

A Virtual Reality Multi-Sensor 3d Reconstruction System



Markus Friedrich, Kyrill Schmid

Lehrstuhl für mobile und verteilte Systeme

LMU München



Ziel (in weiter Ferne 😊)

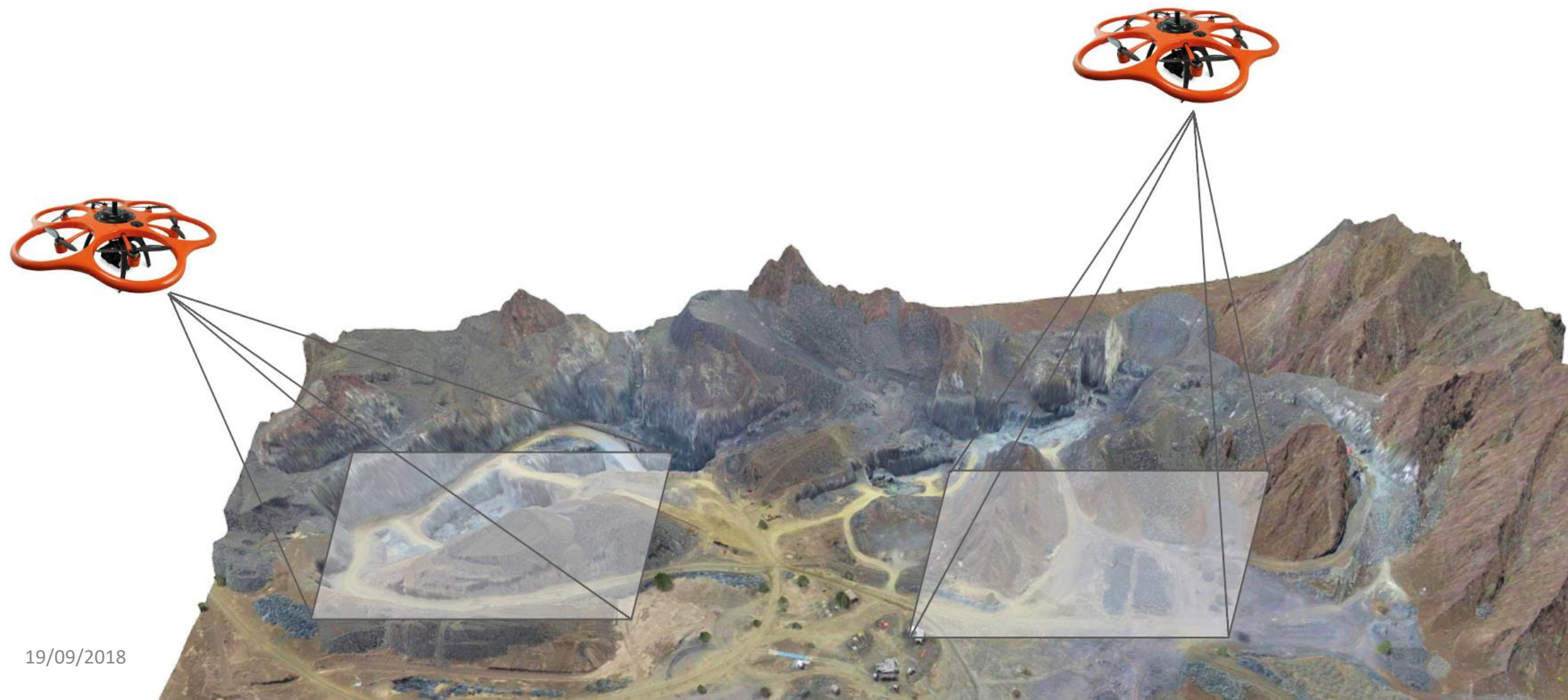
Eine Kollaborative 3d Scanning Plattform.

- Je nach Objektsegment und -oberfläche kommen verschiedene Scan-Geräte zum Einsatz.
- Resultate werden zusammen mit Qualitätsmetriken live eingeblendet (per VR, AR, Smartphones, Laptops).
- Mehrere Personen beteiligen sich an einem Scan-Task.

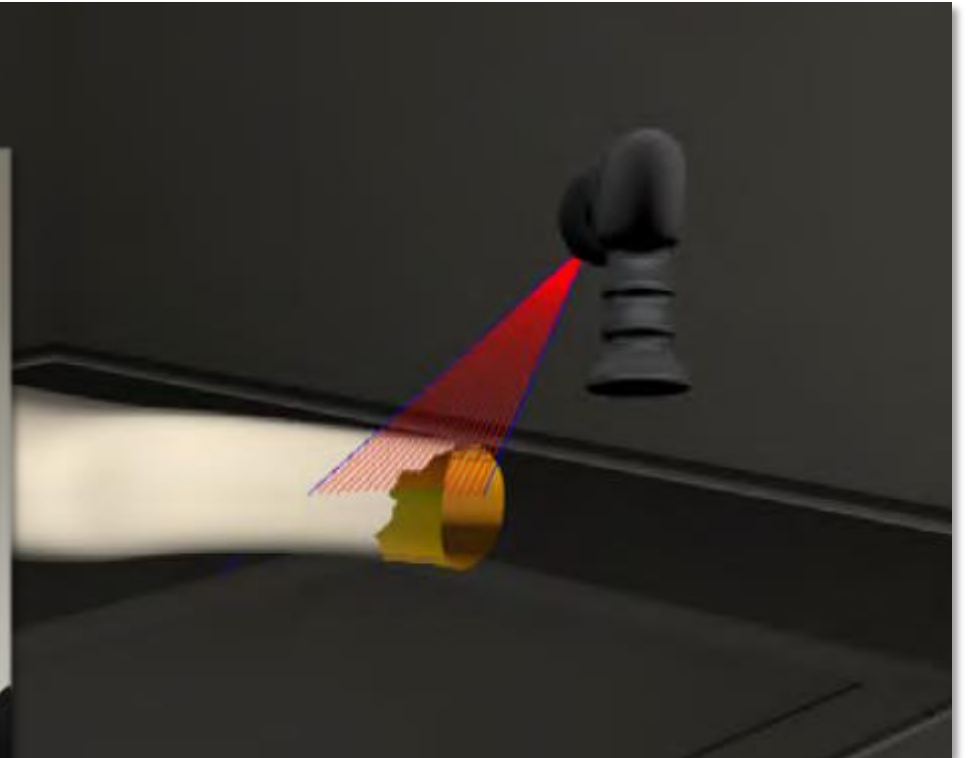


Quelle: <https://financialtribune.com/articles/energy-economy/68067/irans-mapna-unveils-improved-best-in-class-gas-turbine>

Ziel – Outdoor Szenario







Related Work



Schneider et al. „A Virtual-Reality 3d-Laser-Scan Simulation“, 2015

Sensorvielfalt

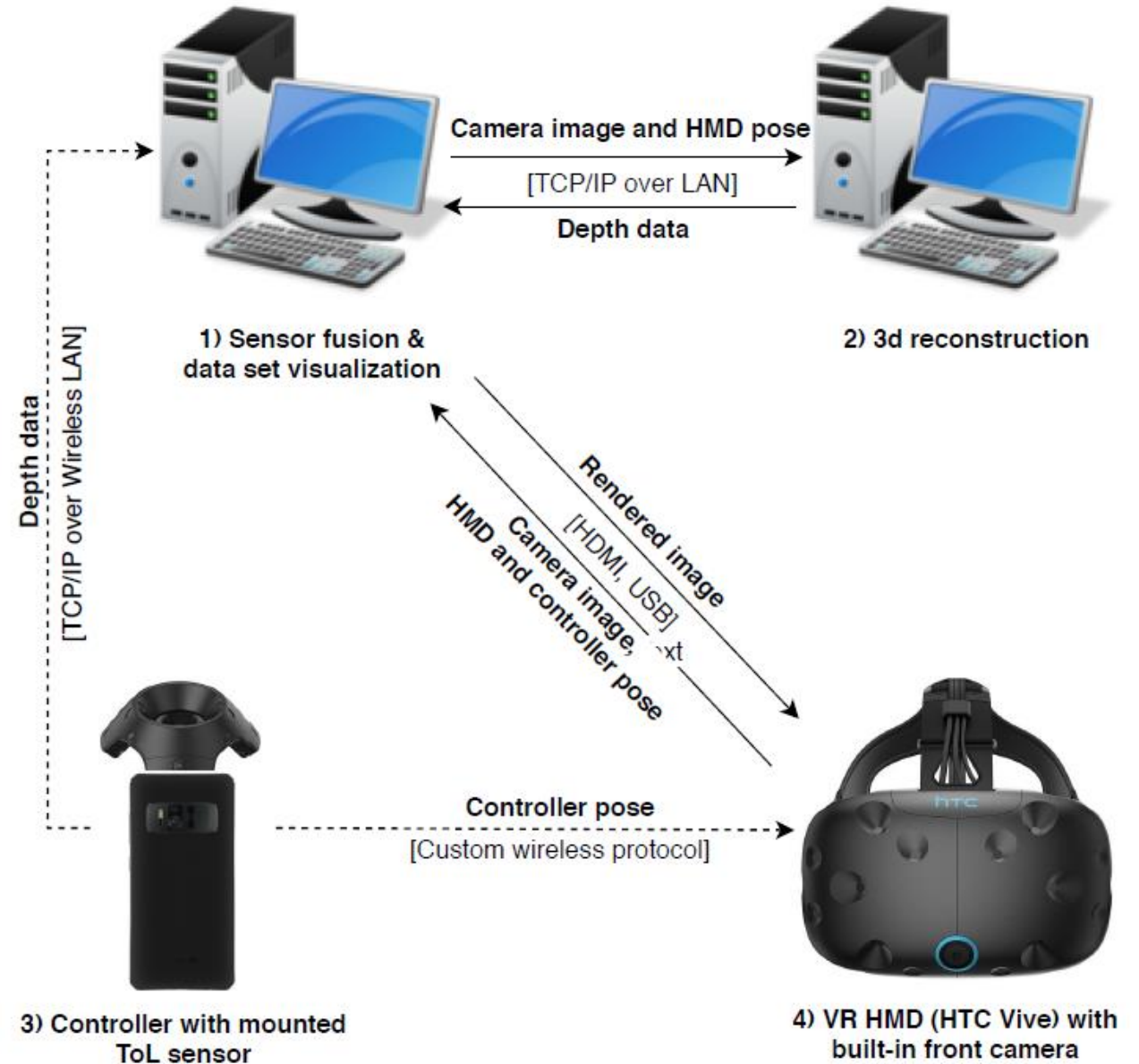
Quelle: Sean Ryan Fanello

Technology	Time Of Flight	Temporal structured light	Spatial structured light	Stereo
Design constraints				
Capture high speed motion	Multi-frame approach → Motion artifacts	Multi-frame approach → Motion artifacts	Possible, but best product available runs at 60fps	200Hz+ depth @ 720p harness mobile phone CMOS sensors
HW complexity	Custom sensor	Highly custom hardware	Simple hardware design	Simple hardware design
Multi-camera configurations (interference)	Some robustness	Bad interference issues	Major interference issues	No interference
Other known limitations	Multi-path artifacts	Short range	Depth estimation inefficient Exception [Fanello et al, CVPR 2016] but per camera calibration & assumes single camera/illuminator	Depth estimation inefficient

=> Rekonstruktionsqualität Qualität stark abhängig von: Oberflächenstruktur, Beleuchtung, Kalibrierung

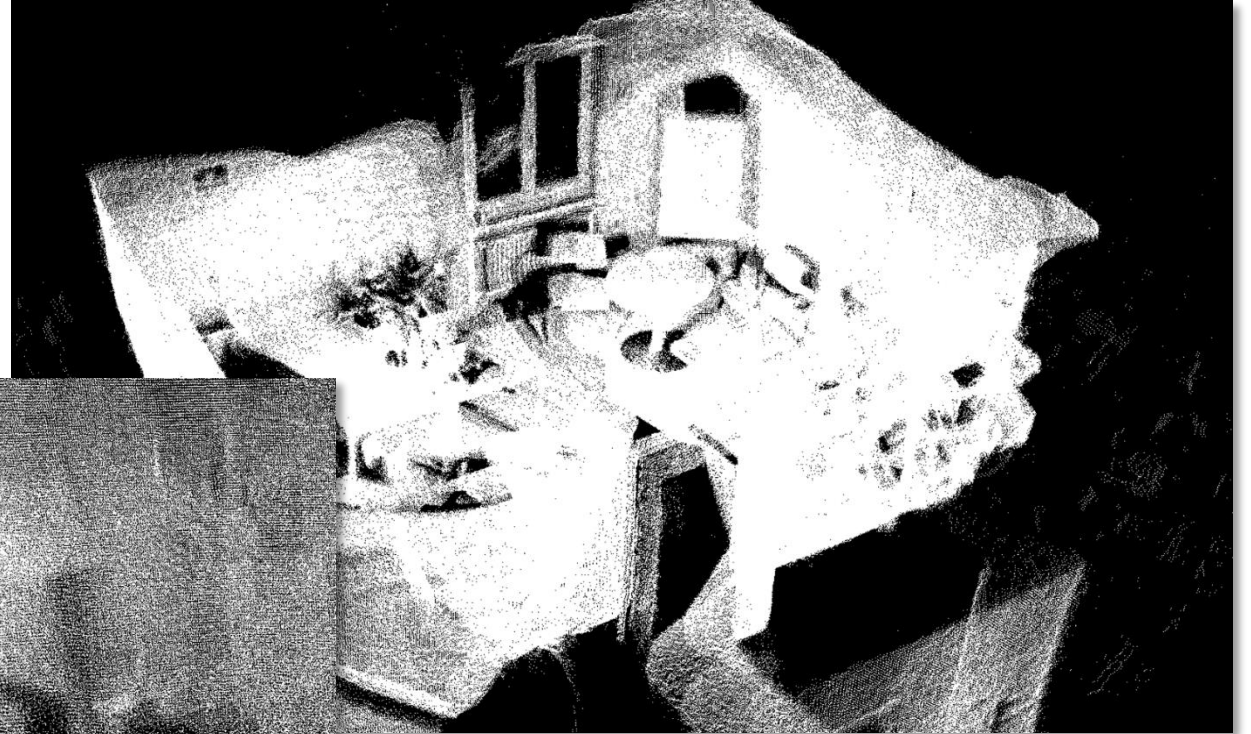
Prototyp - Architektur

- Asus Zenfone AR Time-of-Flight 160x120 sensor (Real3 Infineon)
- HTC Vive Kamera 1280x720 sensor
- **Wichtig:** Wir verwenden hier die Kamera der HTC Vive als Stereo Sensor (Kamera Pose bekannt)



ToF Sensor

6 Mio Punkte



3 Mio Punkte

```
struct PointCloudMessage {  
    uint msg_size , msg_timestamp ;  
    Matrix4x4f pose ;  
    Point4f[] points ; // x,y,z,color  
}
```

60.000 Punkte/s => 900kb/s

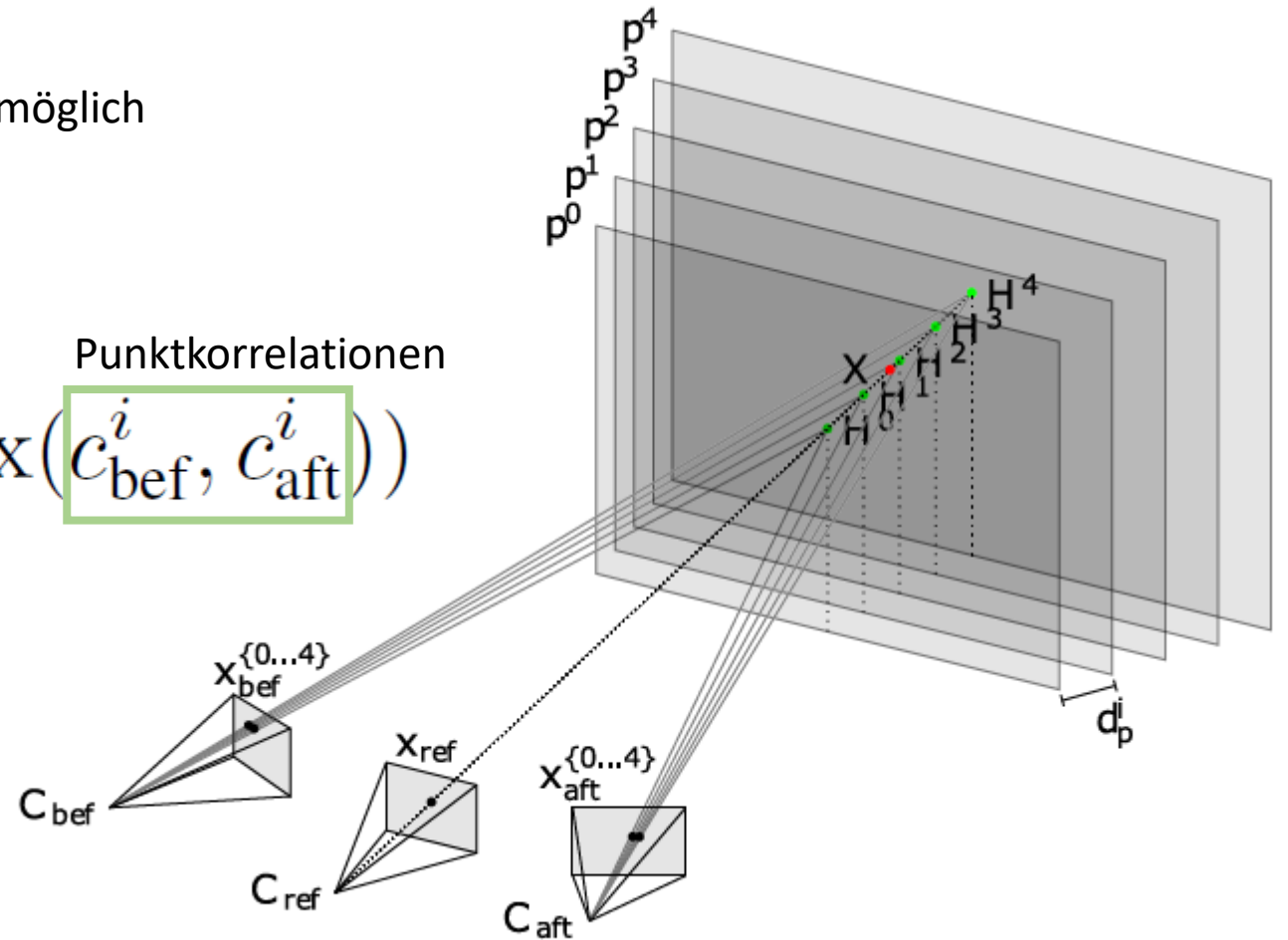
Prototype – Stereo Matching

- Plane Sweep Methode [Col96]
- Gut Parallelisierbar => GPU Implementierung möglich

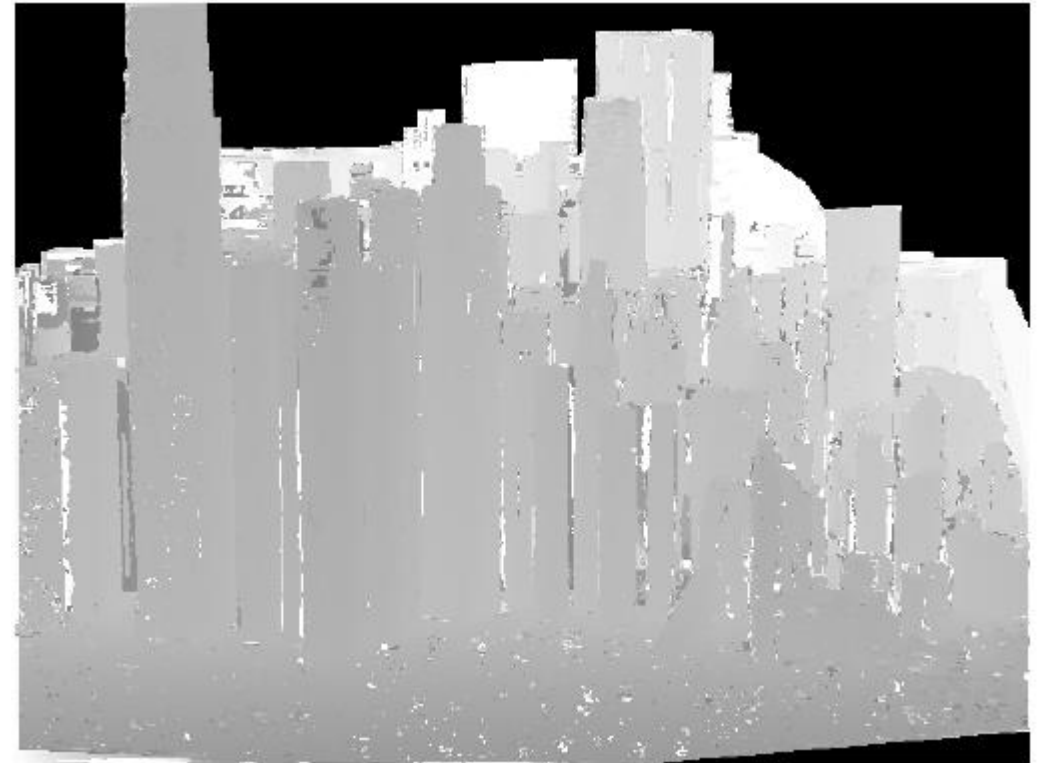
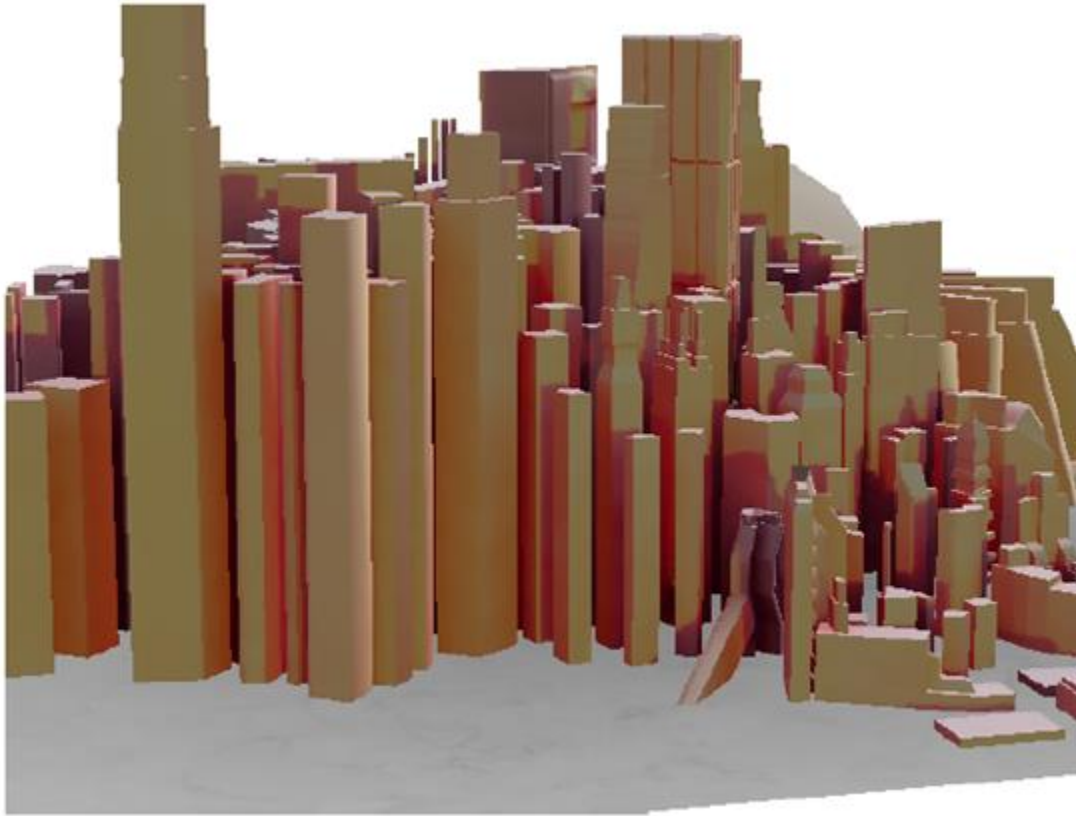
Index der besten Ebene

$$b = \operatorname{argmax}_{i \in \{0, \dots, n_p - 1\}} (\max(C_{\text{bef}}^i, C_{\text{aft}}^i))$$

Anzahl Ebenen



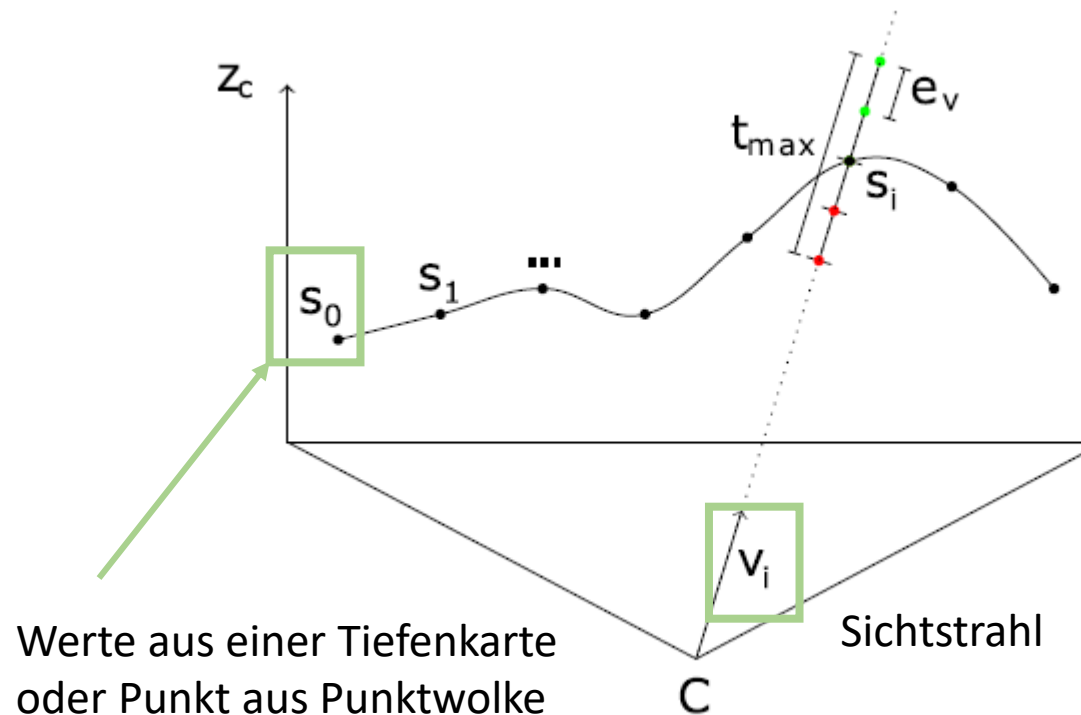
Depthmap Beispiel



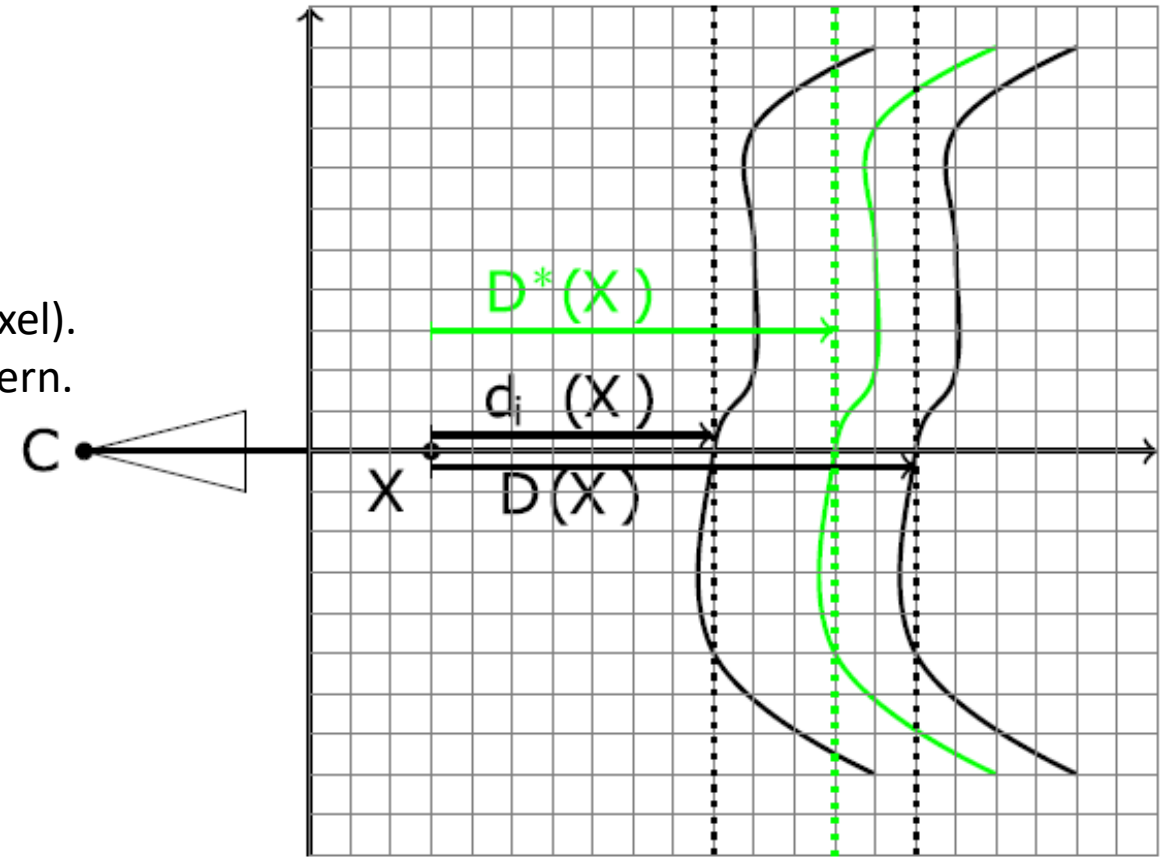
640x480, 1500 sweeping planes, < 100ms

Prototype - Fusion

- Oberfläche wird als **Truncated Signed Distance Function** (TSDF) beschrieben und diskretisiert (Voxel).
- Pro Voxel (X), Weight (W) und Distance (D) speichern.



- Distanz- und Gewichtsupdate (**Fusion**):

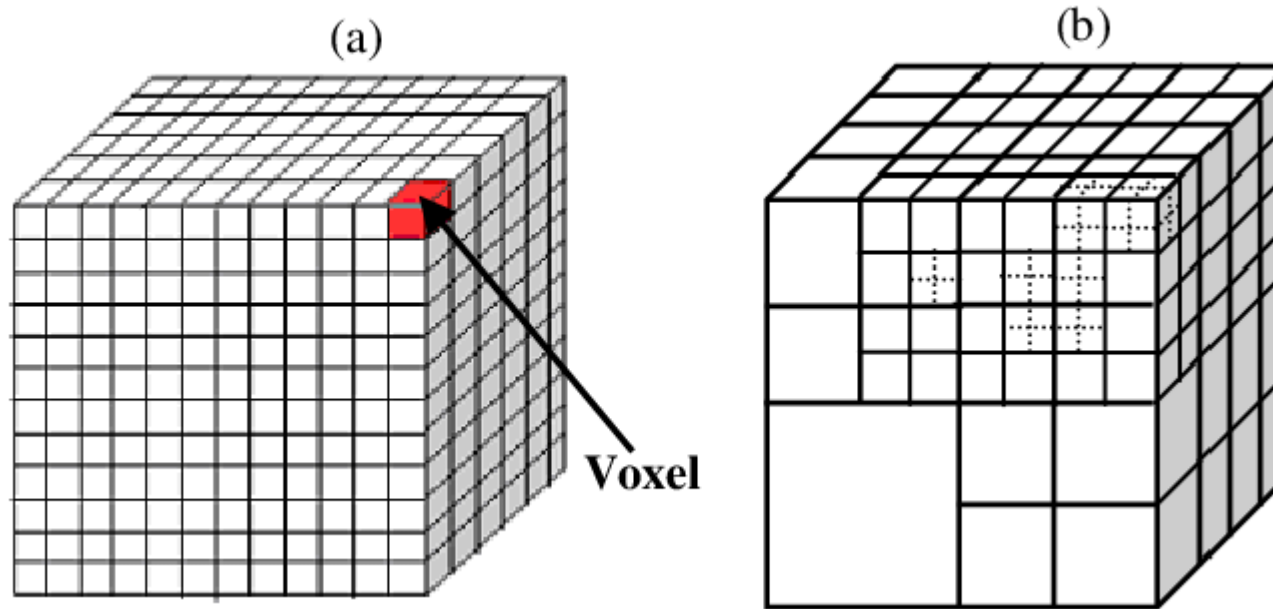


$$D^*(\mathbf{X}) = \frac{W(\mathbf{X})D(\mathbf{X}) + w_i d_i(\mathbf{X})}{W(\mathbf{X}) + w_i}$$

$$W^*(\mathbf{X}) = W(\mathbf{X}) + w_i$$

nach [CL96]

Storage



Grid: Nicht speichereffizient

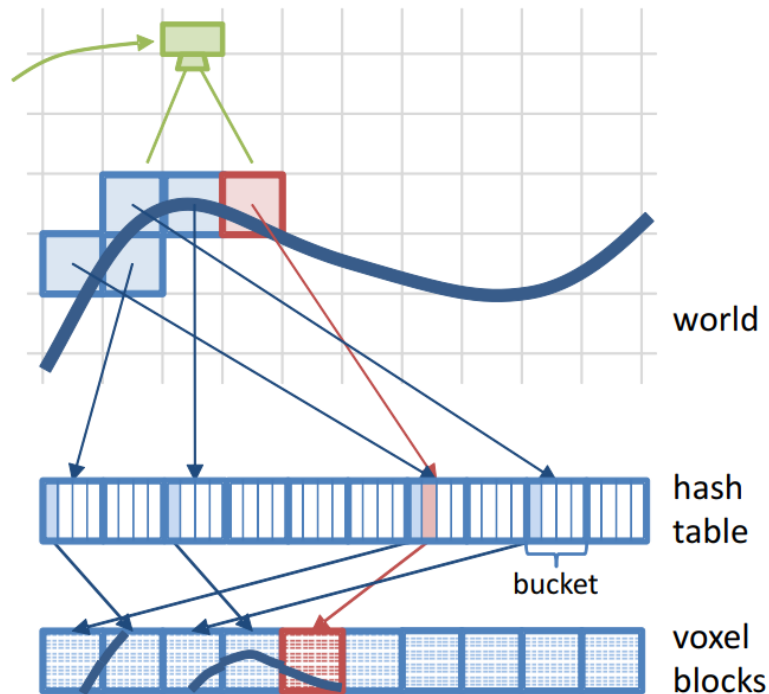
Octree: Komplex in der Handhabung (v.a. auf GPUs)

➔ **Spatial Hashmap**

Prototype - Storage

a, b, c : Große Primzahlen
 (x, y, z) : Punktkoordinaten

Spatial Hashmap



Quelle: [NZIS13]

$$h(x, y, z) = ((xa) \oplus (yb) \oplus (zc)) \bmod n_b \quad [\text{THM+03}]$$



195.000 voxels

300.000 Tiefenwerte => < 10ms

Auch geeignet für Streaming (VRAM <=> RAM)

Aktueller Stand

- Schlüsselemente der Plattform wurden implementiert
 - Plane-Sweep und Voxel-Hashmap für die GPU
 - Übertragungsprotokol für Punktwolken von Asus Zenfone AR
 - Fusion
 - Rudimentäre Visualisierung

Nächste Schritte

- Gesamtsystem zusammenfügen und testen.

Dann:

- Robustere Ansätze zur 3d Rekonstruktion evaluieren (schlechte Lichtbedingungen, schwache Texturierung)
- Evaluation von Kommunikationsstandards für Outdoor Use Case
- Fusion: Loop-Closure & Drift



Quelle: [DA17]

Fragen?

Quellen

- [Col96] R.T. Collins. A space-sweep approach to true multi-image matching. In Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on, pages 358–363, 1996.
- [NZIS13] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger. Realtime3D Reconstruction at Scale using Voxel Hashing. ACM Transactions on Graphics (TOG), 2013.
- [CL96] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In Proceedings of the 23rd annual conference on computer graphics and interactive techniques, pages 303–312. ACM, 1996.
- [THM+03] Matthias Teschner, Bruno Heidelberger, Matthias Müller, Danat Pomeranets, and Markus Gross. Optimized spatial hashing for collision detection of deformable objects. Proceedings of the 18th International Workshop on Vision, Modeling and Visualization, 2003.
- [DAI17] Dai, A., Nießner, M., Zollhöfer, M., Izadi, S., & Theobalt, C. (2017). Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. ACM Transactions on Graphics (TOG), 36(4), 76a.