Photogrammetry & Robotics Lab

Bundle Adjustment – Part I Introduction & Application

Cyrill Stachniss

5 Minute Preparation for Today

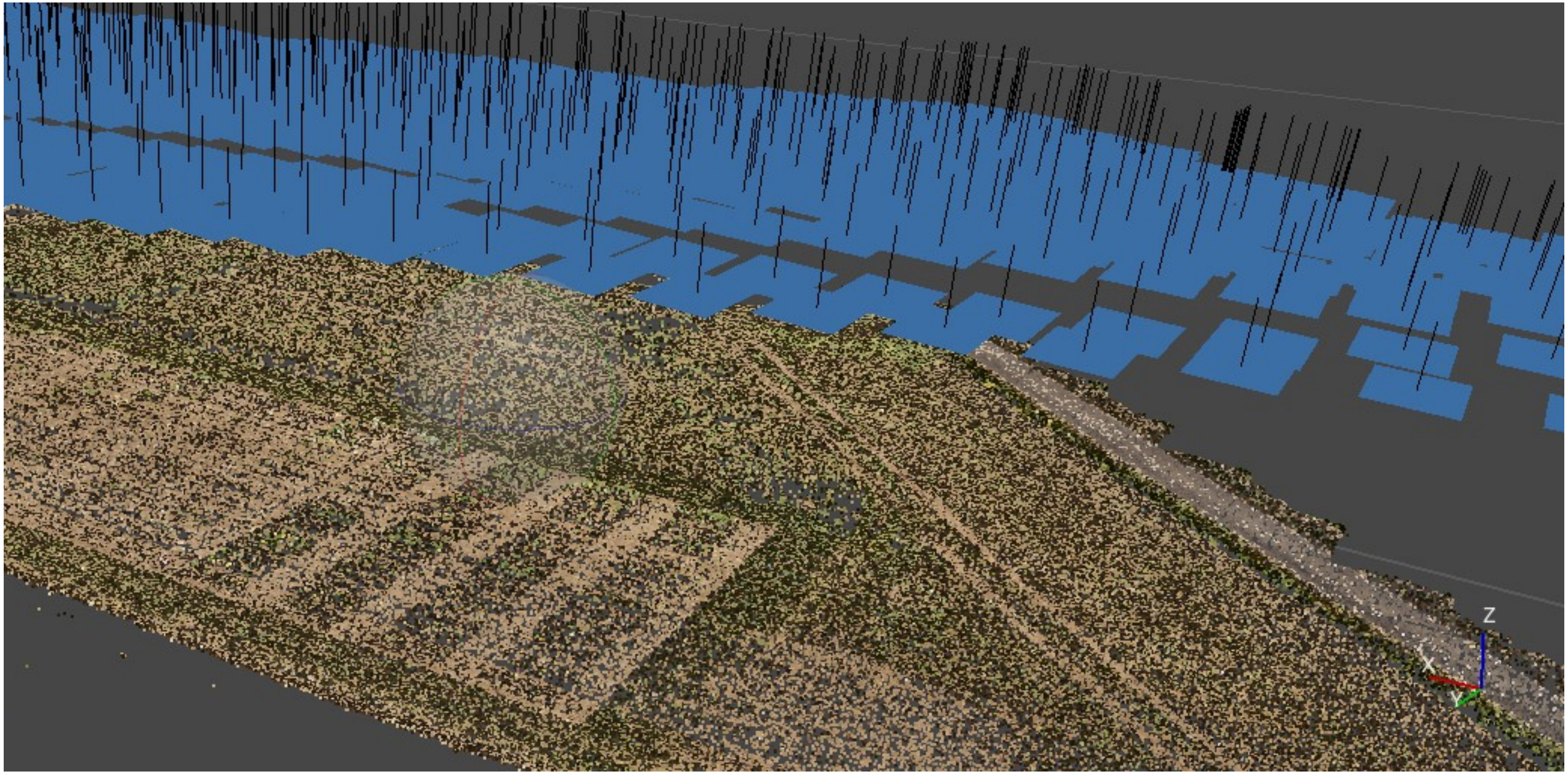
https://www.ipb.uni-bonn.de/5min/

3D Reconstruction

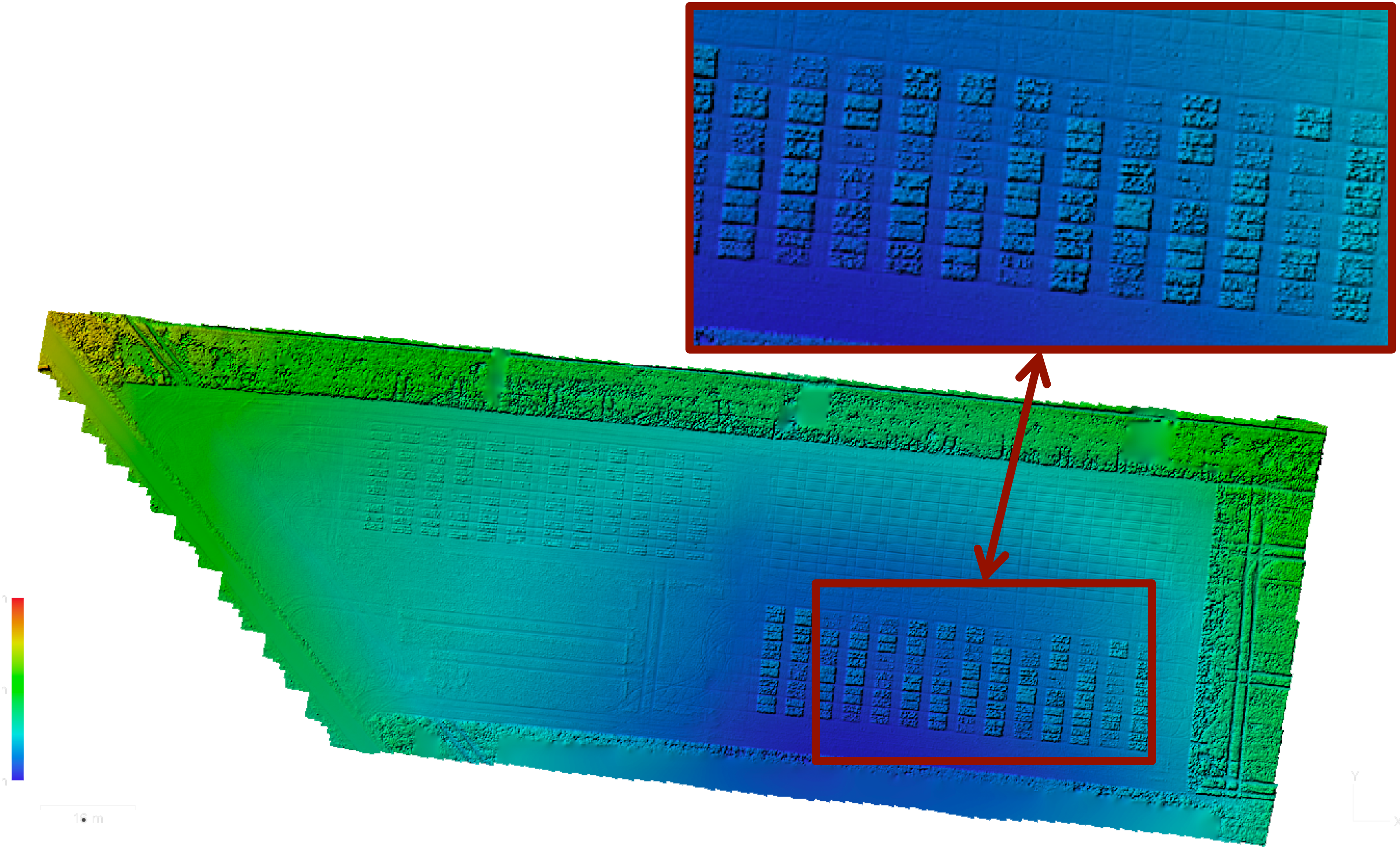
- Cameras can be used as sensors for 3D reconstruction
- So far, we only used **image pairs**
- The next step is to look into 3D **reconstruction using N>2 images**

UAV Mapping Example

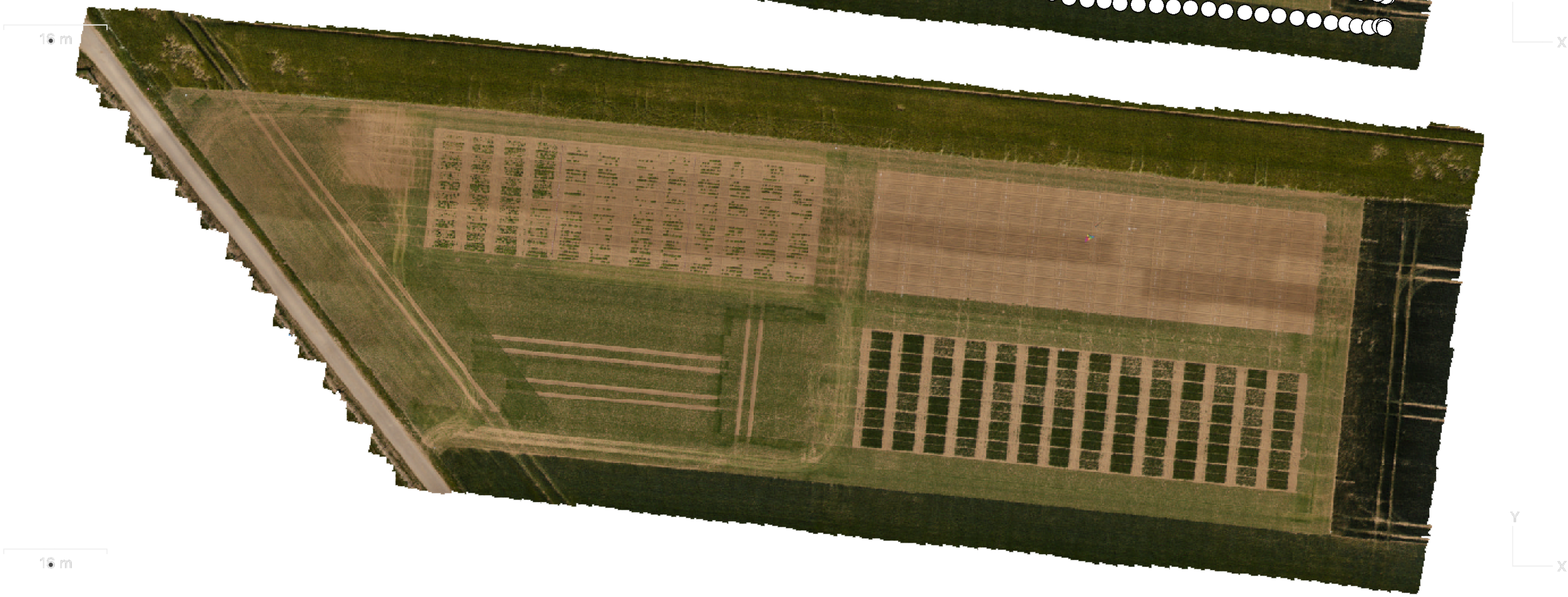
UAV Mapping Example



Digital Surface Model



Orthophoto



Why
Multi-View
Reconstruction?

Why 3D Reconstruction from Multiple Views?

- Multiple images may be needed to cover the whole surface of an object
- Precision requirements are higher than those obtainable by one image pair
- Level of detail higher than one image
- Estimate the motion of a vehicle/robot and the map of the environment

Example

Video courtesy: van Gool / GeoAutomation

Bundle (Block) Adjustment

- Bundle adjustment is a least squares solution for orienting images and estimating 3D point locations
- Bundle “block” adjustment: multiple images are corrected “en bloc”
- Note: often the term “block” is dropped in modern literature

Bundle Adjustment for Aerial Triangulation

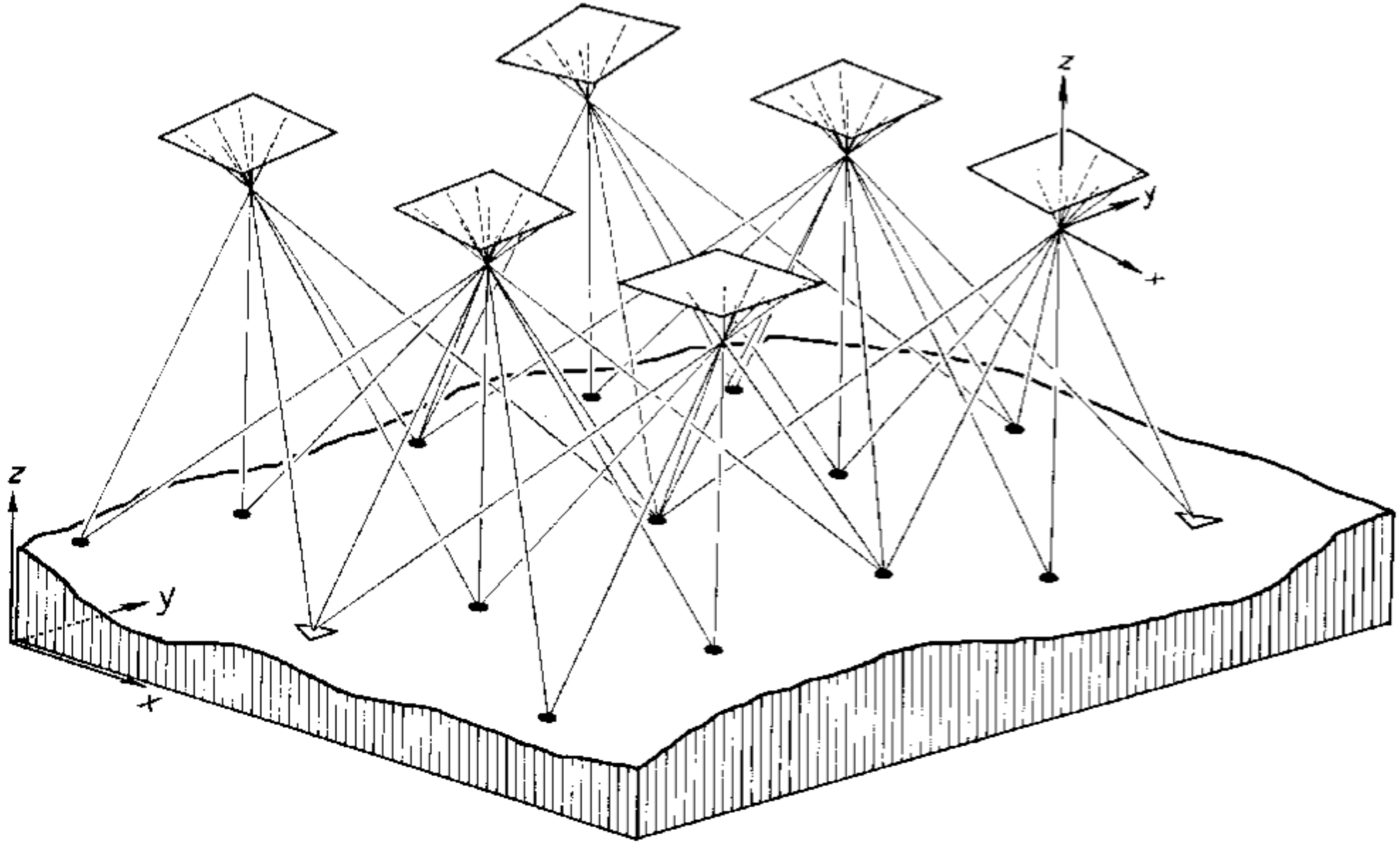


Image courtesy: Ackermann

Aerial Triangulation (AT)

- AT is the task of estimating the 3D location of points using aerial images as well as the camera parameters
- Bundle adjustment is default technique for this problem
- Automated software solutions available

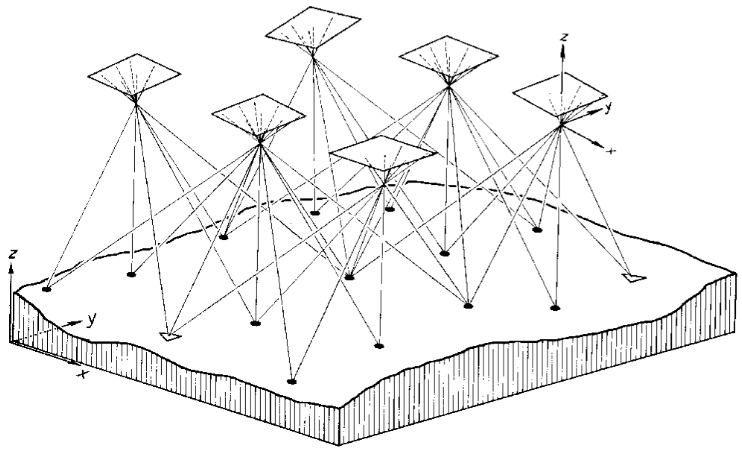
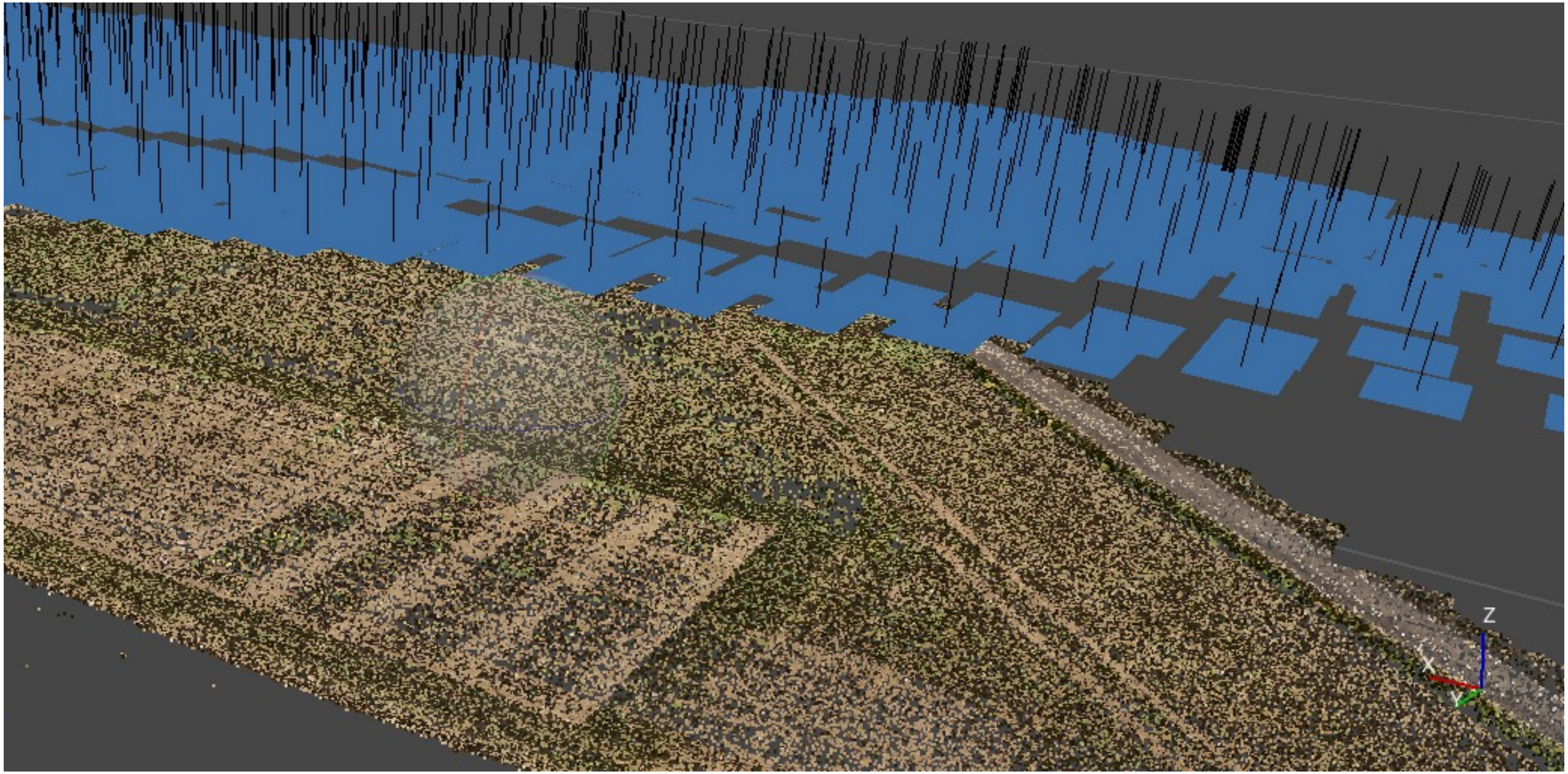


Image courtesy: Ackermann

Bundle Adjustment for Aerial Triangulation



How Does
Bundle Adjustment Work?

Bundle Adjustment

Least squares approach to estimating camera poses and 3D points

Key idea:

- Start with an initial guess
- Project the estimated 3D points into the estimated camera images
- Compare locations of the projected 3D points with measured (2D) ones
- Adjust to minimize error in the images

Reprojection Error

BA is a non-linear least squares approach

$$ {}^a\mathbf{x}'_{ij} + {}^a\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\; {}^a\widehat{\mathsf{P}}_j(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\; \widehat{\mathbf{X}}_i $$

a="arbitrary frame"

img pix.

project. params

distortion

ij = point i observed in image j

corrections

scale factor

projection matrix (w/ non-lin. calib.)

3D point

$$ \text{with } \Sigma_{x_{ij}x_{ij}} \qquad i = 1, ..., I_j; j = 1, ..., J $$

uncertainty in the image coordinates

#points in image j

#images

Reprojection Error

- Non-linear least squares approach

$$ {}^{a}\mathbf{x}'_{ij} + {}^{a}\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\ {}^{a}\widehat{\mathsf{P}}_{j}(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\ \widehat{\mathbf{X}}_{i} $$

$$ \text{with } \Sigma_{x_{ij}x_{ij}}, i = 1, ..., I_j; j = 1, ..., J $$

- Encodes the carmera **projection**
- Error is the distance between each projected 3D and measured 2D point
- Encodes the **collinearity constraint**
- **Known** data association

Unknown Parameters

- Non-linear least squares approach

$$ {}^a\mathbf{x}'_{ij} + {}^a\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\ {}^a\widehat{\mathsf{P}}_j(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\ \widehat{\mathbf{X}}_i $$

Unknowns:

- 3D locations of new points $\widehat{\mathbf{X}}_i$
- 1D scale factor $\widehat{\lambda}_{ij}$
- 6D exterior orientation
- 5D projection parameters (interior o.)
- Non-linear distortion parameters $\boldsymbol{q}$

Reprojection Error

- Non-linear least squares approach

$$ {}^{a}\mathbf{x}'_{ij} + {}^{a}\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\ {}^{a}\widehat{\mathsf{P}}_{j}(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\ \widehat{\mathbf{X}}_{i} $$

$$ {}^{a}\mathbf{x}'_{ij} + {}^{a}\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\widehat{\mathsf{K}}(\boldsymbol{x}_{ij}, \widehat{\boldsymbol{p}}, \widehat{\boldsymbol{q}})\ \widehat{R}_{j}\ [I_3 | -\widehat{\boldsymbol{X}}_{0j}]\ \widehat{\mathbf{X}}_{i} $$

calibration (interior orientation)

exterior orientation

Example

- 10k images, 1k points per image
- Each point seen 10 times on avg.
- **How many unknowns do we have?**

$$ {}^{a}\mathbf{x}'_{ij} + {}^{a}\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\ {}^{a}\widehat{\mathsf{P}}_{j}(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\ \widehat{\mathbf{X}}_{i} $$

Example

- 10k images, 1k points per image
- Each point seen 10 times on avg.
- **How many unknowns do we have?**

$$ {}^{a}\mathbf{x}'_{ij} + {}^{a}\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\ {}^{a}\widehat{\mathsf{P}}_{j}(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\ \widehat{\mathbf{X}}_{i} $$

- DoF observations: 2 x 10k x 1k = 20M
- Unknowns: ~1M (pts) w/ 3 DoF + 10M (scale) + 10k (orientations) w/ 6 DoF
- **~13 Mio unknown dimensions and ~20 Mio observations**

Eliminating the Scale Factors

We can eliminate the per-point scale factor by using Euclidian coordinates (instead of homogenous coordinates)

$$ {}^{a}\mathbf{x}'_{ij} + {}^{a}\widehat{\mathbf{v}}_{x'_{ij}} = \widehat{\lambda}_{ij}\; {}^{a}\widehat{\mathsf{P}}_{j}(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\; \widehat{\mathbf{X}}_{i} $$

$$ {}^{a}\boldsymbol{x}'_{ij} + {}^{a}\widehat{\boldsymbol{v}}_{x'_{ij}} = \frac{{}^{a}\widehat{\mathsf{P}_{1:2}}_{j}(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\; \widehat{\mathbf{X}}_{i}}{{}^{a}\widehat{\mathsf{P}_{3}}_{j}(\boldsymbol{x}_{ij}, \boldsymbol{p}, \boldsymbol{q})\; \widehat{\mathbf{X}}_{i}} $$

Example: ~13M unknowns reduce to ~3M unknowns

Setting Up and Solving the System of Normal Equations

- Standard procedure…
- With unknowns $\boldsymbol{x}$ and observations $\boldsymbol{l}$
- Setup the normal equations

$$A^{\mathsf{T}}\Sigma^{-1}A\,\Delta\boldsymbol{x} = A^{\mathsf{T}}\Sigma^{-1}\Delta\boldsymbol{l}$$

unknowns observations

- This yields the estimate

$$\widehat{\Delta\boldsymbol{x}} = (A^{\mathsf{T}}\Sigma^{-1}A)^{-1}A^{\mathsf{T}}\Sigma^{-1}\Delta\boldsymbol{l}$$

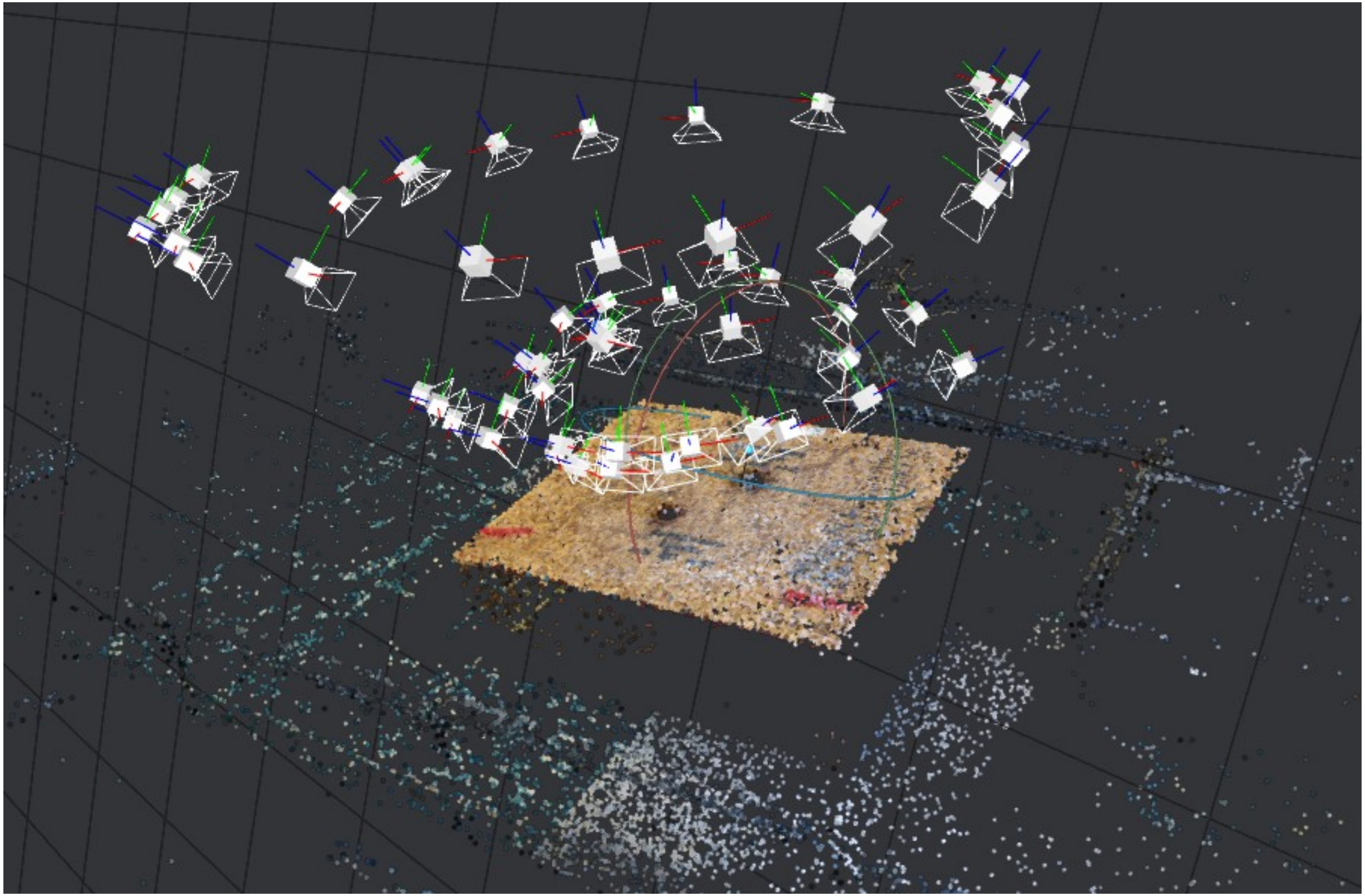
unknowns observations

Numerics of BA

- The resulting linear system gets **huge**
- Out-of-the-box application will not be able to solve BA efficiently

Next lecture: Numerics of BA

Camera Poses and 3D Points



3D Model & Estimated Surfaces

Optimality

- BA is **statistically optimal**
- Exploits all observations and considers the uncertainties and correlations
- Exploits all available information
- Computes orientations, calibration parameters, and point locations with highest precision
- Assumes Gaussian noise
- Requires an initial estimate

Absolute Orientation Through Control Points

Bundle Adjustment and the Photogrammetric Model

- Using only camera images, we obtain a "photogrammetric model"
- No absolute scale information
- Unknown rotation and translation w.r.t. external reference frame
- We need to estimate a similarity transform (absolute orientation)
- Can also be done inside BA

Control Points

- Assume that some of the points are control points (= known cooridnates)
- C.P.s are needed to solve the absolute orientation problem
- How should we treat them?
  Noisy or noise-free?

Control Points: Noisy or Noise-free?

$$\boldsymbol{X}_i + \widehat{\boldsymbol{v}}_{X_i} = \widehat{\boldsymbol{X}}_i \quad \text{with } \Sigma_{X_iX_i}\ ,\ i = 1, ..., I_{CP}$$

provided coord. | correct. | estimates coord. | #control points

Pro and contra for correcting the control points?

BA with Control Points

- **Statistically optimal approach:** considers that C.P.s are noisy
- BA with fixed C.P. **enforces the geometry to match the C.P.**
- Used if the C.P.s cannot be changed (Link to official map data)
- Statistically suboptimal approach: considering fixed C.P.s

Two Step BA with Control Points

1. Step:

- BA with noisy control points
- Search for gross errors using statistical tests

$$T = \widehat{\boldsymbol{v}}_{X_i}^{\mathsf{T}} \Sigma^{-1}_{\widehat{v}_{X_i} \widehat{v}_{X_i}} \widehat{\boldsymbol{v}}_{X_i} \sim \chi^2_3$$

- Eliminate erroneous control points

2. Step:

- BA with fixed control points

How Many Control Points?

- For DLT or P3P solutions, we need 3-6 control points **per image pair**
- **How many control points are needed for bundle adjustment?**

How Many Control Points?

- For DLT or P3P solutions, we need 3-6 control points per image pair
- How many control points are needed for bundle adjustment?
- Only a small number of control points is needed
- Typically full C.P.s at the boundaries
- A central reason for using BA

Initial Guess

Initial Guess

How to obtain the initial guess?

Initial Guess

- Orientation of the image pair with a direct method
- **No direct solution is known for N views**

How to obtain the initial guess?

Computing an Initial Guess

- Compute relative orientation from the first image pair
- Compute orientation for subsequent images through P3P/RRS

Problems?

Computing an Initial Guess

- Compute relative orientation from the first image pair
- Compute orientation for subsequent images through P3P/RRS

Critical issues

- Gross error handling is essential
- Requires enough new points in each image overlap
- No singular/critical configurations

Outliers

Gross Errors / Outliers

What are reasons for gross errors?

Gross Errors / Outliers

Reasons for gross errors

- **Wrong correspondences**
- Wrong point measurements

Gross Errors / Outliers

Reasons for gross errors

- **Wrong correspondences**
- Wrong point measurements

How many observations per point do we need to deal with outliers?

Gross Errors / Outliers

Reasons for gross errors

- **Wrong correspondences**
- Wrong point measurements

Observations per point

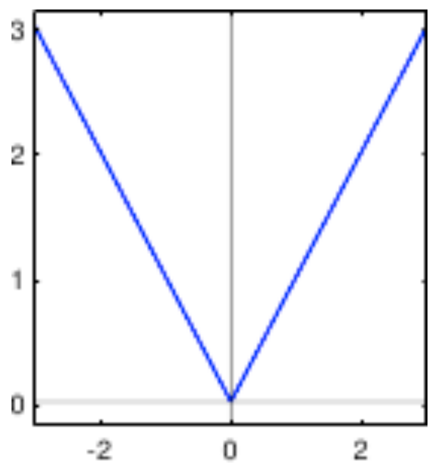
- At **least 4** views to identify the observation with a gross error
- Observed points from **5 to 6 different views** typically yield good estimates

Eliminating Gross Errors

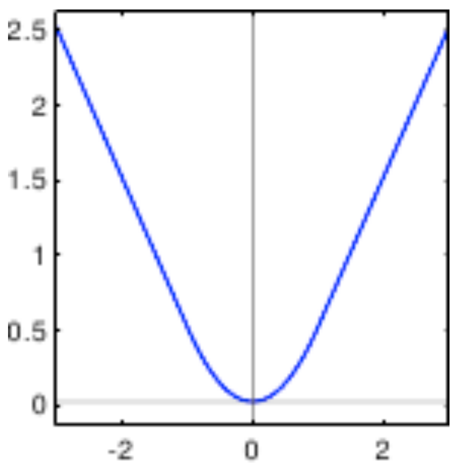
- Decompose the problem into several small blocks of 3-6 images first
- Check for gross errors in the small blocks using statistical tests
- Only consider features that can be tracked consistently in all 3-6 images
- Relative orientation (5-Point algo.) combined with RANSAC
- Eliminate gross errors, then run full BA

Robust Kernels

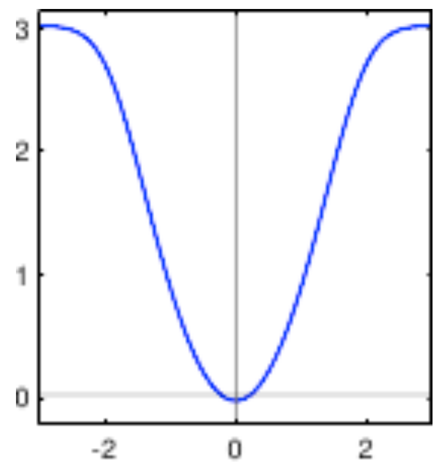
- Instead of using a Gaussian noise model, consider a robustified version
- Reduce "penalty" far away from 0



L1 norm



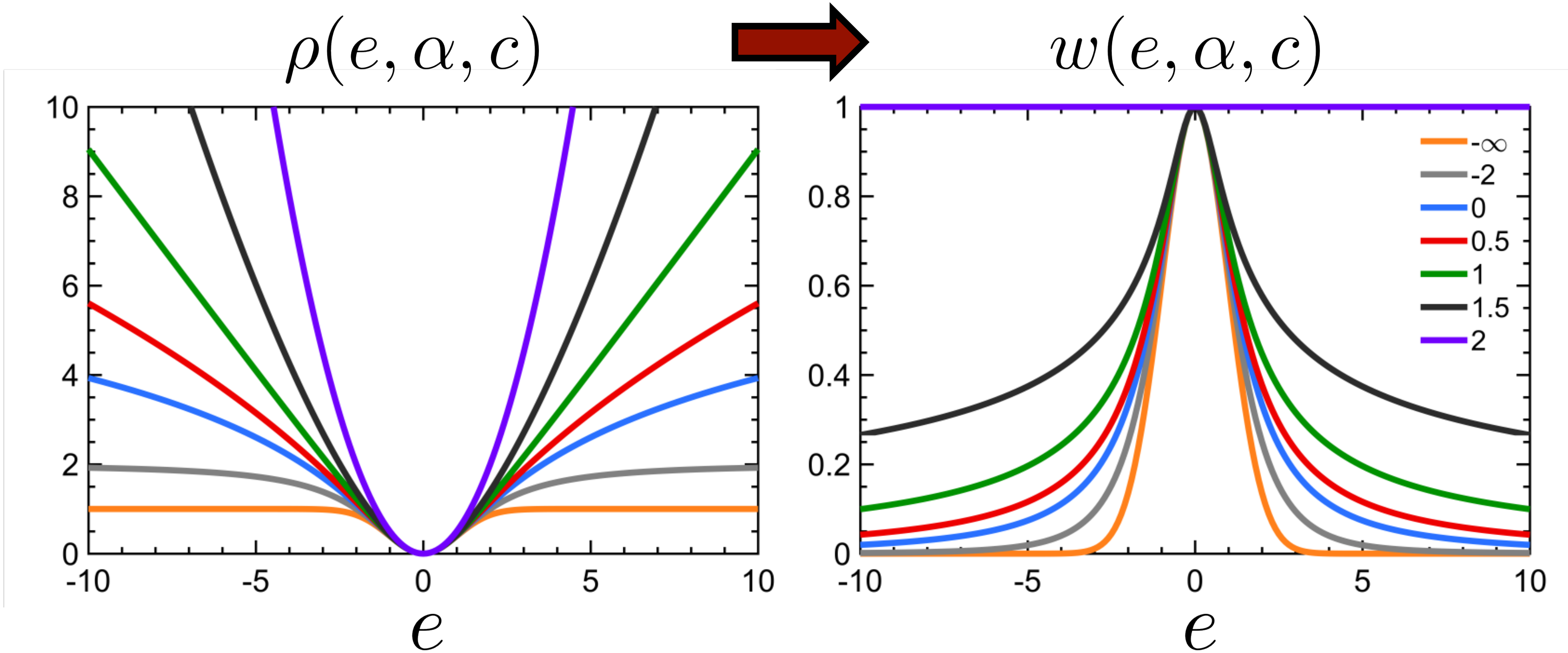
Huber



Blake-Zisserman

Robust Kernels

A robust kernel leads to weighted LS



Source: Barron, A General and Adaptive Robust Loss Function, CVPR 2019

Fully Automated Approaches

- Today, there are fully automated solutions for BA available
- Measuring and identifying corresponding points through features
- Gross error handling
- Computation of an initial guess
- Building and solving the system of linear equations
- **But:** Control point identification may require an operator

Popular Software

Open source software

- Meshroom
  Test Meshroom yourself:
  https://github.com/alicevision/meshroom

Commercial software

- Photoscan
- Pix4D

Quality of the Results

Precision

- We can compute the theoretical precision $\Sigma_{\widehat{x}\widehat{x}} = (A^\mathsf{T} \Sigma_{ll}^{-1} A)^{-1}$
- as well as the empirical precision through the variance factor

$$\widehat{\Sigma}_{\widehat{x}\widehat{x}} = \widehat{\sigma}_0^2 (A^\mathsf{T} \Sigma_{ll}^{-1} A)^{-1}$$

- How good is this precision?

Quality of the Solution

- Precision and reliability are difficult to asses as they depend on the geometric configuration of the scene
- There are some hints however…

Variance Factor

- Does the variance factor take a value around 1?
- $\widehat{\sigma}_0 = 1$ suggests a correct model
- Use a **statistical test** to judge the variance factor
- In practice, a **F-test often fails** due to the high redundancy

$$T = \frac{\widehat{\sigma_l^2}}{\underline{\sigma}_l^2} \sim F(R, \infty) \qquad \text{e.g., } R = 10000$$

Example

- Assumed $\sigma_l = 0.3$
- Estimated $\widehat{\sigma_l} = 0.35$
- Redundancy $R = 10000$

∞ means that σ_l is known!

$$T = \frac{\widehat{\sigma_l^2}}{\sigma_l^2} = 1.36 \sim F(10000, \infty)?$$

$$F(10000, \infty, 0.05) = 1.0234$$

$$\sqrt{F(10000, \infty, 0.05)} = 1.0166 < 1.36$$ **rejected!**

Variance Factor

- Does the variance factor take a value around 1?
- $\widehat{\sigma}_0 = 1$ suggests a correct model
- Use a statistical test to judge the variance factor
- In practice, a F-test often fails due to the high redundancy
- **The uncertainty in the assumed measurement noise has to be considered**

Noise in the Assumed Precision

- Consider uncertainty in precision σ_{σ_l}
- Test statistic

$$T = \frac{\widehat{\sigma_l^2}}{\underline{\sigma}_l^2} \sim F(R, \infty) \rightarrow T = \frac{\widehat{\sigma_l^2}}{\underline{\sigma}_l^2} \sim F(R, R_0)$$

- with $R_0 < \infty$
- The redundancy R_0 can be computed as it is related to the variance through

$$R_0 = \left\lceil \frac{1}{2\sigma_{\sigma_l}^2} \right\rceil$$

Example

- Assumed $\sigma_l = 0.3$
- Estimated $\widehat{\sigma_l} = 0.35$
- Redundancy $R = 10000$
- **Uncertainty in** σ_l**:** $\sigma_{\sigma_l} = 0.2$
- **This yields** $R_0 = \left\lceil (2\sigma_{\sigma_l}^2)^{-1} \right\rceil = 13$

$$T = \frac{\widehat{\sigma_l^2}}{\underline{\sigma_l^2}} = 1.36 \sim F(10000, 13)?$$

$$\sqrt{F(10000, 13, 0.05)} = 1.48 > 1.36$$ **accepted!**

Noise in the Observations

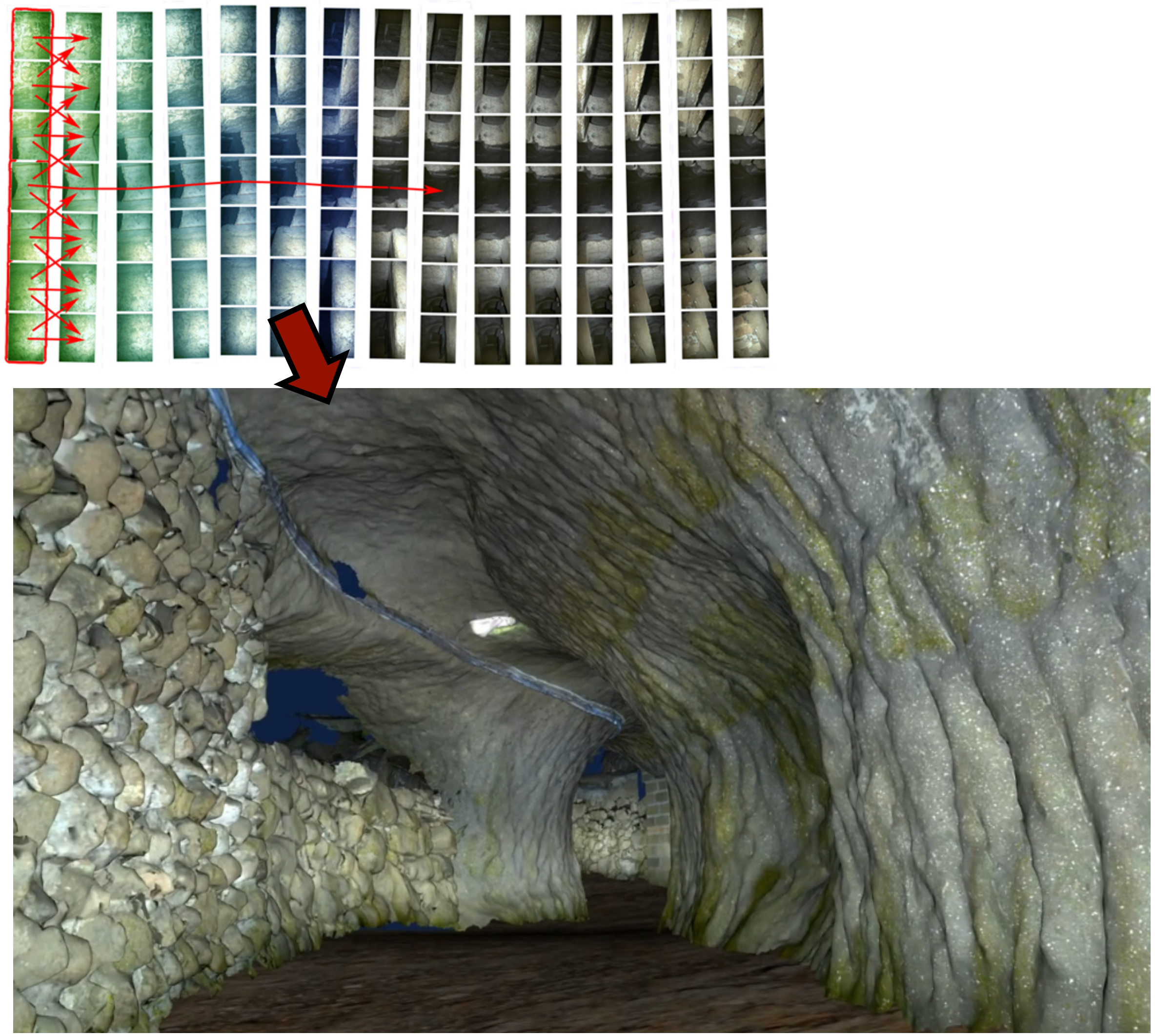
- Under a **correct functional model and no gross errors**: $\widehat{\sigma}_{l_i} = \widehat{\sigma}_0 \sigma_{l_i}$
- The high redundancy in most BA configurations yields a **realistic estimate** $\widehat{\sigma}_{l_i}$

Precision

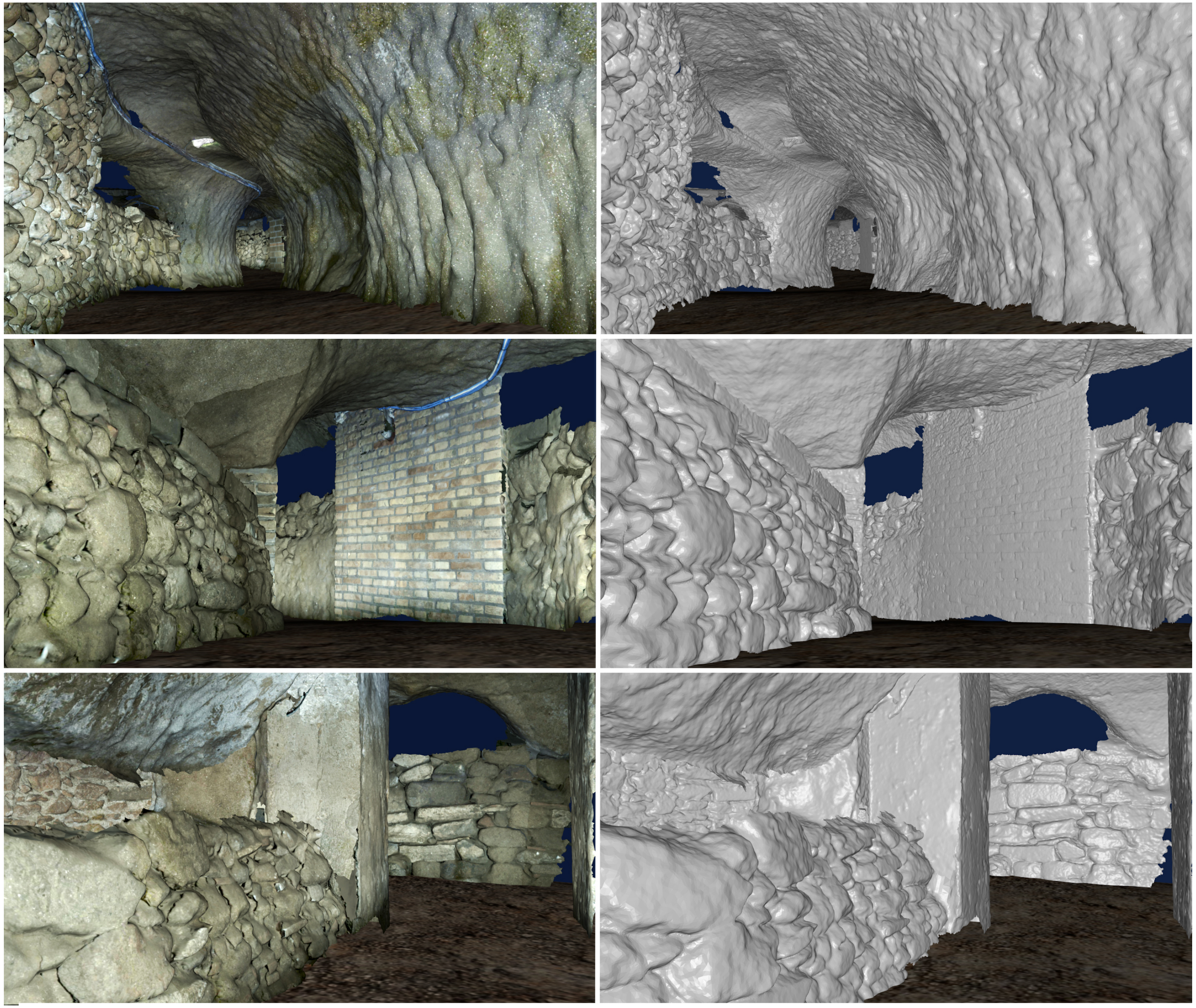
With eliminated gross error, small systematic errors, and $\widehat{\sigma}_0 = 1$, we obtain a realistic estimate for the precision of the parameters

$$\widehat{\Sigma}_{\widehat{x}\widehat{x}} = \widehat{\sigma}_0^2 (A^\mathsf{T} \Sigma_{ll}^{-1} A)^{-1}$$

Example from Robotics: Mapping of Catacombs



Dense Textured 3D Models



Example from Robotics: ORB-SLAM by Mur-Artal

ORB-SLAM

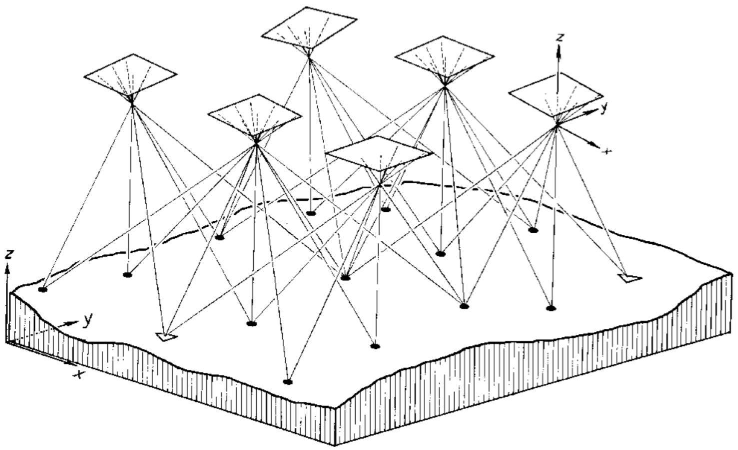
Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós

{raulmur, josemari, tardos} @unizar.es

Instituto Universitario de Investigación en Ingeniería de Aragón Universidad Zaragoza

Universidad Zaragoza 1542

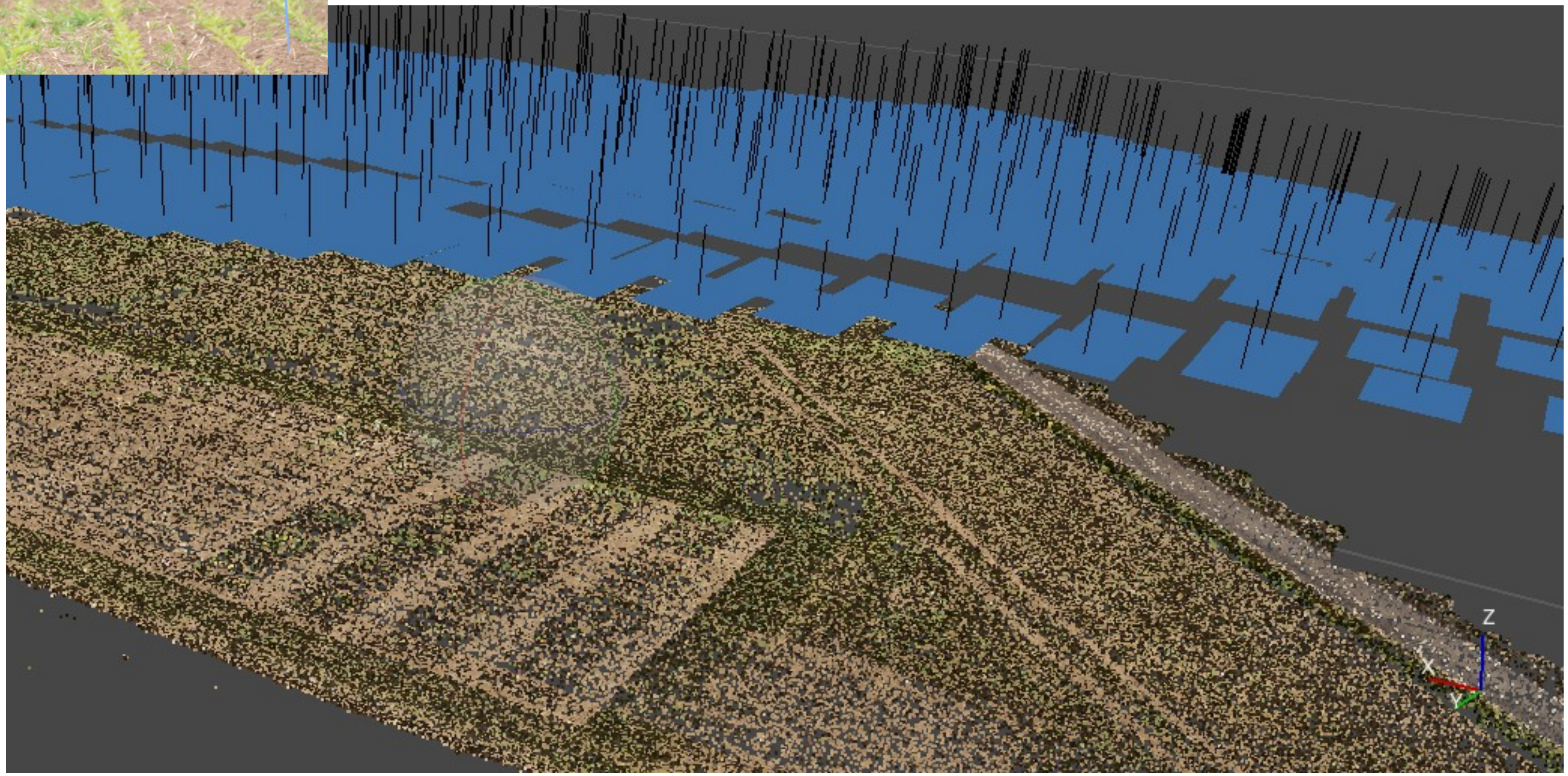
Aero/Arial Triangulation



Note: The term aero triangulation (or aerial triangulation) is derived from the terrestrial method of geodetic triangulation of large areas used for determining reference points using angular measurements with theodolites starting in the early 19th century. The notion triangulation here means the covering of an area by triangles, similar to Delauney triangulation.

Image courtesy: Ackermann

UAV Mapping Example using BA



Typical Setup

- Area is covered in stripes
- Typical overlaps: 60/20 up to 90/80

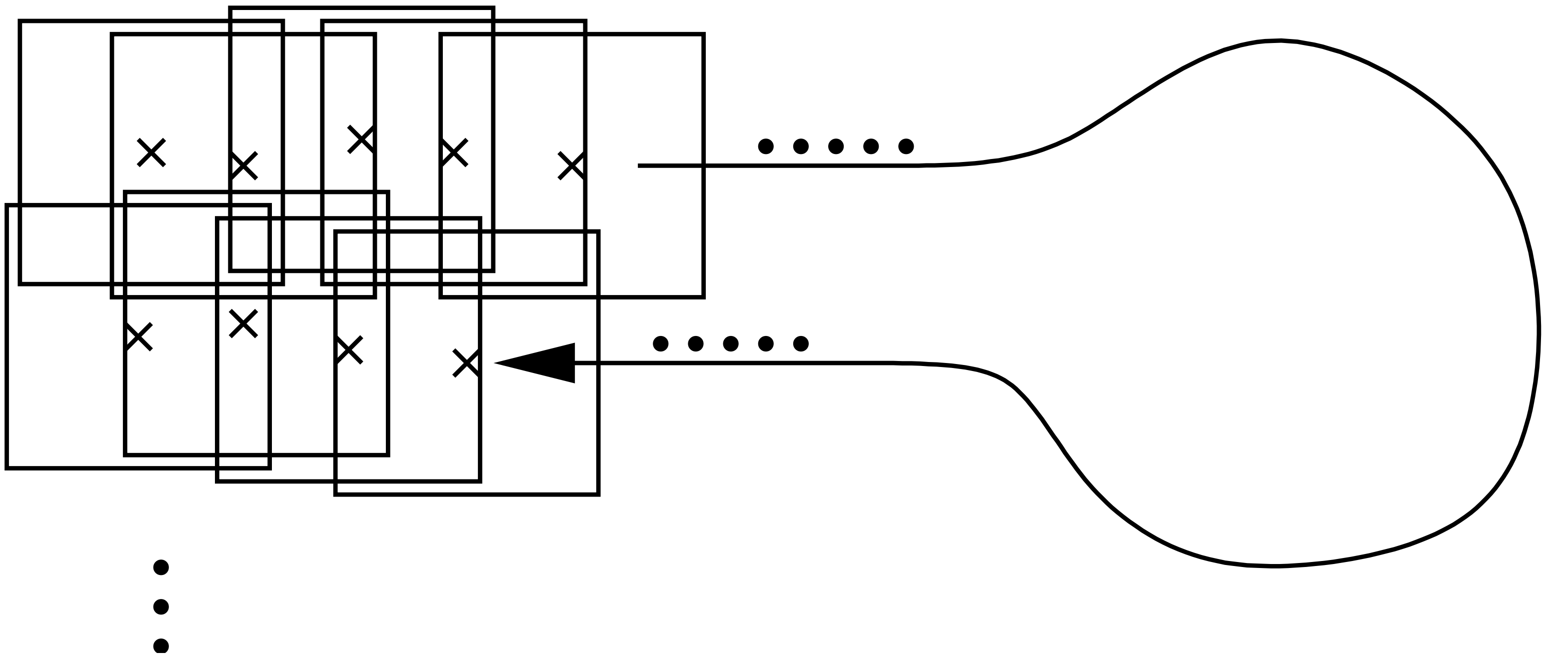


Image courtesy: Förstner

Typical Setup

Area to be mapped is covered in stripes

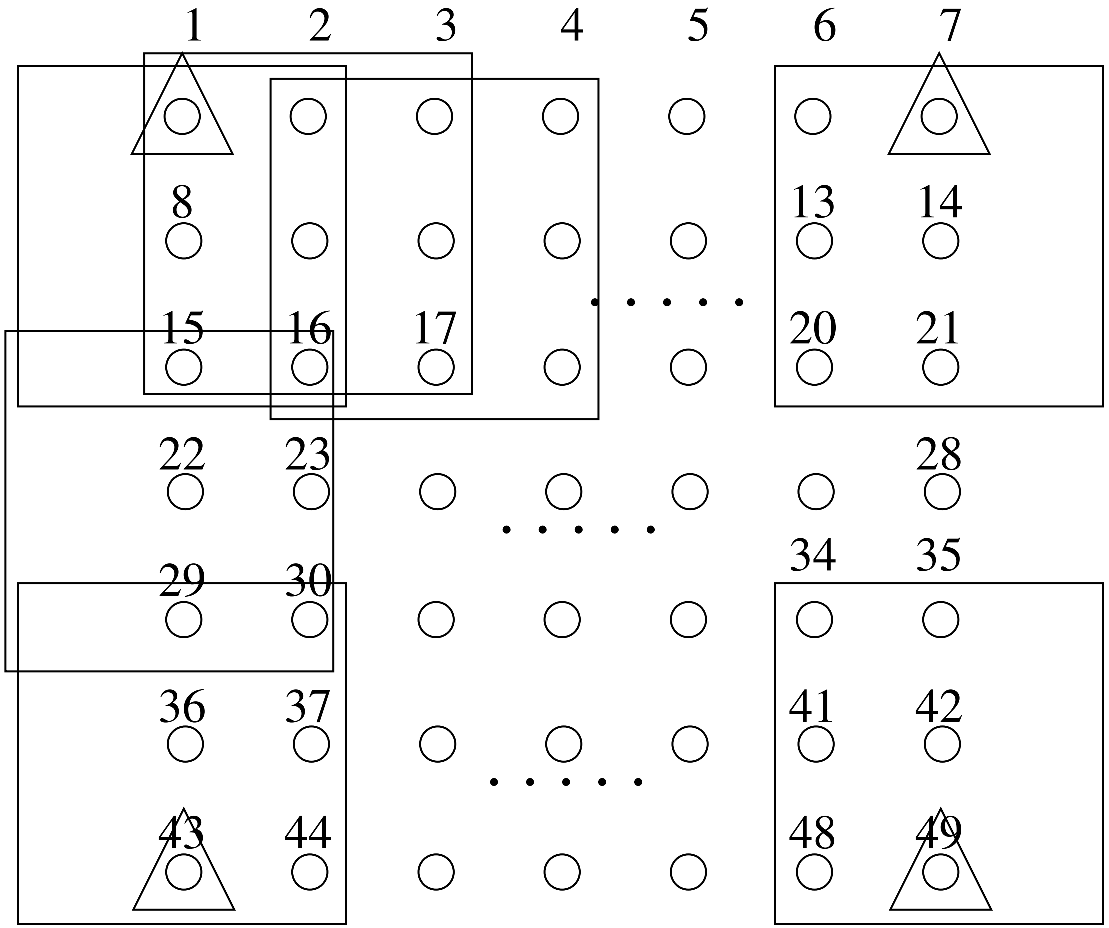
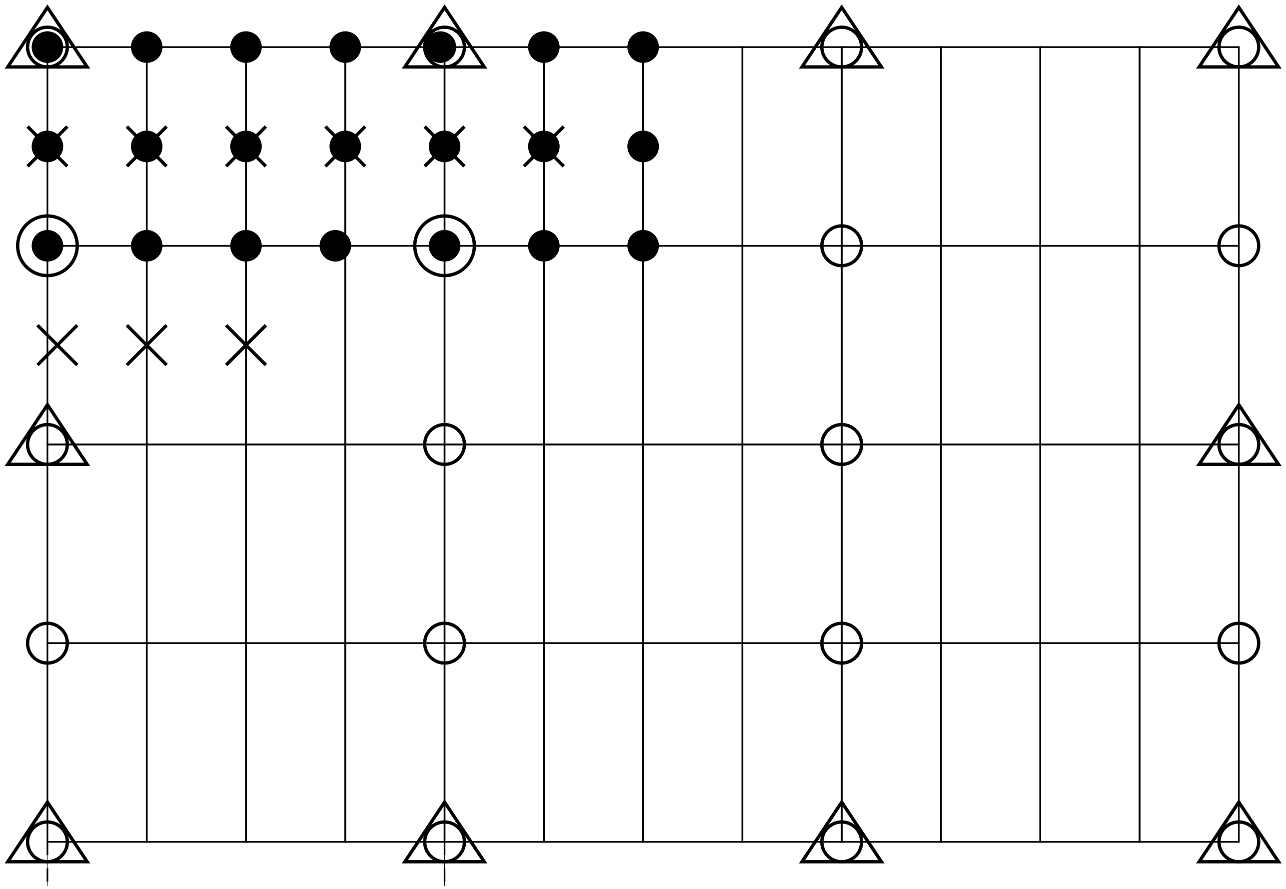


Image courtesy: Förstner

Resulting Geometry



△ Full C.P.

○ Height C.P.

• Tie points

× Projection centers

Image courtesy: Förstner

Resulting Geometry

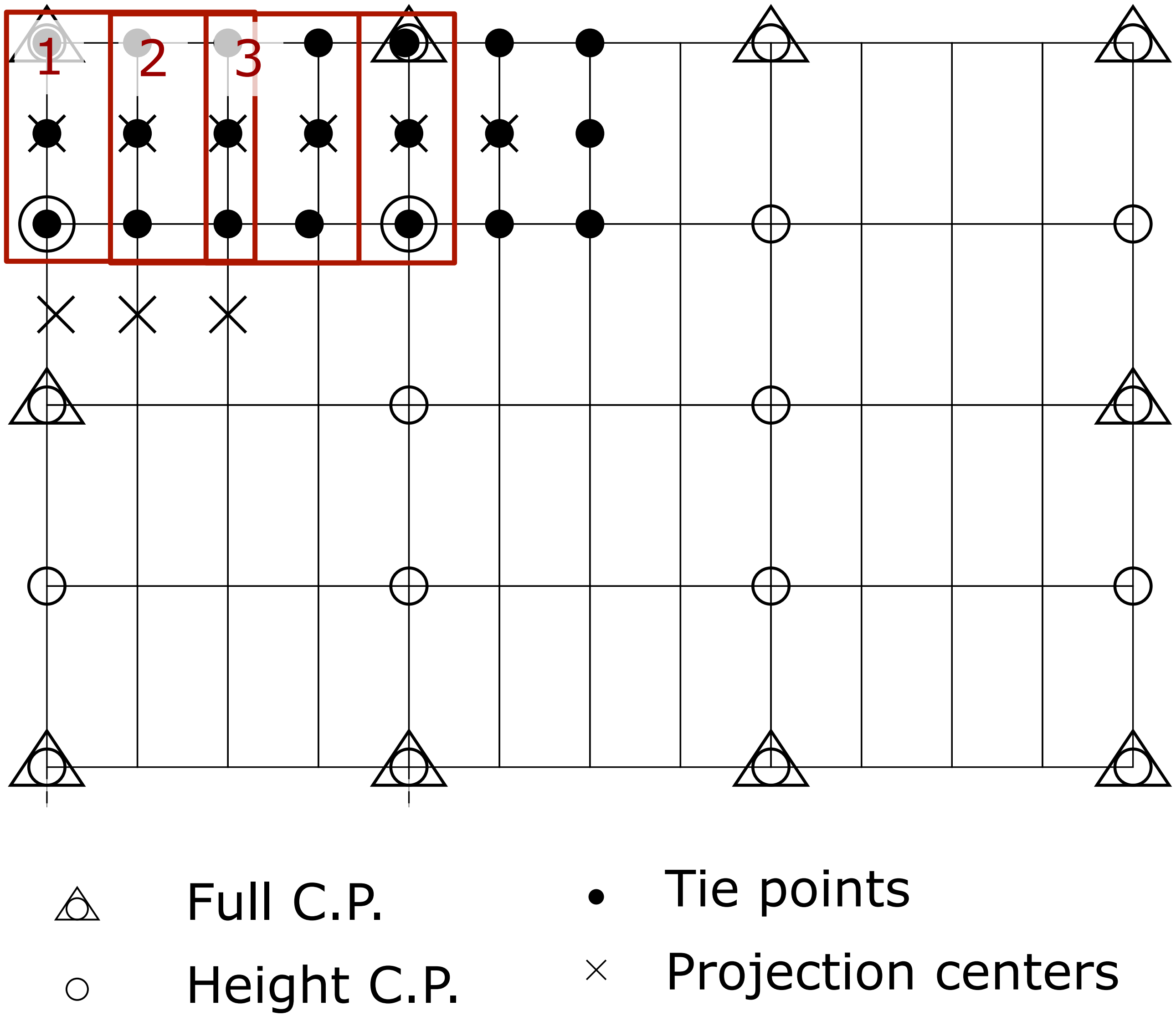
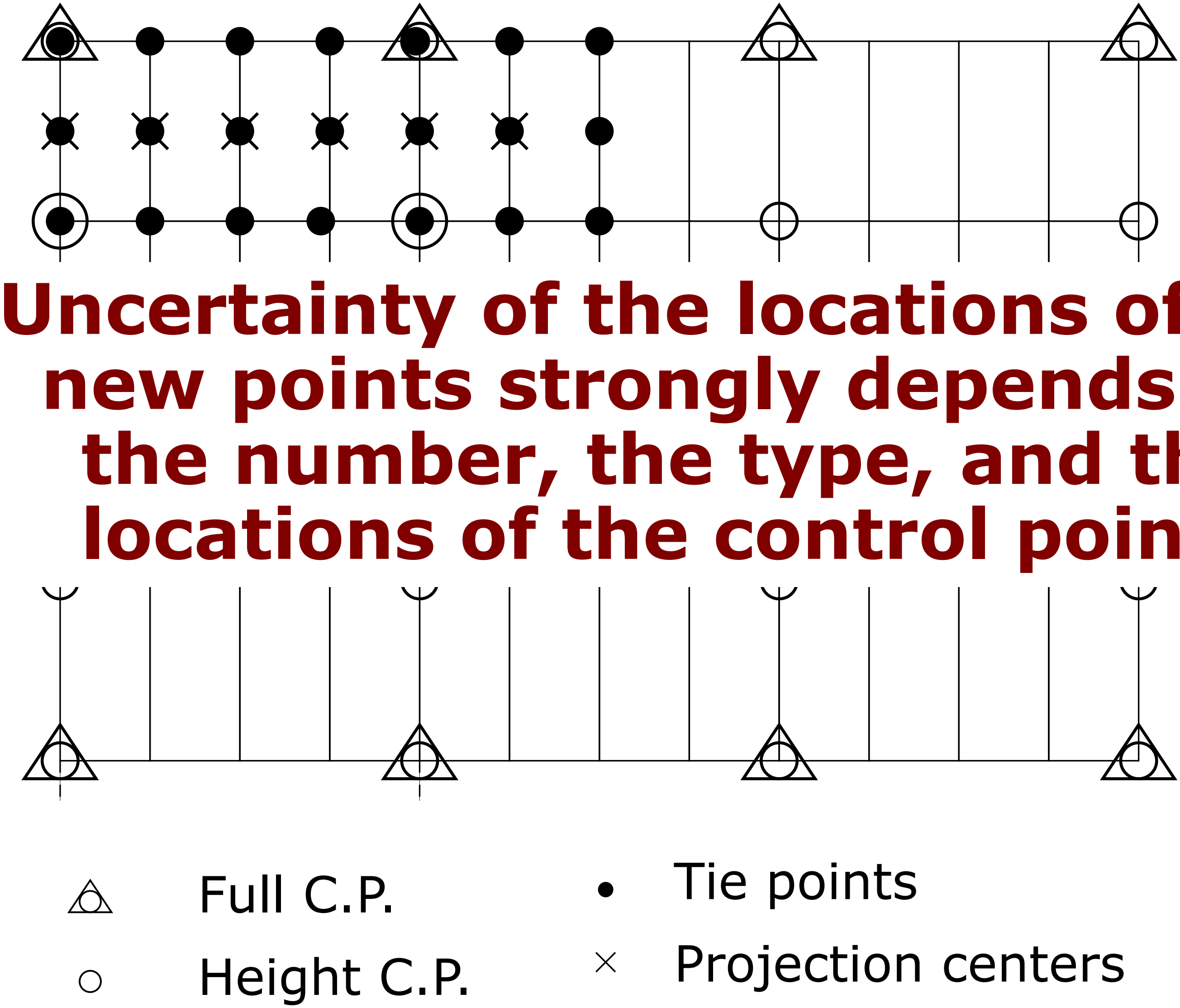


Image courtesy: Förstner

Resulting Geometry



Uncertainty of the locations of the new points strongly depends on the number, the type, and the locations of the control points

(D)GPS + IMU + Camera

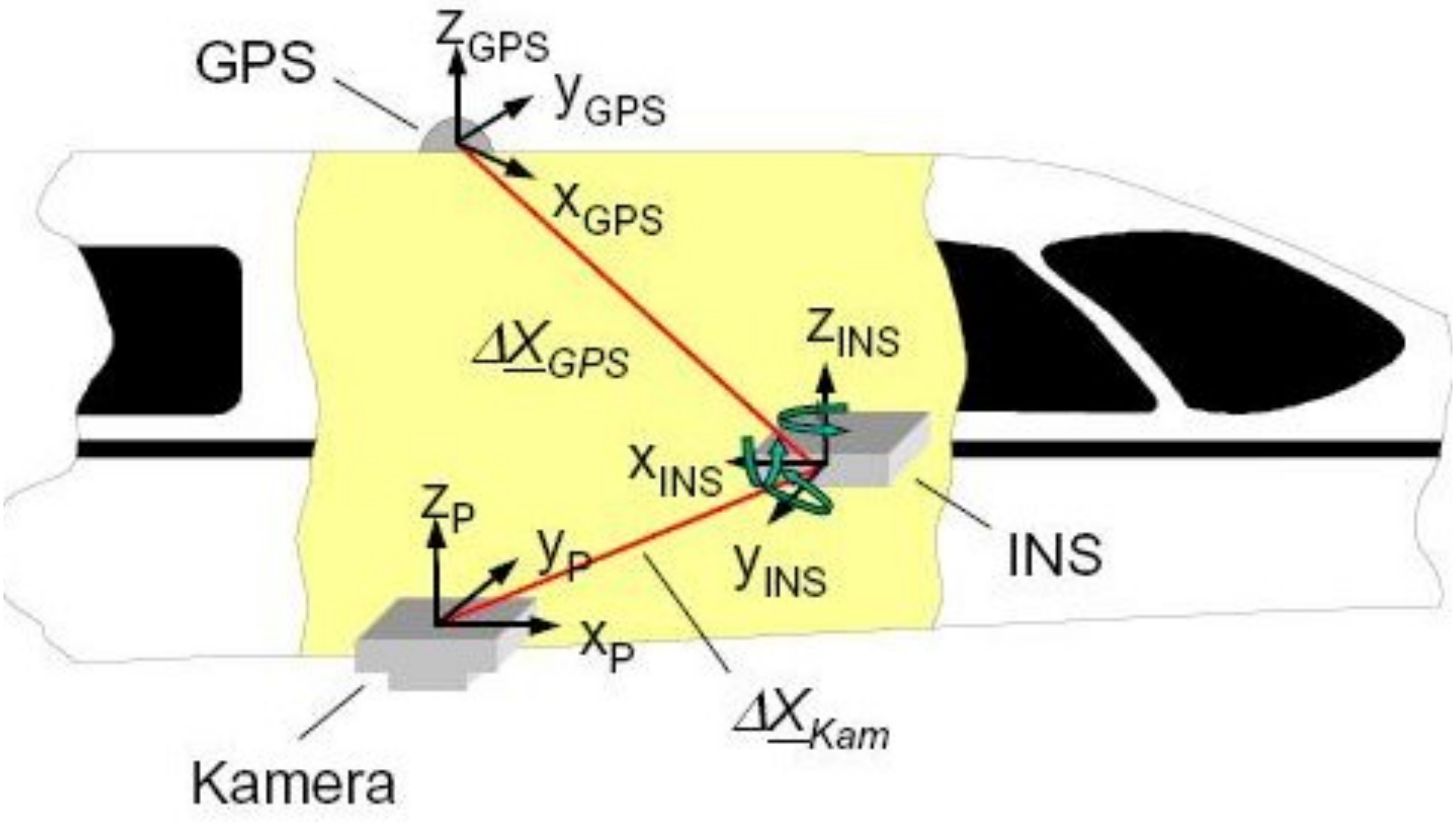


Image courtesy: Cramer

GPS/IMU Integration

- With D-GPS+&IMU, we obtain a 5-10cm accuracy for the camera pose
- “Control points” in the cameras
- Roll, pitch, yaw information from IMU
- Information is added as additional observations for the BA procedure

GPS/IMU Integration

- With D-GPS+&IMU, we obtain a 5-10cm accuracy for the camera pose
- “Control points” in the cameras
- Roll, pitch, yaw information from IMU
- Information is added as additional observations for the BA procedure

Do we need ground control points at all?

Few Control Points For…

- Dealing with calibration errors for example in the camera constant (temperature issues)
- Systematic errors
- Difference of coordinate system (GPS c.s. vs. targeted map c.s.)

How to Find CPs in the Images?

How to Find CPs in the Images?

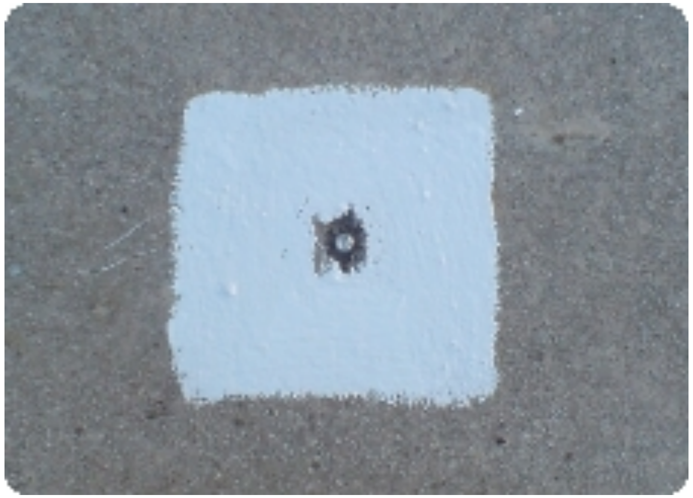


Image courtesy: StädteRegion Aachen

Aero/Arial Triangulation

- Photogrammetric measurements lead to precise estimates at large scales
- AT results in estimates with up to
  $\sigma_{XY} \approx 2.5\,cm$
- Large overlaps are suggested
- CP setup: CPs at the borders and Height-CP along the stripes
- Combination with GPS/IMU is standard today ("CPs in the projection centers")

Summary

- Bundle Adjustment = least squares solution to relative and absolute orientation considering uncertainties
- Minimizes the reprojection error
- Gold standard approach
- Statistically optimal solution under certain assumptions
- Solving the least squares problem requires numerical tricks (see Part 2)

Literature

- Förstner, Wrobel: “Photogrammetric Computer Vision”, Ch. 14.1-14.3
- Triggs, McLauchlan, Hartley, Fitzgibbon: “Bundle Adjustment — A Modern Synthesis”
- Hartley & Zisserman: “Mutiple View Geometry”

Slide Information

- The slides have been created by Cyrill Stachniss as part of the photogrammetry and robotics courses.
- I tried to acknowledge all people from whom I used images or videos. In case I made a mistake or missed someone, please let me know.
- The photogrammetry material heavily relies on the very well written lecture notes by Wolfgang Förstner and the Photogrammetric Computer Vision book by Förstner & Wrobel.
- Parts of the robotics material stems from the great Probabilistic Robotics book by Thrun, Burgard and Fox.
- If you are a university lecturer, feel free to use the course material. If you adapt the course material, please make sure that you keep the acknowledgements to others and please acknowledge me as well. To satisfy my own curiosity, please send me email notice if you use my slides.

Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de, 2014