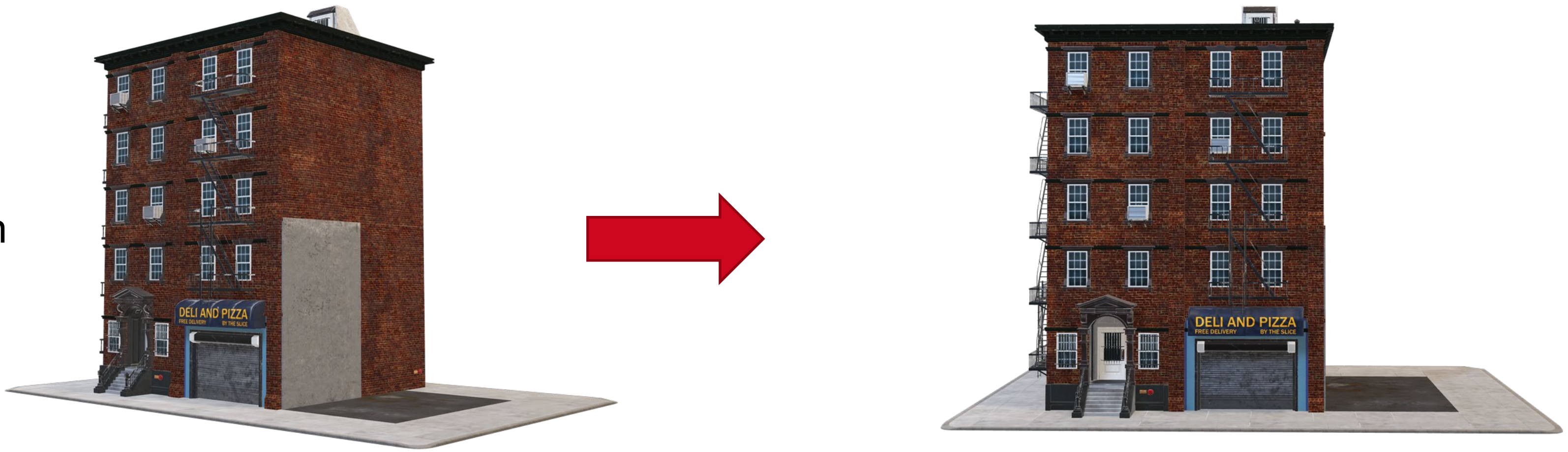
Automated Image Rectification of Perspective Distortions using Machine Learning

Oldenburger 3D-Tage

Aswin Lal, Elisa Beringmeier, Baris Özcan, Chongjie Kang, Steffen Marx, Jörg Blankenbach |
31.01.2024

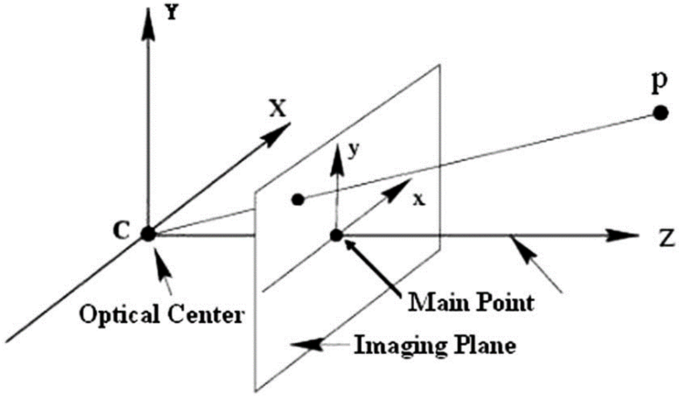
gia Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Contents



Background

- Central projection of camera: leads to perspective distortion
- For planar surfaces: evident from oblique views
- Parallel lines converge in image: vanishing lines/points



Central projection as a simplified pinhole camera model

Collinearity equations:

$$x - x_0 = -c \frac{R_{11}(X - X_0) + R_{21}(Y - Y_0) + R_{31}(Z - Z_0)}{R_{13}(X - X_0) + R_{23}(Y - Y_0) + R_{33}(Z - Z_0)}$$

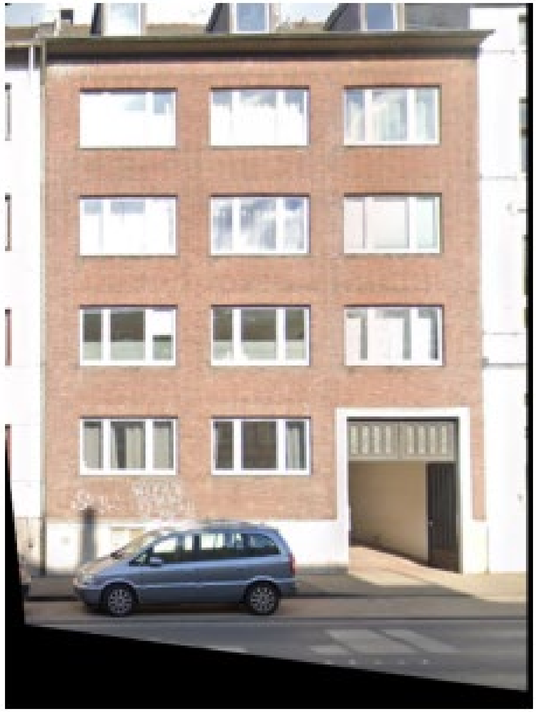
$$y - y_0 = -c \frac{R_{12}(X - X_0) + R_{22}(Y - Y_0) + R_{32}(Z - Z_0)}{R_{13}(X - X_0) + R_{23}(Y - Y_0) + R_{33}(Z - Z_0)}$$

Perspective distortion: Building façade
(Source: Google Street View)

Motivation

- Goals of image rectification:
  - better visual perception
  - simpler measurements within the image
  - uniform view of the object
- Fields of application:
  - Remote sensing
  - Cartography
  - Geography
  - Architecture
  - Construction


Distorted image


Rectified image

Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Introduction

Motivation

- Traditional photogrammetric methods:
  - based on planar projective transformation
  - manual or semi-automated
  - requires known (control) points
  - elaborate, time-consuming, prone to human-error, …
- Our objective:
  - Automation of rectification procedure:
    - using artificial intelligence
    - based on monocular images

$$X = \frac{A_1 \cdot x + A_2 \cdot y + A_3}{A_7 \cdot x + A_8 \cdot y + 1} \qquad Y = \frac{A_4 \cdot x + A_5 \cdot y + A_6}{A_7 \cdot x + A_8 \cdot y + 1}$$

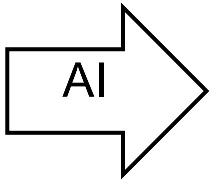
x, y = image coordinates
X, Y = object coordinates
A_i = 8 transformation parameters

Inverse formulas:

$$x = \frac{(A_5 - A_8 . A_6) \cdot X_P + (A_3 . A_8 - A_2) \cdot Y_P + (A_2 \cdot A_6 - A_3 \cdot A_5)}{(A_4 \cdot A_8 - A_5 \cdot A_7) \cdot X_P + (A_2 \cdot A_7 - A_1 \cdot A_8) \cdot Y_P + (A_1 \cdot A_5 - A_2 \cdot A_4)}$$

$$y = \frac{(A_6 \cdot A_7 - A_4) \cdot X_P + (A_1 - A_3 \cdot A_7) \cdot Y_P + (A_3 \cdot A_4 - A_1 \cdot A_6)}{(A_4 \cdot A_8 - A_5 \cdot A_7) \cdot X_P + (A_2 \cdot A_7 - A_1 \cdot A_8) \cdot Y_P + (A_1 \cdot A_5 - A_2 \cdot A_4)}$$


Distorted image

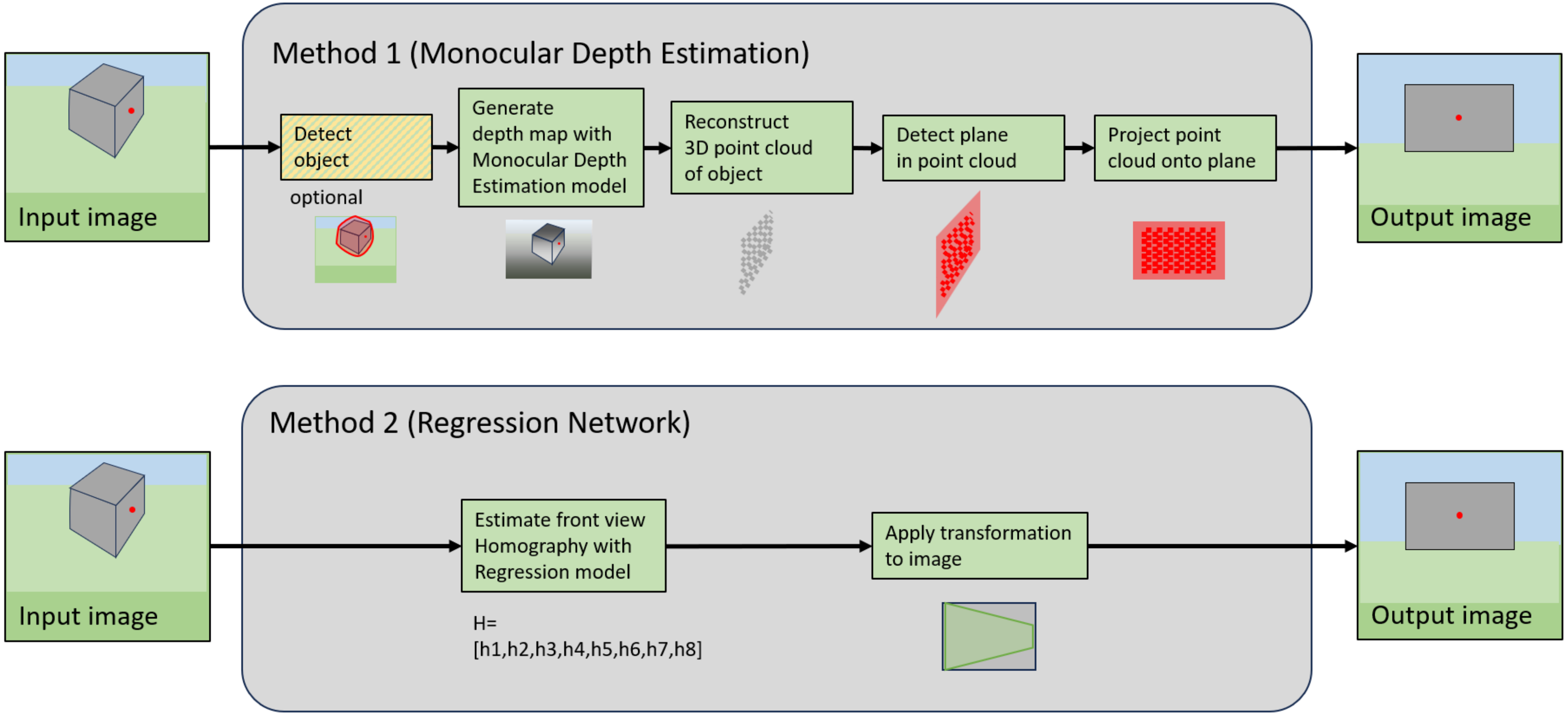



Rectified image

Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Methods

Overview

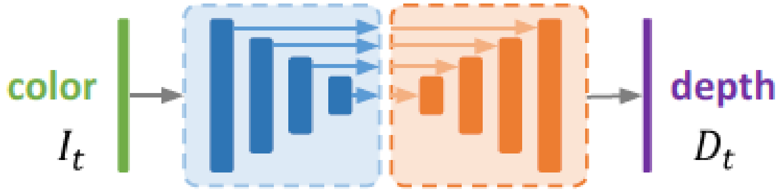


Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme

RWTH AACHEN UNIVERSITY

Monocular Depth Estimation

- Depth prediction for single images
  - Estimation of depth values for each pixel
  - Different models
    - Laina et al. 2016
    - Eigen et al. 2014
    - Godard et al. 2019

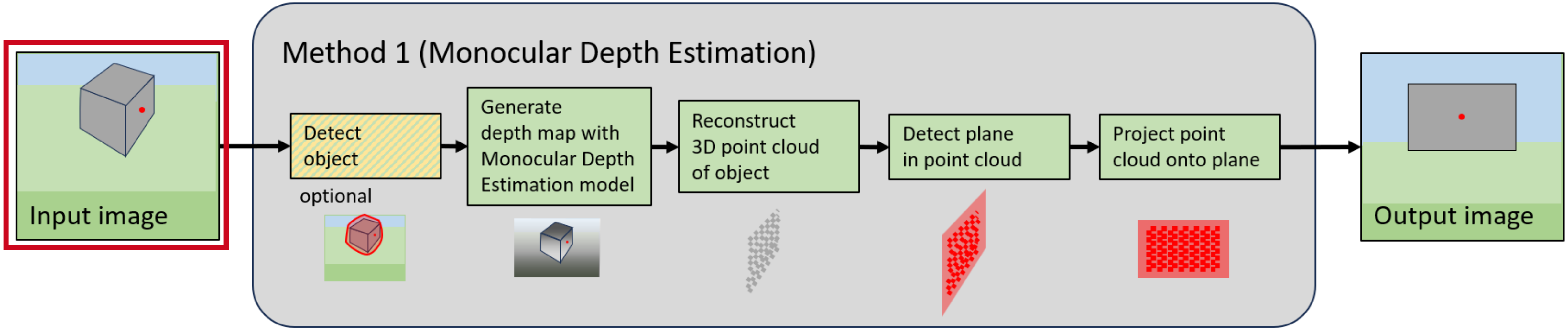


Depth Network: MonoDepth2 (Godard et al., 2019)

MonoDepth2 (Godard et al., 2019)

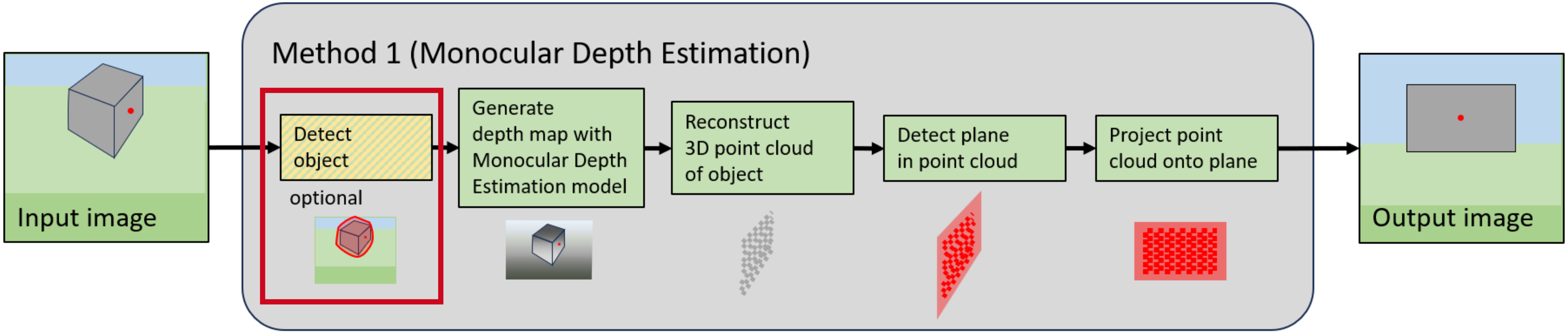
Method 1: Monocular Depth Estimation

Workflow



Method 1: Monocular Depth Estimation

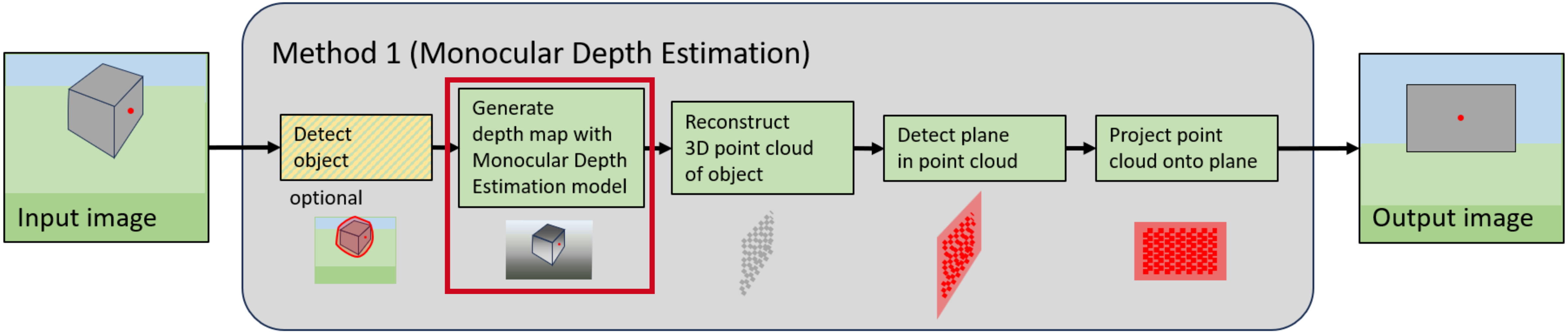
Workflow

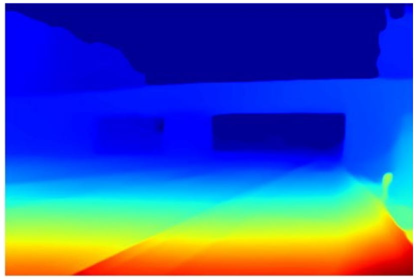


gia Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Method 1: Monocular Depth Estimation

Workflow

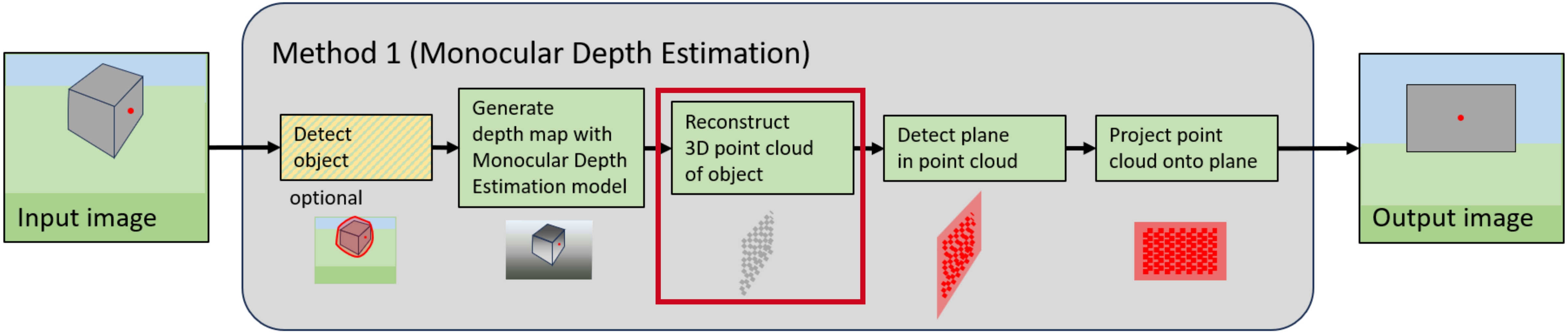


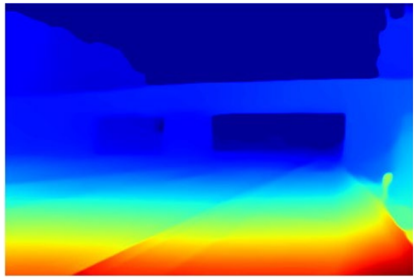


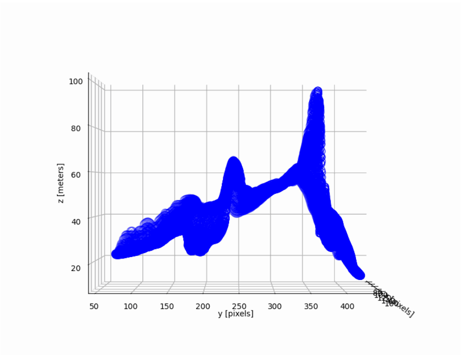
Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Method 1: Monocular Depth Estimation

Workflow

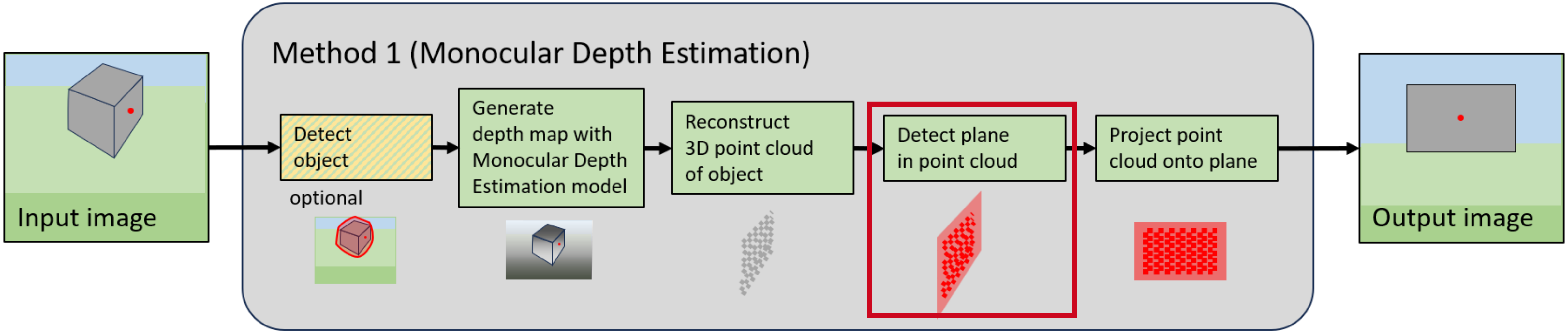


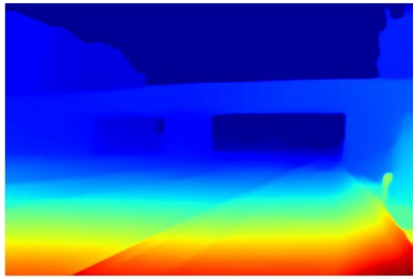


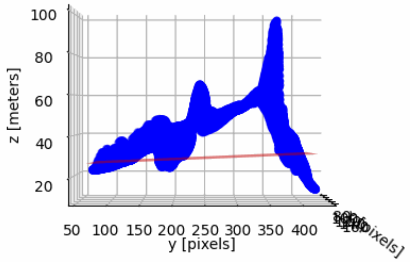


Method 1: Monocular Depth Estimation

Workflow

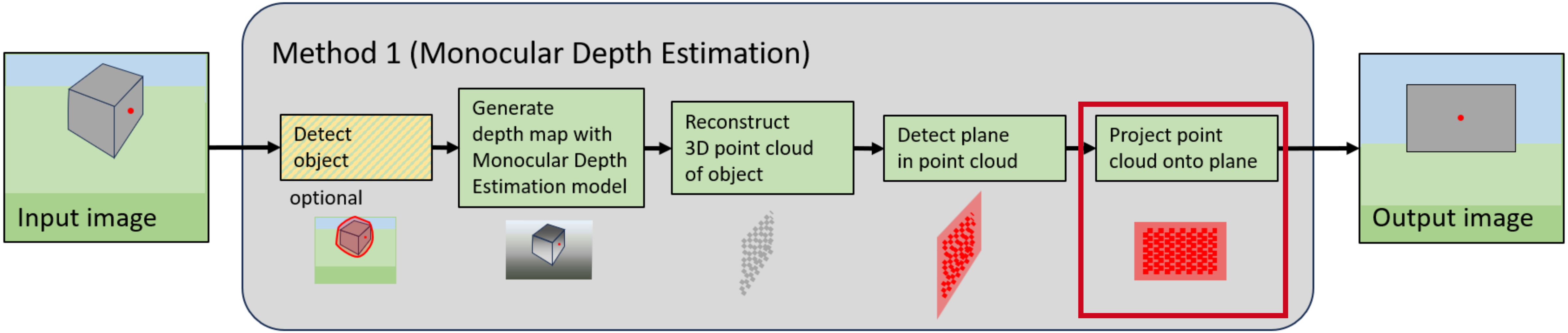


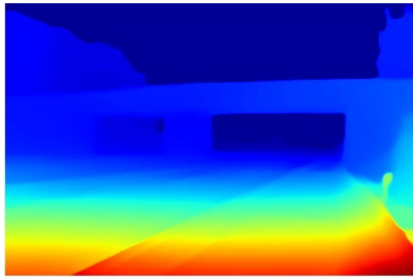


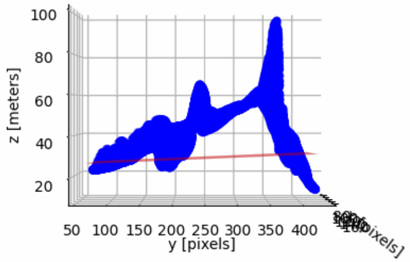


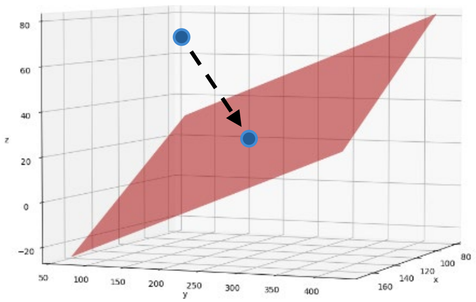
Method 1: Monocular Depth Estimation

Workflow



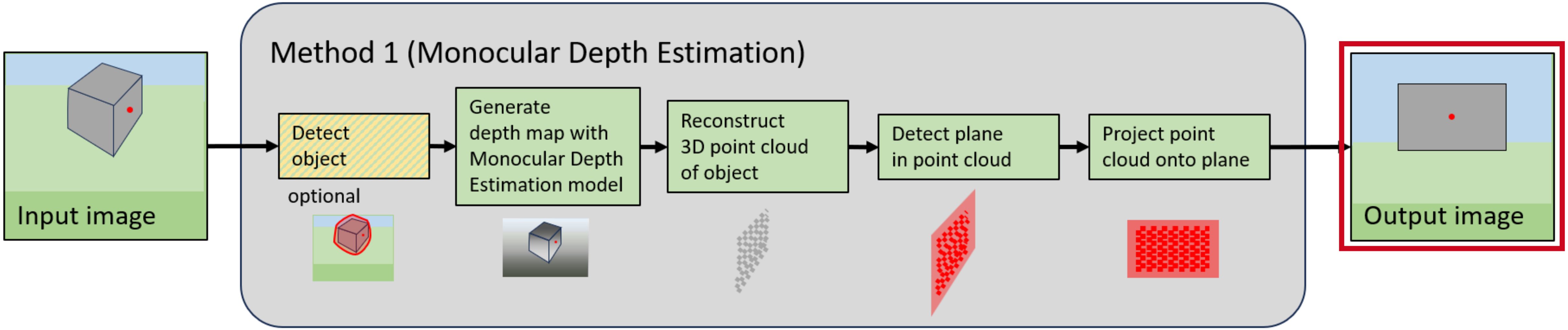


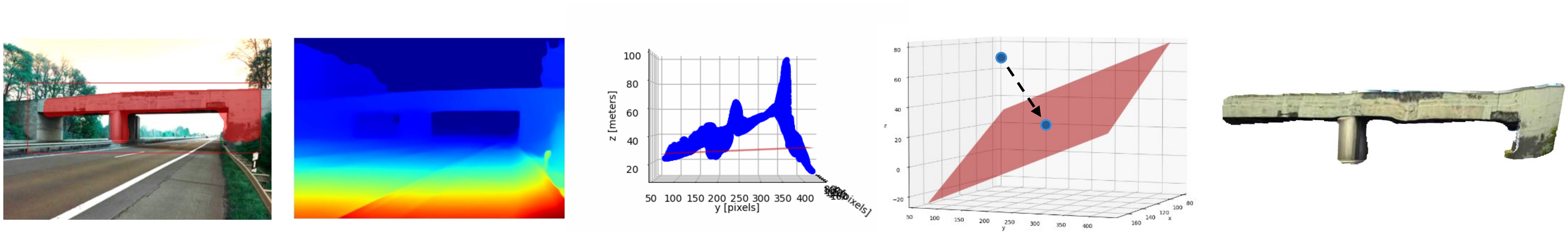




Method 1: Monocular Depth Estimation

Workflow





gia Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Method 2: Regression Network

Homography

- Relates two different views of the same object plane
- Given by a 3×3 matrix with 8 degrees of freedom:
  - $H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}$
    - h_i: rotation
    - h_i : translation
    - h_7 and h_8 : perspective transformation in the x and y directions
    - h_9 = 1 (fixed)
- Requires at least 4 point correspondences:
  - Feature Detectors (SIFT, SURF, …)

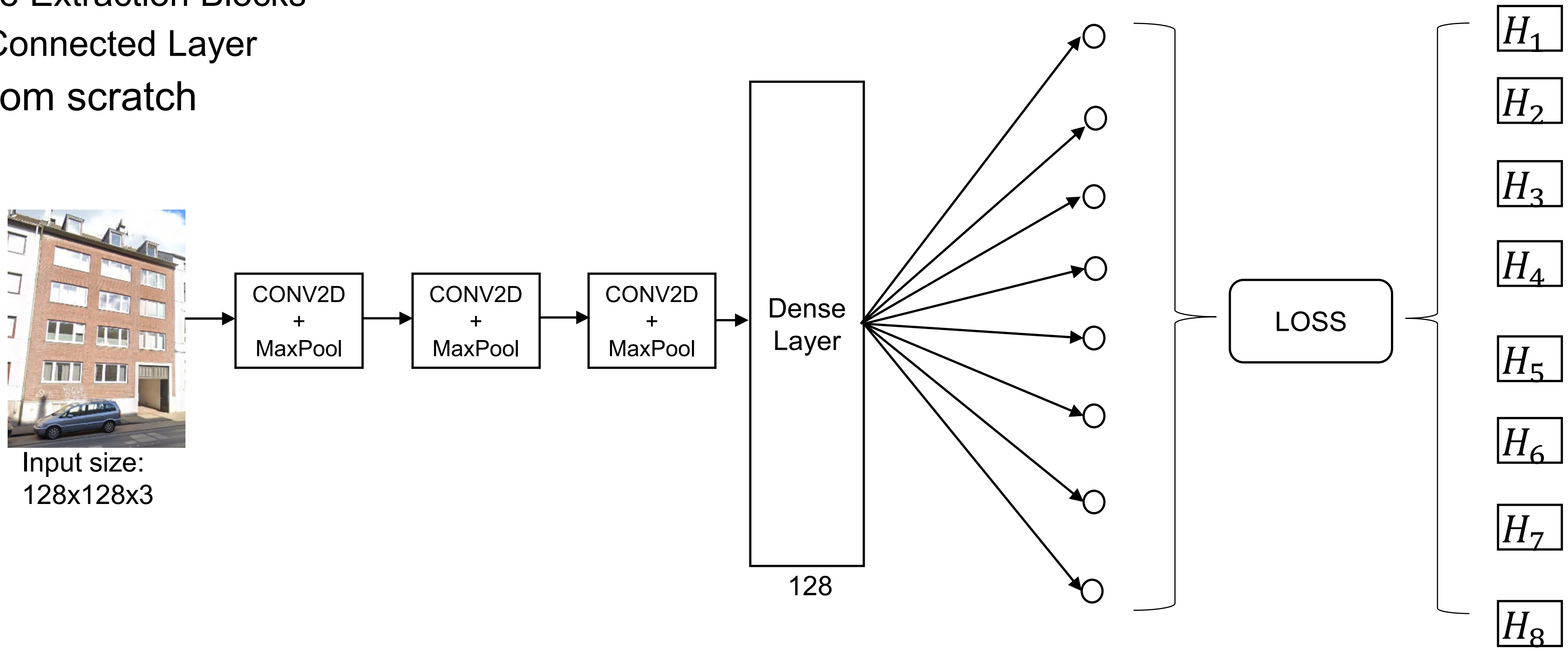
Matching of corresponding points using two images using SIFT

Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Method 2: Regression Network

Model

- Custom made CNN architecture
  - 3 Feature Extraction Blocks
  - 1 Fully-Connected Layer
- Trained from scratch

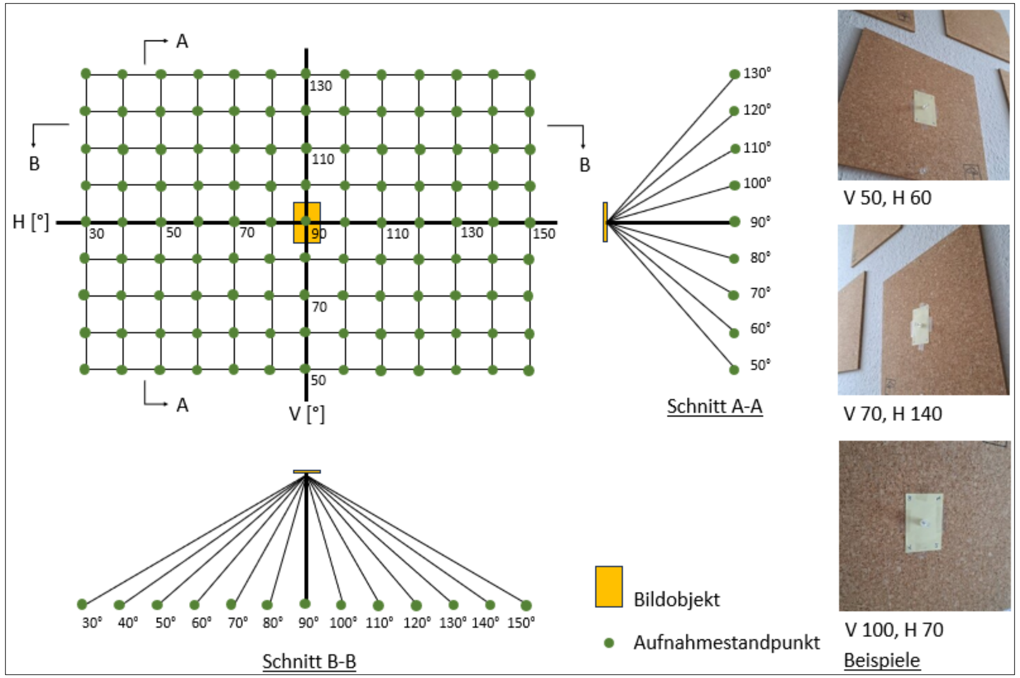


gia Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Method 2: Regression Network

Dataset Generation

- Training data
  - Input: Perspectively distorted images
  - Label: Homography parameters for rectified view



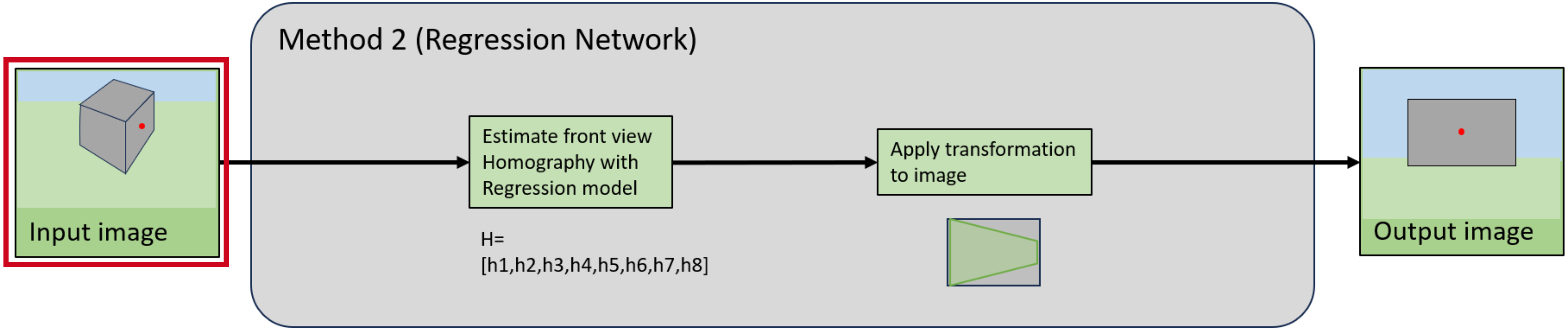
H=
[.7976e-01, -1.2159e-02,
7.4422e+00, -8.1867e-02,
8.4388e-01, 1.1808e+01,
-1.2194e-03, 9.6331e-05]

H=
[5.9155e-01, -1.8284e-03,
1.8212e+00, -2.0740e-01,
6.0980e-01, 2.8614e+01,
-3.03494e-03, 1.0305e-04]

H=
[1.7207e+00, 5.1040e-02,
4.2351e-01, 3.7136e-01,
1.0512e+00, -5.9037e-01,
5.8006e-03, 2.6404e-04]

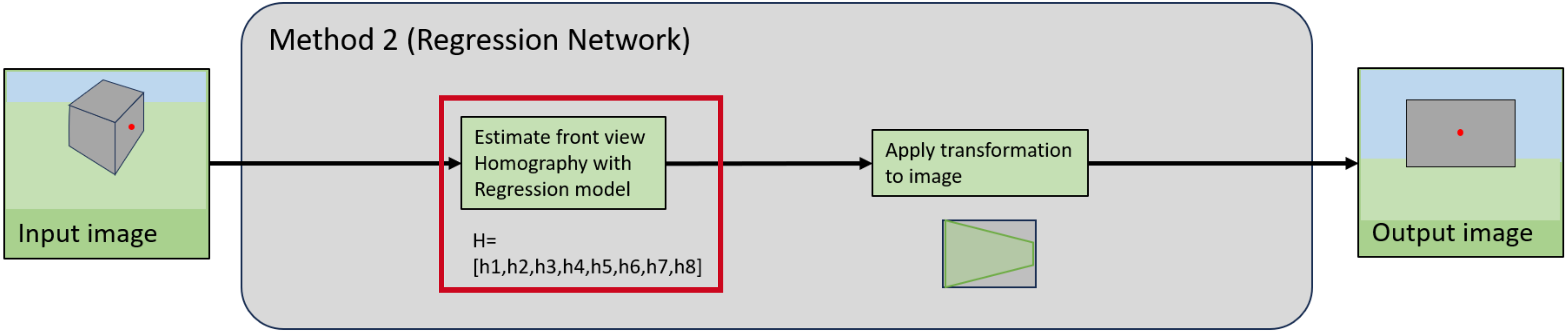
Method 2: Regression Network

Workflow



Method 2: Regression Network

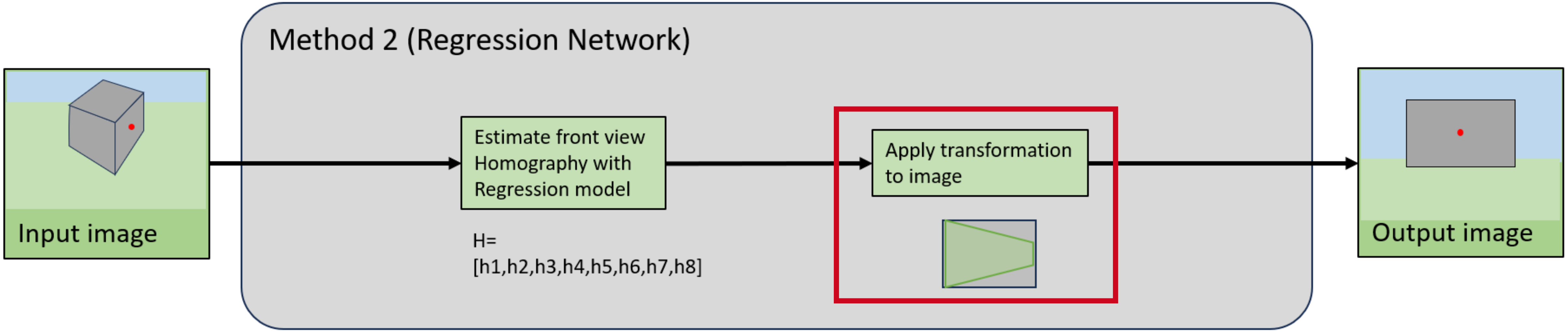
Workflow



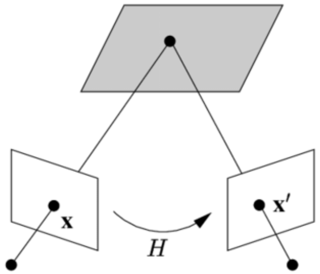
H=
[5.9155e-01, -1.8284e-03,
1.8212e+00, -2.0740e-01,
6.0980e-01, 2.8614e+01,
-3.03494e-03, 1.0305e-04]

Method 2: Regression Network

Workflow

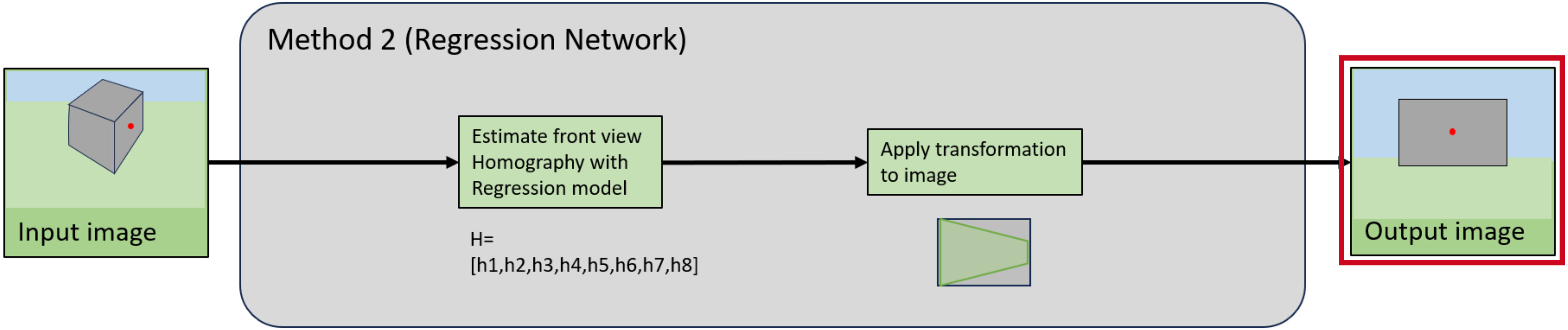


H=
[5.9155e-01, -1.8284e-03,
1.8212e+00, -2.0740e-01,
6.0980e-01, 2.8614e+01,
-3.03494e-03, 1.0305e-04]

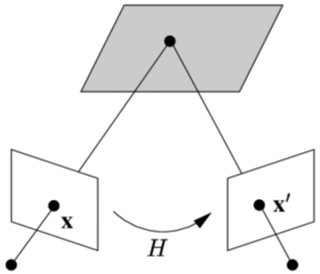


Method 2: Regression Network

Workflow

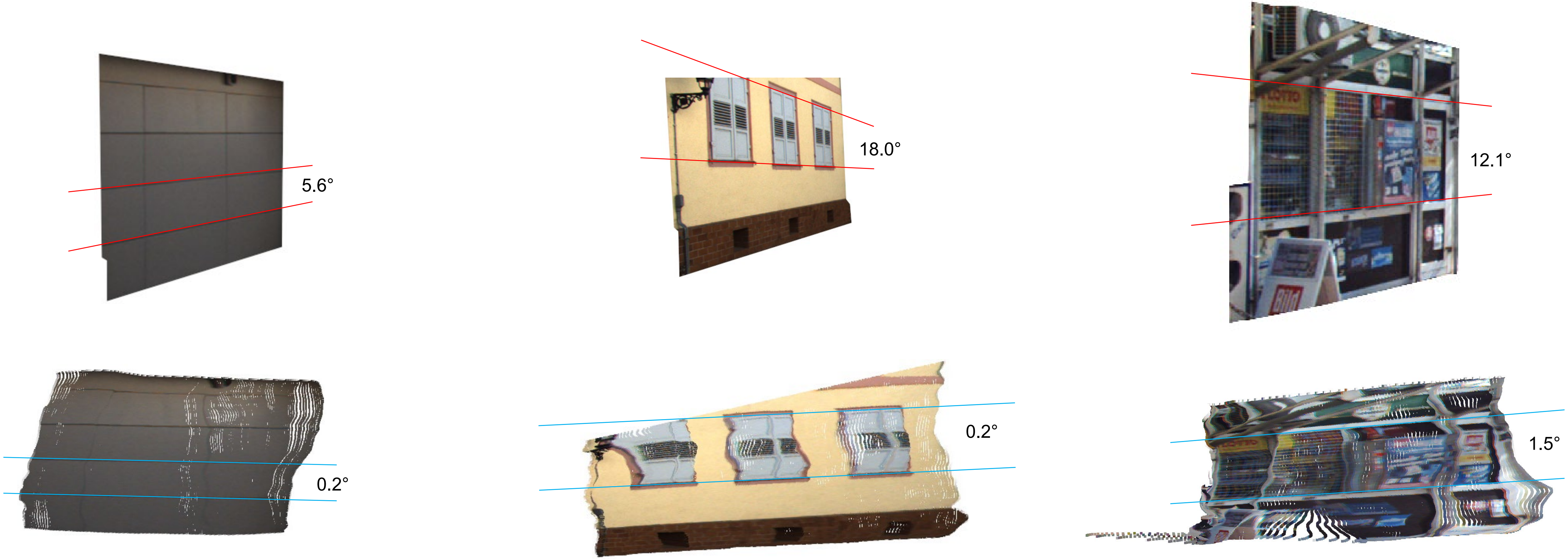


H=
[5.9155e-01, -1.8284e-03,
1.8212e+00, -2.0740e-01,
6.0980e-01, 2.8614e+01,
-3.03494e-03, 1.0305e-04]



Results

Method 1: Monocular Depth Estimation



Results

Method 2: Regression Network

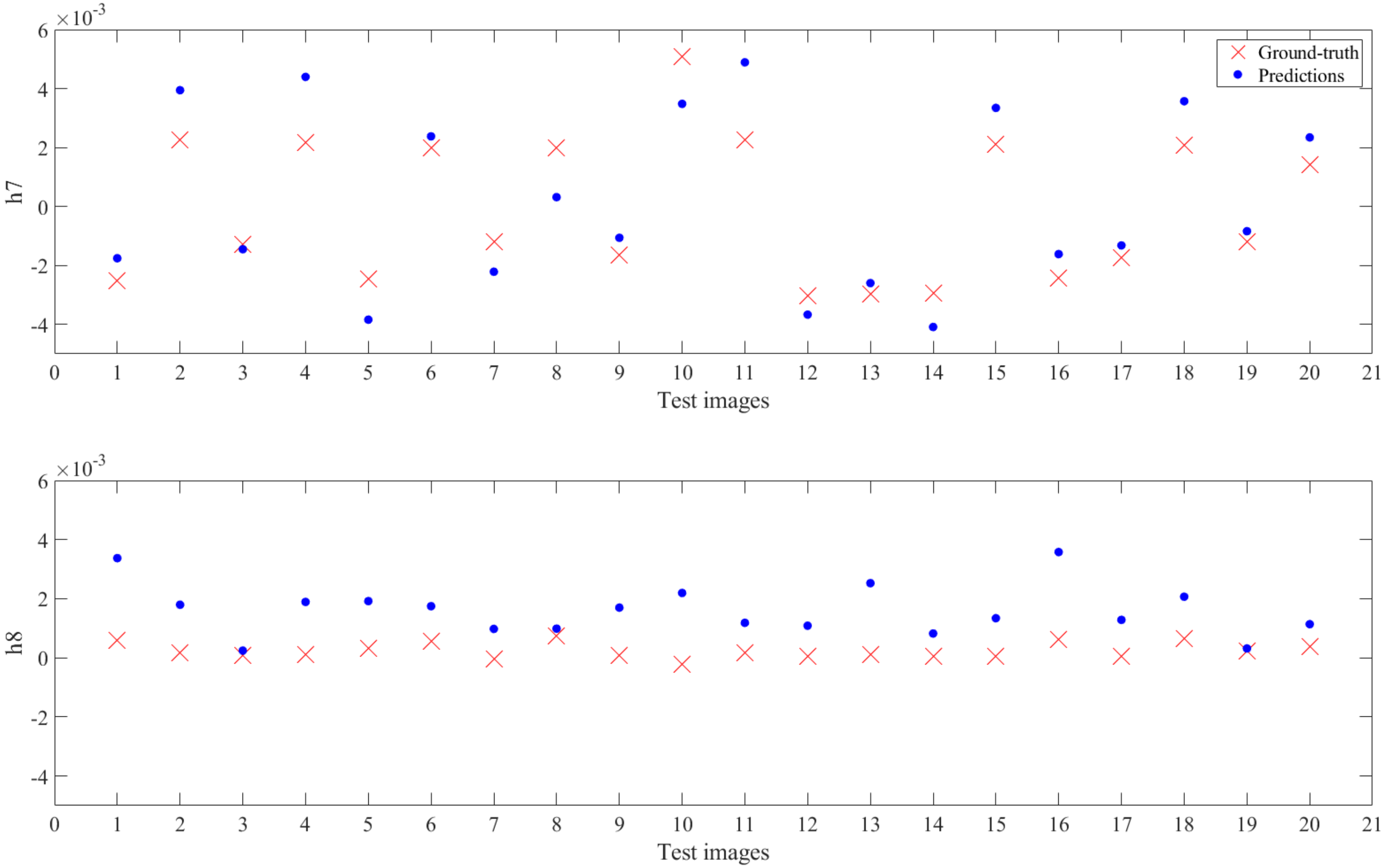
Results

Method 2: Regression Network

Results

Evaluation of predicted homography parameters

- Average RMSE = 0.2028
- Could be misleading!
- Parameter-wise evaluation necessary
- Deviations in h_7 and h_8: poor results

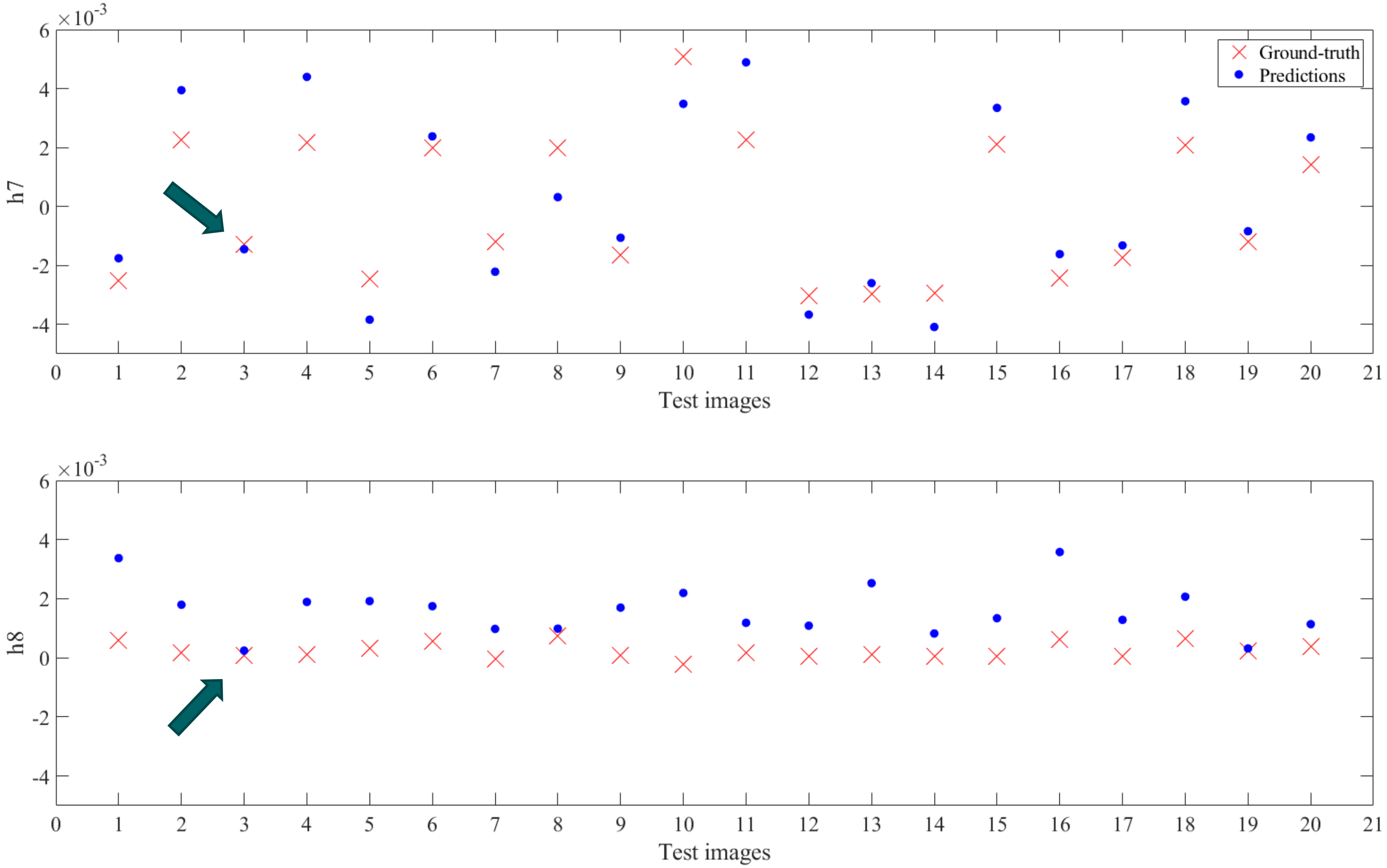


Comparison of h_7 and h_8 with the ground-truth

Results

Evaluation of predicted homography parameters

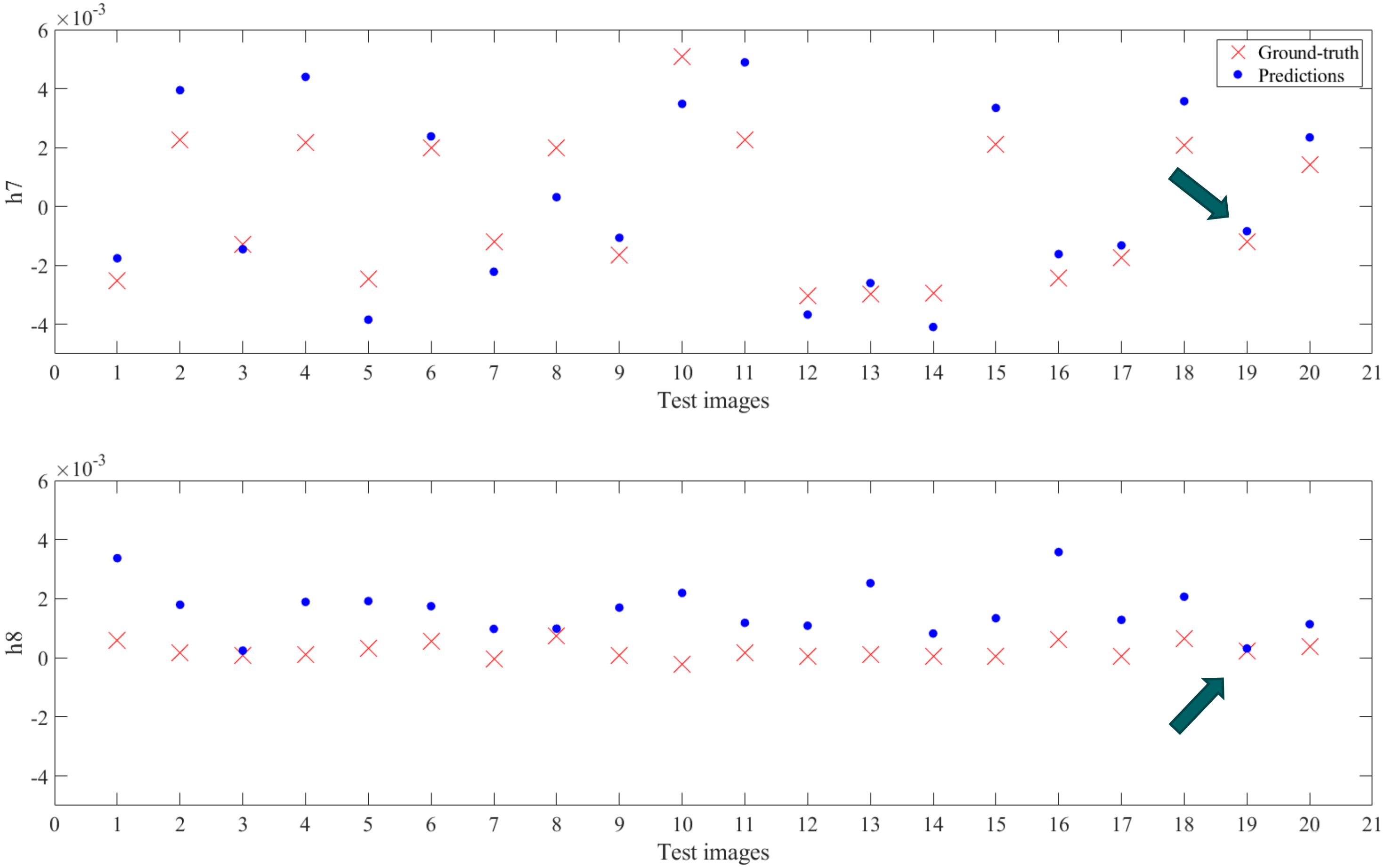
Test image 3: A good result



Comparison of h_7 and h_8 with the ground-truth

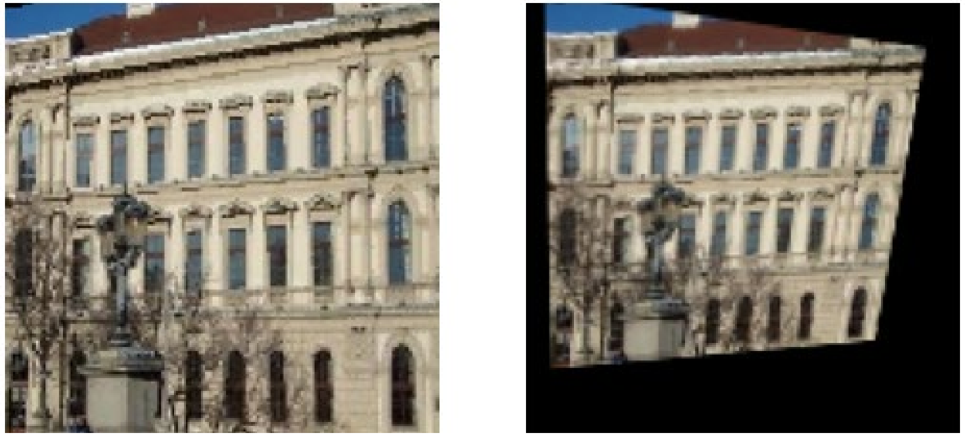
Evaluation of predicted homography parameters

Test image 19: A good result

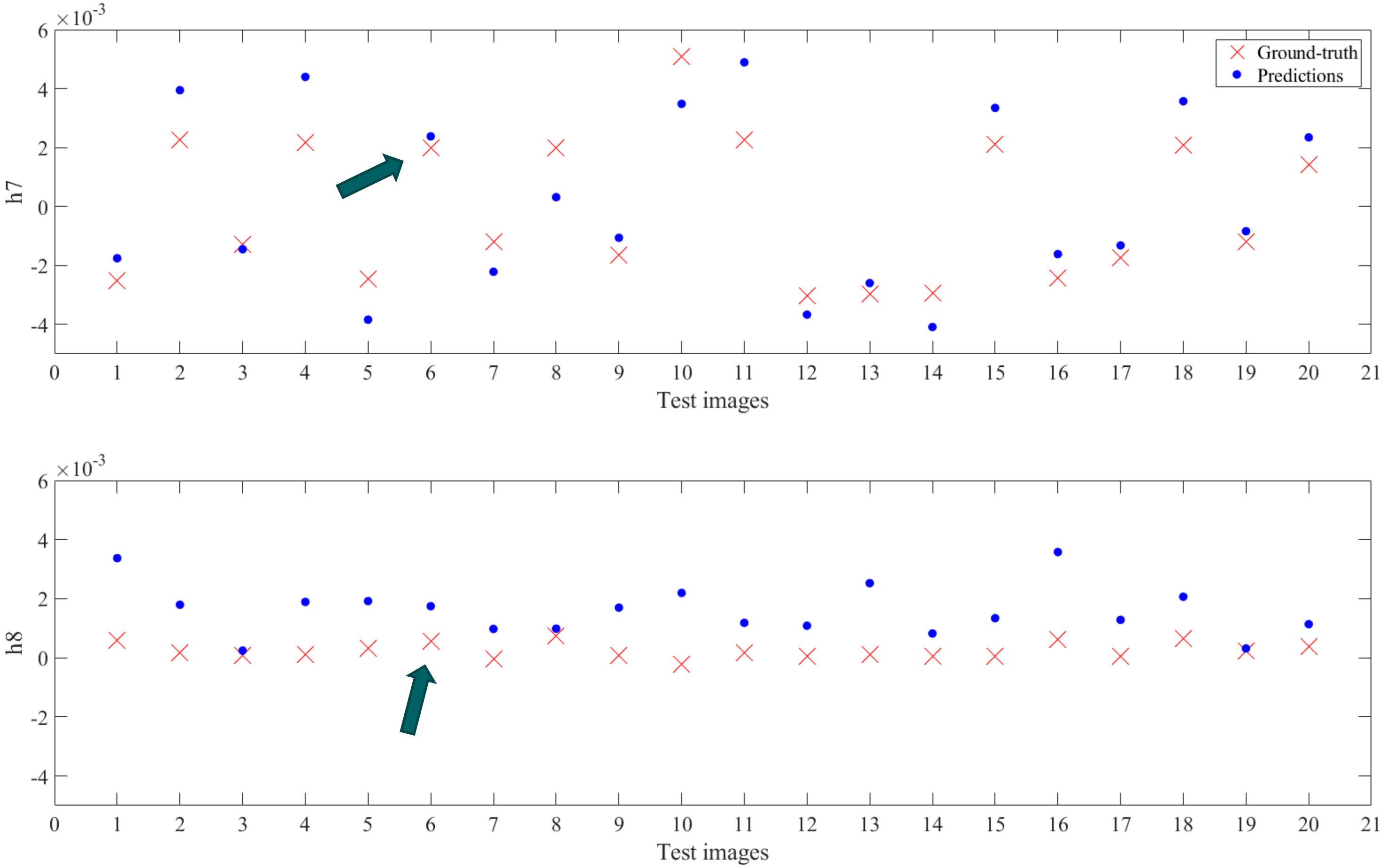


Comparison of h_7 and h_8 with the ground-truth

Evaluation of predicted homography parameters



Test image 6: A bad result

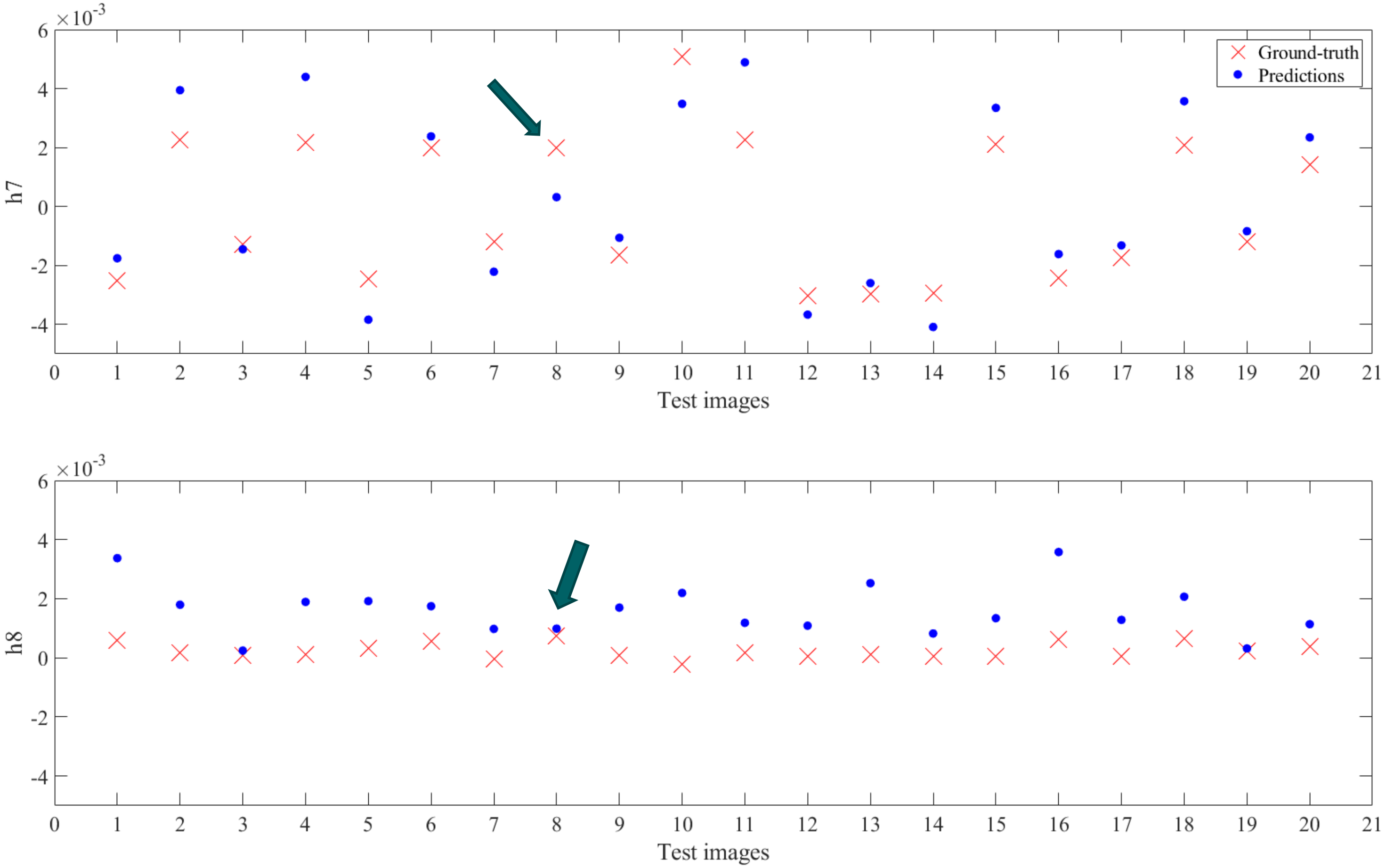


Comparison of h_7 and h_8 with the ground-truth

Results

Evaluation of predicted homography parameters

Test image 8: A bad result

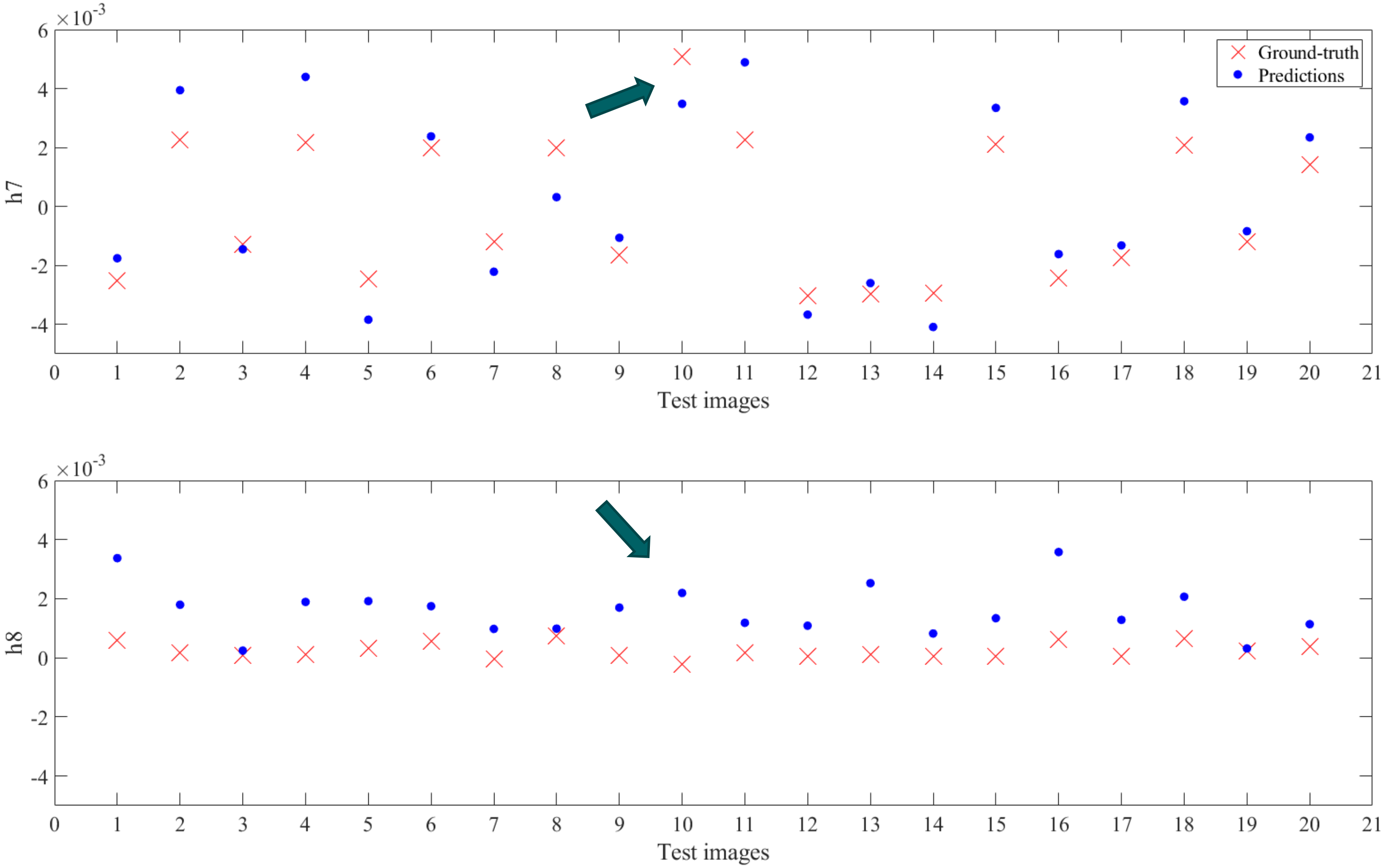


Comparison of h_7 and h_8 with the ground-truth

Results

Evaluation of predicted homography parameters

Test image 10: A bad result



Comparison of h_7 and h_8 with the ground-truth

Discussion

Assessments

- Method 1 (Monocular Depth Estimation)
  - Qualitative results:
    - show curvy lines (especially vertical lines)
    - heavily depend on quality of predicted point clouds
  - Quantitative results
    - quite good (< 2°)
- Method 2 (Regression Network)
  - Qualitative results:
    - vary largely (very good to bad)
    - rectification goes in the right direction
  - Quantitative results:
    - the accuracy of h_7 and h_8 are crucial
    - even small deviations can lead to poor results

Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Summary and Outlook

- Two workflows for perspective rectification using Machine Learning:
  - based on Monocular Depth Estimation
  - based on a Regression Network
- Promising results for both workflows
  - necessary to improve methodologies
- Future Work:
  - Improved Monocular Depth Estimation methods
    - Use more sophisticated models
    - Conduct Fine-Tuning
  - Constrained Regression
    - Incorporate constraints to the predicted Homography parameters

Work in progress…
Stay tuned!

Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme
RWTH AACHEN UNIVERSITY

Vielen Dank für Ihre Aufmerksamkeit

References

- Tyleček, R., & Šára, R. (2013). Spatial Pattern Templates for Recognition of Objects with Regular Structure. In Proc. GCPR. Saarbrucken, Germany.
- Map of Aachen, Germany. Google Maps (2019)[8]
- Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J. (2019). "Digging into Self-Supervised Monocular Depth Prediction." In The International Conference on Computer Vision (ICCV), October 2019.
- Geiger, A., Lenz, P., & Urtasun, R. (2012). "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite." In Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
- Luhmann, Thomas et al. "Close-Range Photogrammetry and 3D Imaging." (2013).
- Abdulla, W. (2017). "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow." GitHub repository. [Online]. Available: https://github.com/matterport/Mask_RCNN
- Title picture: bmp city: Accessed on 26.01.2024, Street Perspective Drawing - bmp-city

Results

Method 1: Monocular Depth Estimation