3D Rekonstruktion aus Luftbildern

Balthasar Teuscher

Inhalt

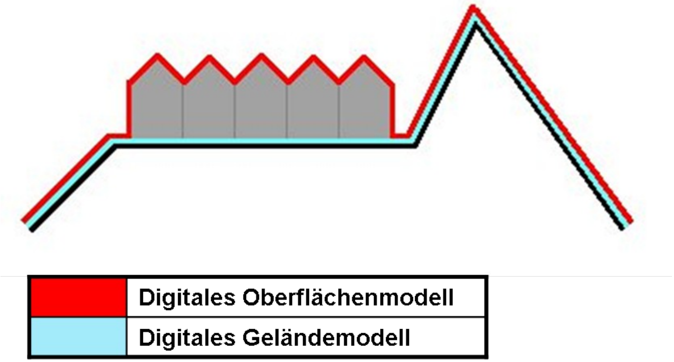
- Ausgangslage & Problemstellung
- Literature & State of the Art
- Datengrundlage & Datenaufbereitung
- Modelling & Netzwerkarchitektur
- Ergebnisse & Evaluation

Ausgangslage

- 3D Trend
  - Autonomes Navigieren, SLAM, AR, VR, GIS, BIM
  - Simulation, Gaming (Ray Tracing)
- Grundlage
  - Increasing computational capabilities (GPU)
  - Neuartige Algorythmen (Machine Learning)

Problemstellung

- 3D Rekonstruktion
  - Erhöhen der Dimensionalität
  - Digitales Oberflächenmodel (DOM)
- Traditioneller Workflow
  - Kamera Orientierung
  - Triangulation / stereo matching
    - Semiglobal matching (SGM) / Patch matching using CNN



State of the Art

- Robust Vision Challenge
  - stereo, multi-view stereo (MVS), optical flow, single image depth prediction, semantic segmentation and instance segmentation
- Fokus auf Robotik/SLAM, Autonomes Navigiern
- Datensätze: Street level view / synthetisch

Literatur

- Revisiting Single Image Depth Estimation: Toward Higher Resolution Maps with Accurate Object Boundaries (Hu, 2018)
- Indoor scene

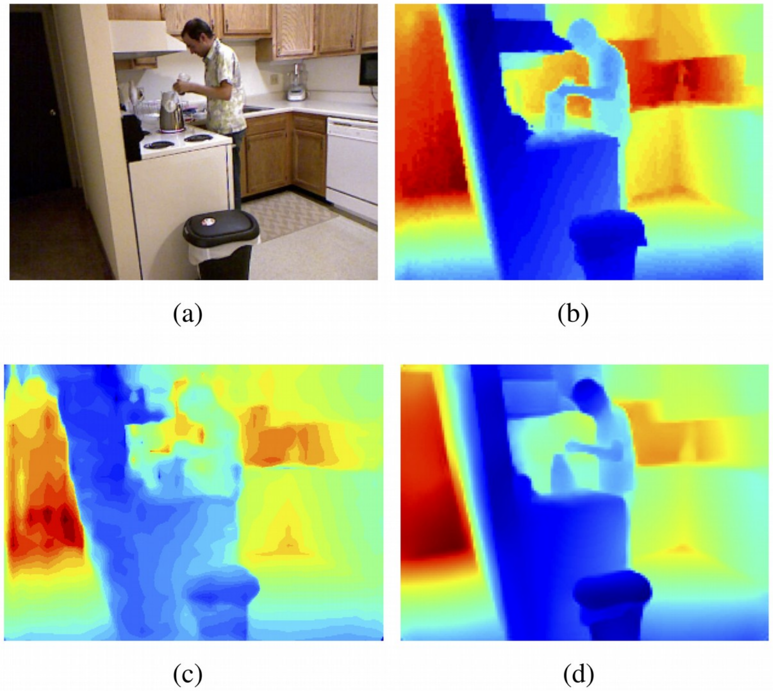


Figure 1. Example comparison of estimated depth maps; (a) RGB input, (b) ground truth depth, (c) the current state-of-the-art [10], and (d) our method.

Literatur

- Large-Scale Semantic 3D Reconstruction: An Adaptive Multi-resolution Model for Multi-class Volumetric Labeling (Blaha, 2016)
- Oblique imagery

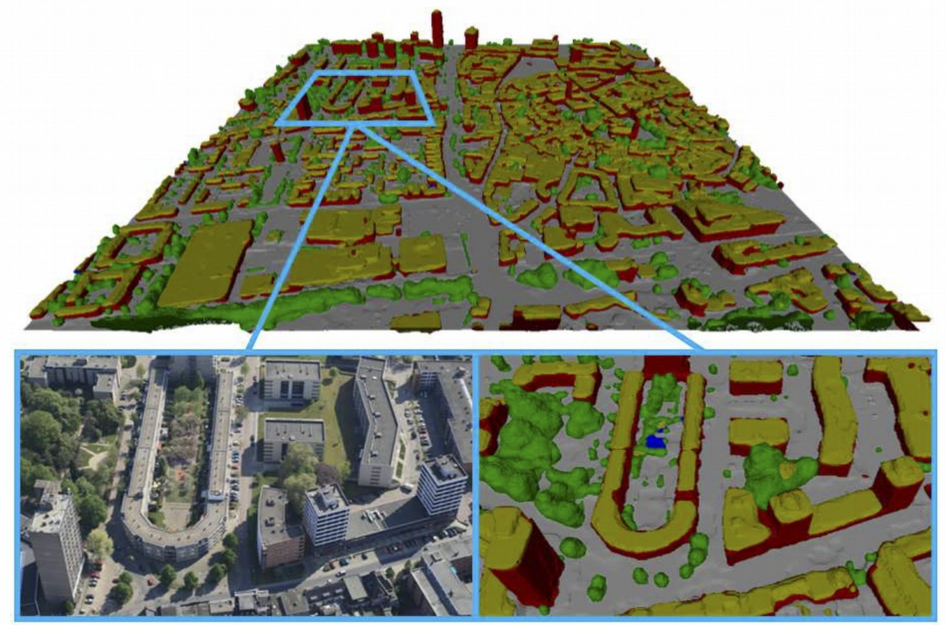


Figure 1: Semantic 3D model of the city of Enschede generated with the proposed adaptive multi-resolution approach.

Literatur

- Learning Shape Priors for Single-View 3D Completion and Reconstruction (Wu & Zhang, 2018)
- Objects

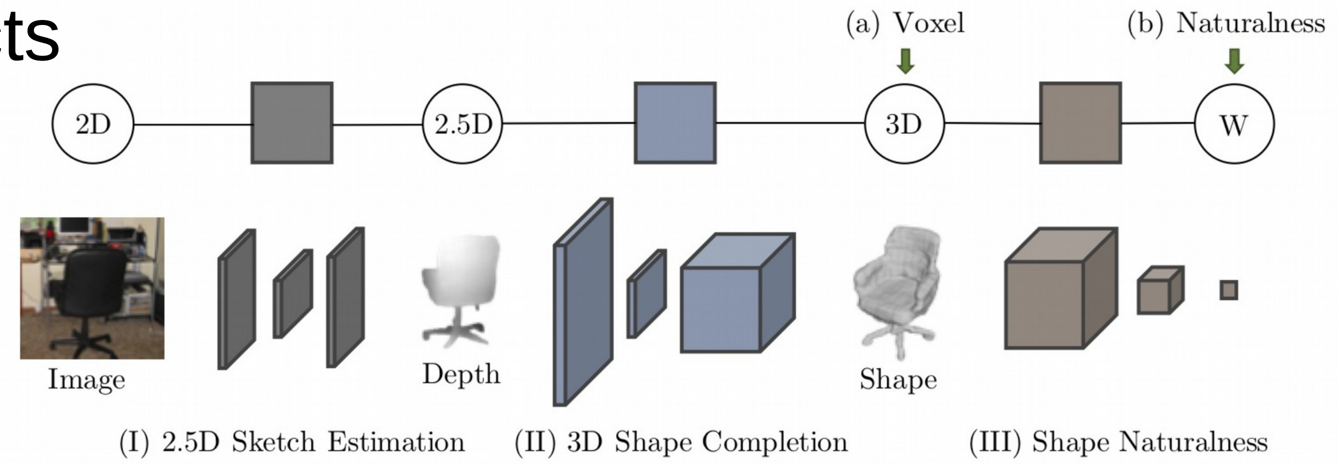


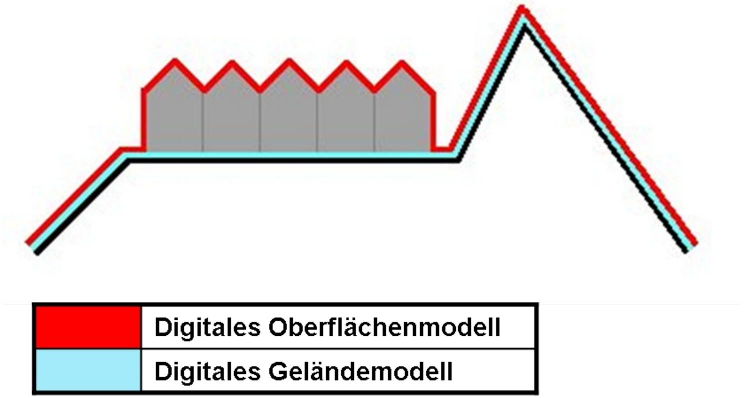
Figure 1: Our model for 3D shape reconstruction

Datengrundlage

- Luftbilder
  - Rasterdaten
  - RGB, 10 cm Auflösung, 34 GB
- Blockmodell
  - Vektordaten
  - TIN (LoD 1), 3 Mio Triangles
  - Dachmodel (LoD 3), 0.5 Mio Polygons

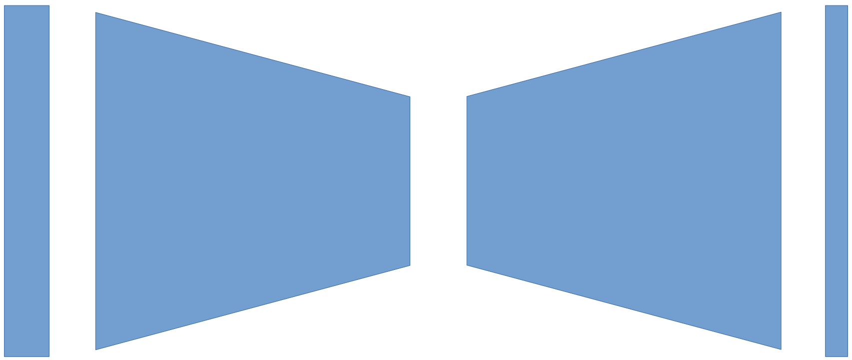
Datenaufbereitung

- Extrahieren einer “height map”
  - Differenz zwischen Dach und Boden
  - Ground truth um den loss zu berechnen
- Interpolation
  - Linear, nearest neighbour
- Tiling
  - ca. 8500 tiles, 1024x1024 pixel



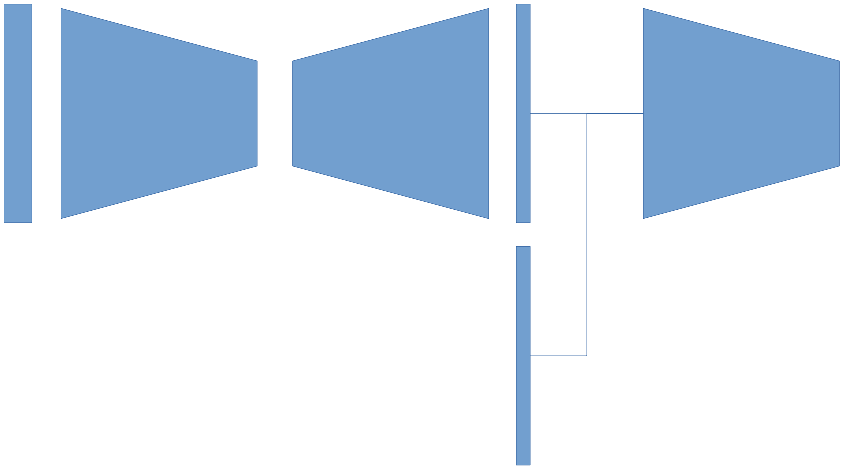
Modeling

- Autoencoder (CNN)
  - Input: RGB 1024x1024
  - Downsampling: 4 x Conv Layer mit Kernel 3x3 > leaky relu
  - Upsampling: 4 x Deconv Layer mit Kernel 4x4 > relu, tanh
  - Outputput: Grayscale 1024x1024
  - Stride = 2, Padding = same
  - Adam Optimizer
  - Instance-norm > Batch-norm

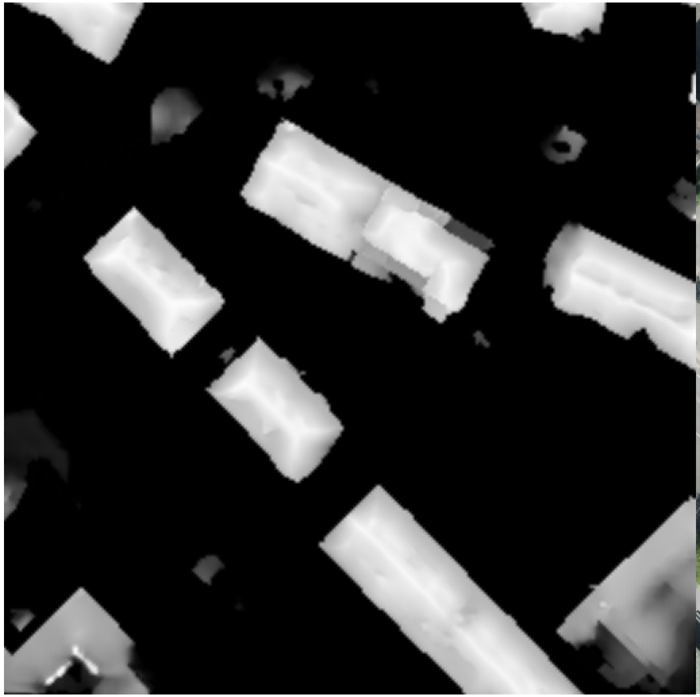


Netzwerkarchitektur

- GAN
  - Wie Autoencoder plus Discriminator
  - Batchnormalization, Batchsize=16
- Discriminator
  - Ground truth vs. output
  - Sigmoid Aktivierung
  - Loss: L2 + L_{GAN}

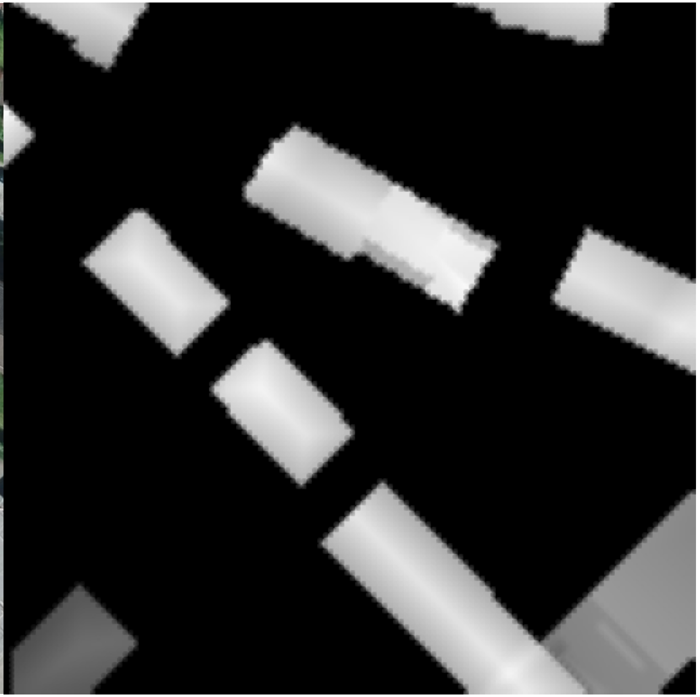


Ergebnisse



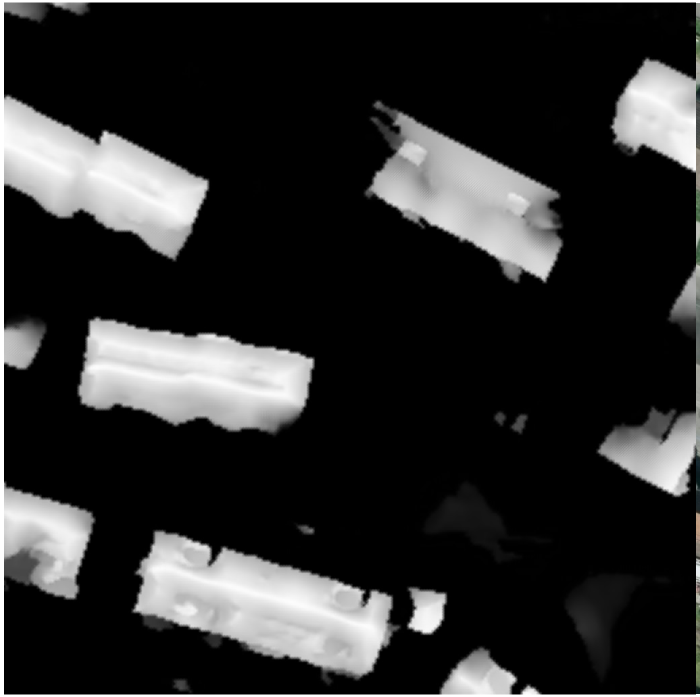
Output

Input



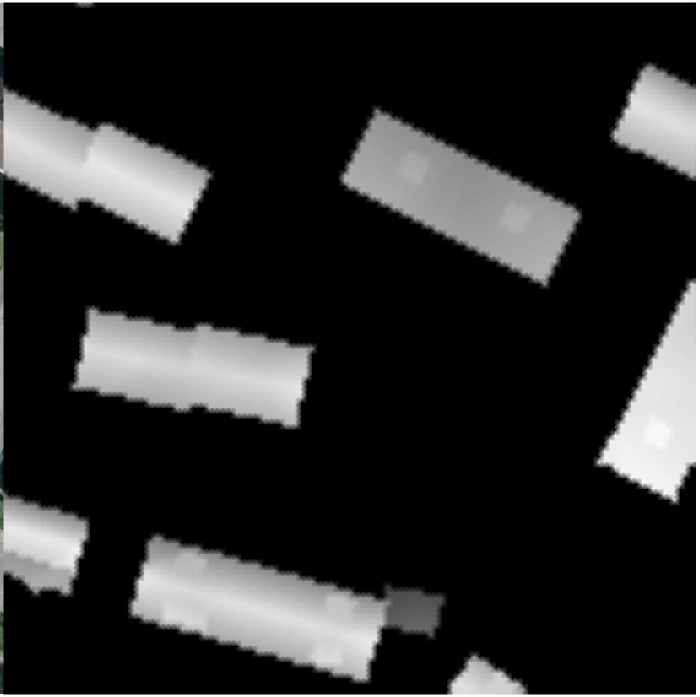
Ground Truth

Ergebnisse



Output

Input



Ground Truth

Showcase

Evaluation

- Mean Error
  - Train (128): 4.4 Meter
  - Test (128): 5.2 Meter
  - Test (1024): 8.7 Meter
- Optimierungsmöglichkeiten
  - Mehr Daten, Augmentation
  - Akkuratere Ground Truth