation.
- The concept of the influence curve is a differential one as can be seen from the definition (cf. also eq. (1.8). This is a hint that very large errors cannot be guaranteed to be eliminated by a ML-type estimate. An example is the linear regression with the L_1-norm, which can lead to arbitrary wrong results if a gross error occurs at a leverage point (cf. the discussion after eq. (1.4)).
- The concept of the influence curve only refers to a single observation. This indicates that multiple outliers cannot be guaranteed to be handled, i. e. the estimator break down in the presence of already a few outliers.
- The weight function in eq. (1.9) depends on the residuals. As the residuals may have extremely different standard deviation due to $\sigma_{\widehat{\varepsilon_i}} = \sigma_{l_i}\sqrt{r_i}$, errors at leverage points cannot be detected easily, the lower bound for detectable errors is even a factor $1/\sqrt{r_i}$ larger than eq. (1.4). Replacing the argument by the standardized residual z_i does not really help. This indicates that *robust estimation requires a homogeneous geometry which can only be checked using diagnostic tools.*

Estimators with High Breakdown Point

The need for handling multiple outliers motivated the search for high break down estimators. The *break down point* ε^* of an estimator is defined as the

smallest fraction of contamination that can cause the estimator to take values arbitrarily far from the true value. The break down point of the least squares estimator is $1/n$, the break down point for the L_1norm depends on the number u of unknown parameters: the median ($u = 1$) has break down point 50 %, while for $u \geq 2$ the break down point is $1/n$. Thus both types of estimators do not allow for more than one outlier in the general case, which of course is a theoretical statement.

Actually there exist estimators which in general have a break down point close to 50 %. A prominent one is the least median square estimator (LMedS, cf. [12]). Instead of the *sum* of the squared residuals the *median* is minimized, or equivalently

$$\text{med}\,(|b_i - f_i(\widehat{\boldsymbol{p}})|) \rightarrow \min \tag{1.10}$$

The breakdown point is $\varepsilon^* = ([n/2] - p + 2)/n$ which approaches 50 % for $n \rightarrow \infty$.

In the 1D-case the LMedS-estimator can easily be realized by sorting the data and looking for the shortest interval containing 50 % of the data. For $u > 1$ a random sampling needs to be performed: Minimal sets $\boldsymbol{l}^m$ of u observations are chosen randomly, from which by inversion of $\boldsymbol{f}$ an estimate $\widehat{\boldsymbol{p}}^m = \boldsymbol{f}^{-1}(\boldsymbol{l}^m)$ is derived, an estimate which has to be confirmed by a certain fraction of the observations not used in this estimation step.

Unfortunately the number of trials necessary to find a set of non-contaminated observations grows exponentially with u, making the method only feasible for small u. But then it has shown to be extremely powerful for solving vision tasks (cf. e. g. [14], [8], [17]).

Comparison: When comparing the strategies for eliminating outliers one may have a look at the state space of correct and false observations: There are $\binom{k}{n}$ possibilities to select k outliers out of n observations, indicating the search space to have 2^n entries. Classical testing of residuals and iterative elimination corresponds to a best first search for false observations, possibly allowing a backtracking, the randam sampling strategy of LMedS-estimators corresponds to a bottom up search for good observations. Clustering methods, such as the Hough-transform not discussed here, correspond to a complete search.

All these methods are symbolic in the sense, that the system can identify symbolically which observations are assumed to be correct or false. This is in contrast to the subsymbolic method of ML-type estimators where by a reweighting scheme the observations take part in the estimation to some degree, depending on their weight. Obviously symbolic, thus search methods show high algorithmic complexity which in principal is exponential in n, restricting their application to problems with a few parameters, with the advantage of being able to explain their procedure and result. Subsymbolic methods are much more efficient, but are restricted in handling large errors and cannot explain their behaviour.

1.4 Open Questions

The following questions seem to be essential for the development and the use of robust estimation techniques and for the promotion of performance evaluation of vision algorithms.

1. **The problem of modelling unforeseen outliers.**

   Several procedures assume a well defined outlier model:

   - Are there procedures which can handle unforseen disturbances?
   - Is it possible to find such procedures?
   - How can *background*, *clutter*, *other objects* be modelled in a generic way in order to at least be able to identify such parts/objects?

2. **The problem of evaluating robust procedures.**

   Robust procedures inherently are nonlinear.

   - Is there a conceptual basis for predicting the performance of robust procedures under realistic conditions?
   - Is there a methodology to choose, link, and apply robust procedures?

3. **The problem of symbolic and subsymbolic procedures.**

   Robust estimation techniques may be classified into subsymbolic and symbolic procedures, indicating whether the state space of observational values is continuous or discrete. E. g. reweighting schemes are subsymbolic, heuristic search techniques are symbolic.

   - Under which conditions are techniques conceptually superior to another?
   - Which role has *understanding* the decisions of a robust technique for image analysis/understanding? (e. g. when being able to *name* the constraints which have been violated by the data)
   - Is there a framework to mix both types of techniques in order to exploit their individual advantages?

4. **The problem of documentation of robust procedures.**

   Several robust procedures for increasing the performance have been proposed and will be proposed. The following questions seem to hold for any kind of vision algorithm which is meant to solve part of an engineering problem.

   - Is there a commonly accepted notion of quality/performance of a vision algorithm?
   - What should researchers report if they have designed a new vision algorithm?
   - What type of generic tests/experiments should be performed to illustrate/prove the quality/superiority of a new vision algorithm?

References

1. W. Baarda. *Statistical Concepts in Geodesy*, Volume 2 of Series *4*. Netherlands Geodetic Commission, 1967.
2. W. Baarda. *A Testing Procedure for Use in Geodetic Networks*, Volume 5 of Series *1*. Netherlands Geodetic Commission, 1968.
3. A. Blake. Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction. *IEEE T-PAMI*, 11(1):2–12, 1989.
4. W. Förstner. Reliablitiy and Discernability of Extended Gauss-Markov Models. In F. Ackermann, editor, *Seminar "Mathematical Models of Geodetic/Photogrammetric Point Determiation with regard to Outliers and Systematic Errors"*, 79–103, München, 1983. Deutsche Geodätische Kommission, A98.
5. W. Förstner. Reliability Analysis of Parameter Estimation in Linear Models with Applications to Mensuration Problems in Computer Vision. *Computer Vision, Graphics & Image Processing*, 40:273–310, 1987.
6. W. Förstner. Image Analysis Techiniques for Digital Photogrammetry. In *Schriftenreihe d. Institut für Photogrammetrie, Universität Stuttgart, Heft 13*. Institut für Photogrammetrie, Universität Stuttgart, 1989.
7. W. Förstner. Diagnostics and Performance Evaluation in Computer Vision. In *Proc. Performance versus Methodology in Computer Vision, NSF/ARPA Workshop, Seattle*, 11–25. IEEE Computer Society, jun 1994.
8. Roth G. and Levine M. D. Random Sampling for Primitive Extraction. In *Workshop on Robust Computer Vision, Seattle*, 352–366, 1990.
9. D. Geiger and A. Yuille. A Common Framework for Segmentation. *IJCV*, 6(3):227–243, 1991.
10. F. R. Hampel, E. M. Ronchetty, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach based on Influence Functions*. Wiley NY, 1986.
11. P. J. Huber. Between Robustness and Diagnostics. In W. Stahel and S. Weisberg, editors, *Directions in Robust Statistics and Diagnostics*, 121–130. Springer verlag, 1991.
12. Rousseeuw P. J. and Leroy A. M. *Robust Regression and Outlier Detection*. Wiley NY, 1987.
13. Y. G. Leclerc. Constructing Simple Stable Descriptions for Image Partitioning. *IJCV*, 3(1):73–102, 1989.
14. P. Meer, D. Mintz, and A. Rosenfeld. Least Median of Squares Based Robust Analysis of Image Structure. Technical Report CAR-TR-490, University of Maryland, Computer Vision Laboratory, Center for Automation Research, College Park, MD 20742, Mar. 1990.
15. D. Pregibon. A DIY Guide to Statistical Strategy. In *Artificial Intelligence and Statistics*, 389–400. Addison-Wesley, 1986.
16. W. Schickler. Feature Matching for Outer Orientation of Single Images Using 3-D Wireframe Controlpoints. In *Int. Arch. f. Photogrammetry and Remote Sensing, Vol. 29, III*, 1992.
17. B. G. Schunck. Robust Computational Vision. In *Proc. of IEEE Workshop on Robust Computer Vision*, 1–18. IEEE Computer Society Press, Sept/Oct. 1990.