



## **BenCOS**

Performance Evaluation and Benchmarking of  
Algorithms or Systems for  
Calibration, Orientation and Surface Reconstruction

Beijing, Oct 15th, 2005

Wolfgang Förstner



## **Outline**

- **Motivation**
- **Notions**
- **Calibration and Orientation**
- **Surface Reconstruction**
- **Test data and statistical parameters**
- **Establishing teams**

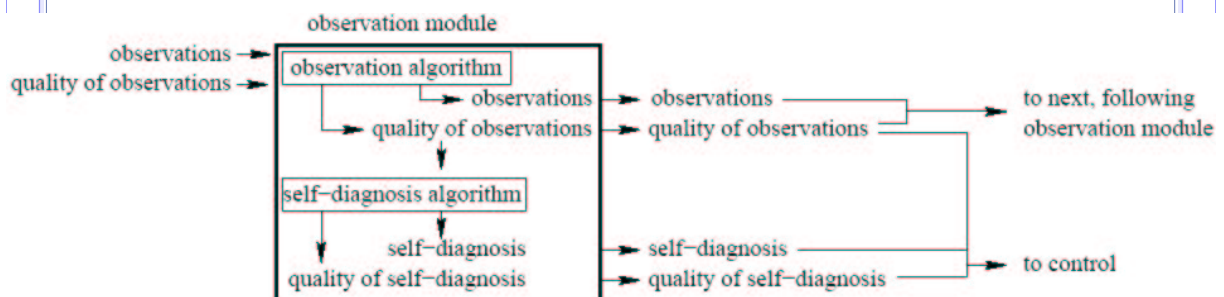


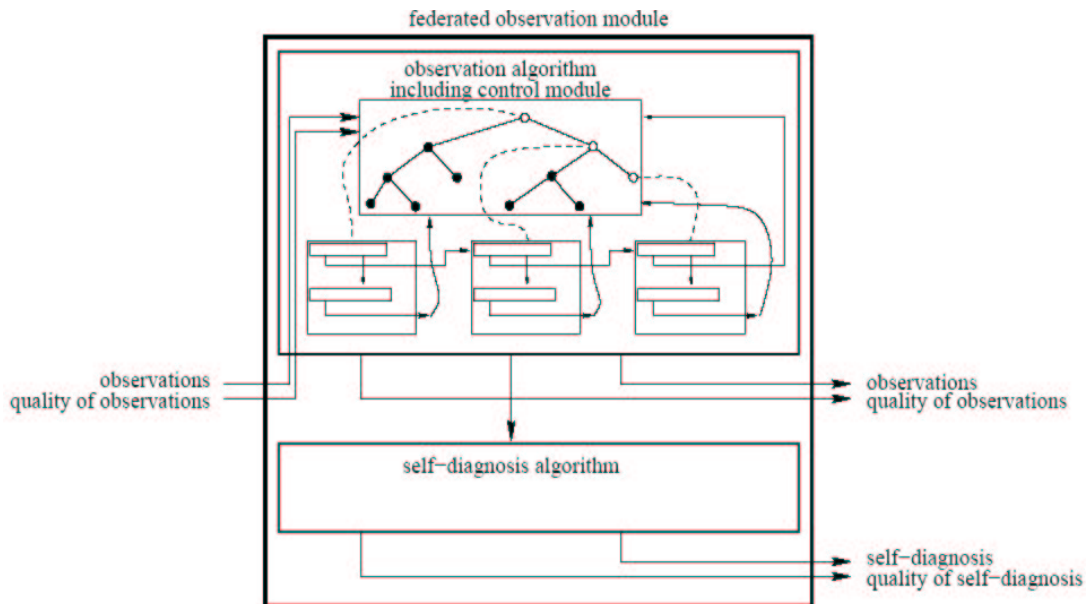
- **Performance**
- **Performance Characteristics**
- **Performance Evaluation**
- **Benchmarking**



## The situation: Self-diagnosis in vision/observation modules

- **Observation/measurement/parameter/class**
- **Quality of observation (e. g. precision)**
- **Self-diagnosis (e. g. certainty of success)**
- **Quality of self-diagnosis (e. g. variance of precision)**





- **Control development of modules (internal)**
  - Prove increase of performance
  - Find weakest points
  - Select best module
- **Increase value of modules (external)**
  - Provide self-diagnostic features
  - Characterise performance
  - Report on benchmarks
  - Prove superiority



- **Performance**

- Properties relevant for an application
- Related to time, space, money, ...
- Examples:
  - Precision: Covariance matrices, percentiles, ...
  - Accuracy: including bias, referred to some reference, ...
  - Robustness: detectability, sensitivity, successrate
  - Speed

- **Requires theory:**

consistent representation of uncertainty



- **Performance Characteristics**

- Description of performance =  $f(\text{boundary conditions})$
- No *comparison* to evaluation
- Examples
  - Specification sheet:
    - precision =  $f(\text{image quality})$
    - Reliability =  $f(\text{overlap})$
  - Meta data: what designed for, ist best use, ...
    - Designed for: indoor images
    - Best use: only edges, no texture

- **Requires empirical studies**

- Estimation of parameters characterizing performance
- Simulated, partially simulated and real data



- **Performance Evaluation**
  - Usefulness (for an application)
  - Admissability (w. r. t. a specification)
  - „Costs“ (availability, reference to a CPU)
- **Requires specifications**
  - Applications
  - Requirements
  - Budget



- **Benchmarking**
  - = comparing performance of several modules
  - Same module at different stages
  - Modules of different authors
- **Requires agreement**
  - On performance measures
  - On test data
- **Part of characterization**
- **Not performance evaluation**



### Views on calibration, orientation and reconstruction

- **Parameter estimation**
- **Detection and matching processes**
- **Function estimation**



- **Parameter Estimation**
  - Limited and known number of parameters  
→ complete specification of reference data
  - First and second order statistics  
→ Testing theory for deciding on differences
- **Benchmarking is easy (still work)**



- **Detection and matching processes**
  - Unknown number of observations  
→ no explicit specification of reference data possible
  - Spatial statistics of 2D and 3D-point clouds  
→ incomplete statistical evaluation
- **Benchmarking becomes tricky**  
Agreement on performance measures?



- **Surface Estimation**
  - Part of infinite space ( $\mathbb{R}^2$ )  
→ statistics awkward
  - Complex topology (holes, occlusions)  
→ mixture of continuous and discrete representations
  - Possibly full 3D  
→ 2 ½ D not sufficient, no graph surfaces
  - Possibly mixture of algebraic and general surfaces
- **Benchmarking requires reduction**
  - Difficult agreement on representative points/lines
  - Reduction of statistics difficult



Strecha 2003



## Suggestions

- **Agreements**
- **Performance measures**



- **Compare parameters**
  - Calibration parameters
  - Orientation parameters
- **Characterize detection and matching indirectly**
  - Robustness
  - Flexibility
  - Completeness



- **Precision:**

Effect of random errors: variances, internal
- **Accuracy:**

Including bias: mean deviations, significance tests  $\leftrightarrow$  reference
- **Sensitivity:**

effect of model errors onto result: leave one-out test (image)
- **Parameters:**
  - Calibration and orientation
  - not 3D-points



## Suggestions

- **Agreements**
- **Performance measures**



- **Compare parameters**
  - Grid of surface (position and normal)
  - Grid of occlusion boundaries (position and normal)
- **Restrictions**
  - Separate tests for 2 ½ and 3D-surfaces
  - Well behaved surfaces
    - no fractals, clouds, transparent surfaces
    - But shiny, ...
- **Characterize detection and matching indirectly (cf. ↑)**
  - Robustness
  - Flexibility
  - Completeness



- **Same as for orientation**
  - Precision
  - Accuracy
  - Sensitivity



- **Following Maimone 1996**
  1. Synthetic, error free, full control
  2. Synthetic, random errors, full control
  3. Partially synthetic full control
  4. Real data, full control
  5. Real data, partial control
  6. Real data, no control
- **Program check: 1-4**
- **Quantitative performance characteristics: 2-5**
- **Qualitative performance characteristics: 2-6**
- **Benchmarking: 1-6**



- **Theoretical precision**  
For planning
- **Empirical precision without ground truth**  
For internal evaluation (self-diagnosis)
- **Empirical precision with ground truth (accuracy)**  
For performance characteristics

similar:

- **Theoretical sensitivity**
  - **Empirical sensitivity without ground truth**
  - **Empirical sensitivity with ground truth**
- (Förstner 2001)



- **Predicted by variance propagation**  
Cramer-Rao bound from inverse normal equation matrix

$$C_{pp} = (A' C_{ll} A)^{-1}$$

$A$  = Jacobian = **Design** matrix,

$C_{ll}$  = assumed covariance matrix of observations

$C_{pp}$  = predicted covariance matrix of parameters

- **No real data required**
- **May be replaced by Monte Carlo**  
→ information about effect of non-linearities
- **Yields full distribution**
- **Often good approximation by covariance matrices**



- **Exploits internal redundancy**
  - Geometric constraints from multiple view geometry
  - Geometric constraints from object structure
- **Estimated precision from residuals is most simple**
  - One number for evaluation difference data  $\leftrightarrow$  model
- **Test on significance of parameters**
  - Calibration (camera), additional constraints (object)
- **Local estimates**
  - Local precision (e. g. of points or images) = f(residuals and design)
- **Variance component estimation**
  - estimates variances of groups of observations
  - May be used to estimate  $\lambda$  in regularization schemes



- Necessary for model checking**  
*compares data-model fit*
- **Evaluate differences**  $d = p_{ref} - p_{est}$
  - **Exploit knowledge of covariance structure**
    - $T = d' C_{dd}^{-1} d \sim \chi^2$
    - Mahalanobis distance
    - $C_{dd} = C_{dd,ref} + C_{dd,est}$
    - takes uncertainty of reference data into account
  - **Applied to subvectors or functions of parameters**
    - Points
    - Object parameters
    - Calibration or orientation parameters
    - Subregions



- **Expected effect of model errors onto**
  - Bias and covariance
  - Possibly distribution
  - Reference to precision (e. g. multiples of st. dev.)
- **Leave-one out test**
  - Individual observations (leverage points)
  - Individual parameters (model errors)
  - Groups of observations (points, images)
  - Groups of parameters (calibration, object constraints)
- **Realizable by Kalman-Filter**
  - Add observations = leave out constraints: positive  $C_{ll}$
  - Delete observations = add constraints: negative  $C_{ll}$   
→ Efficient computation



- **Real effect of model errors onto result**
  - Without ground truth:  
Change of parameters, subset of parameters
  - With ground truth  
difference of parameters to ground truth
- **Leave-one-out test**
- **Realizable as Kalman-Filter**



### Role of Participants

- **Initialization**

- Provide images (test data)  
Yields broad sample,
- Provide meta-data  
calibration, boundary conditions, possibly reference data,

- **Test**

- Provide results (data, result of self-diagnosis)  
On all data sets → cross validation
- Provide meta-data (answers to query)



### Role of Coordinator

- **Specification**

- Formats  
sizes of images, calibration, orientation data, ..
- Selection criteria for images

- **Check/Determination of ground truth**

- e. g. based on matches of all the participants
- Self-calibrating bundle adjustment
- Artificial images  
Computer Graphics
- Semi-natural images  
Artificial geometry with natural texture

- **Determination of performance measures**

- **Comparison**



- **Reference data**  
**High resolution / redundancy data set**  
Example:
  - > 30 images of one object
  - Resolution > 2000 x 2000 pixels
  - High overlap
- **Test data**  
**Subset with lower resolution**  
Example
  - 4-6 images of object
  - Resolution < 700 x 700 pixels
  - Lower overlap
- **Expected quality ratio: > 2**



- [Henrik I. Christensen, Wolfgang Förstner: Performance characteristics of vision algorithms, Machine Vision and Applications, V.9 n.5-6, p.215-218, 1997](#)
- *Förstner, W.: 10 Pros and Cons Against Performance Characterization of Vision Algorithms, [ECCV Workshop on Performance Characteristics of Vision Algorithms](#), 1996*
- *Förstner, W.: Image Preprocessing for Feature Extraction in Digital Intensity, Color and Range Images, Summer School on [Data Analysis and the Statistical Foundations of Geomatics](#), Chania, Crete, Greece, 1998, Springer Lecture Notes on Earth Sciences*
- *[Wolfgang Förstner](#): Generic Estimation Procedures for Orientation with Minimum and Redundant Information, [Calibration and Orientation of Cameras in Computer Vision](#), [A. Gruen](#) and [T. S. Huang](#) (Eds.), Springer 2001*
- *[Wolfgang Förstner](#): Computer Vision and Photogrammetry --- Mutual Questions: Geometry, Statistics and Cognition, in: [Bildtechnik/Image Science](#), Swedish Society for Photogrammetry and Remote Sensing, 2002 (pp. 151-164)*
- *[M. Maimone](#) and [S. Shafer](#): [A Taxonomy for Stereo Computer Vision Experiments](#), [ECCV Workshop on Performance Characteristics of Vision Algorithms](#), pp 59-79.*