



Institut für Kartographie und Geoinformatik | Leibniz Universität Hannover

Workshop zum FDM in den IngWissenschaften

FDM im GRK i.c.sens

Darmstadt 08./09. März 2018

Colin.Kuntzsch@ikg.uni-hannover.de



GRK i.c.sens - collaboration and integrity in dynamic sensor networks

▶ <https://www.icsens.uni-hannover.de>

▶ DFG-Graduiertenkolleg über 2x 4.5 Jahre

▶ 5 Institute, **Studiengang Geodäsie und Geoinformatik**

▶ Forschungsschwerpunkt autonome Fahrzeugnavigation

- Umgebungserfassung mit Automotive-Sensoren

- Themen

- Selbstlokalisierung (z.B. GNSS)

- Objektdetektion und –klassifikation (z.B. Photogrammetrie)

- Mapping (z.B. kinematisches Laserscanning)

▶ technischer Koordinator

- Betreuung IT, Experimente, **FDM**, ...

Research
Training
Group
i.c.sens



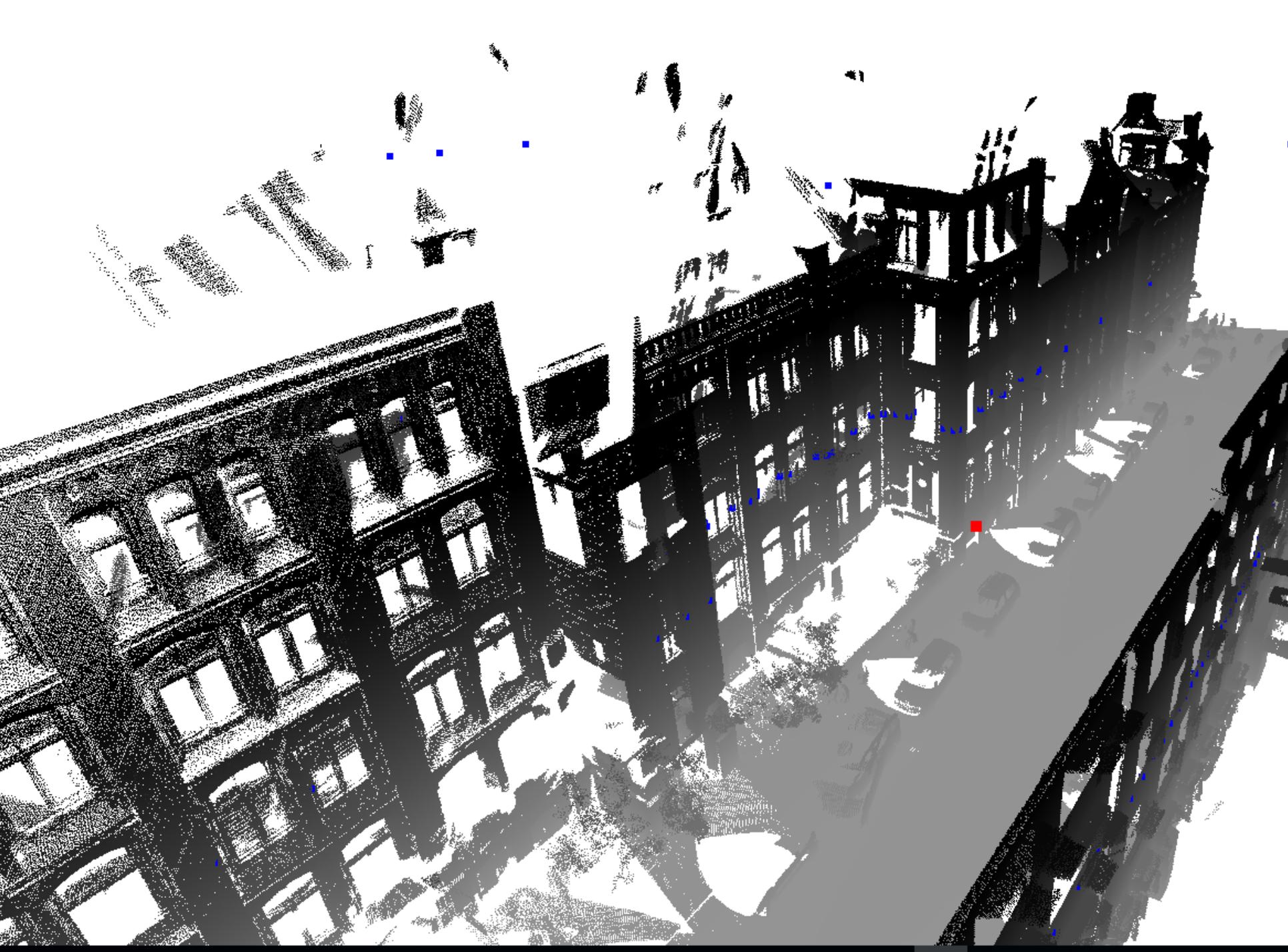
Forschungsdaten

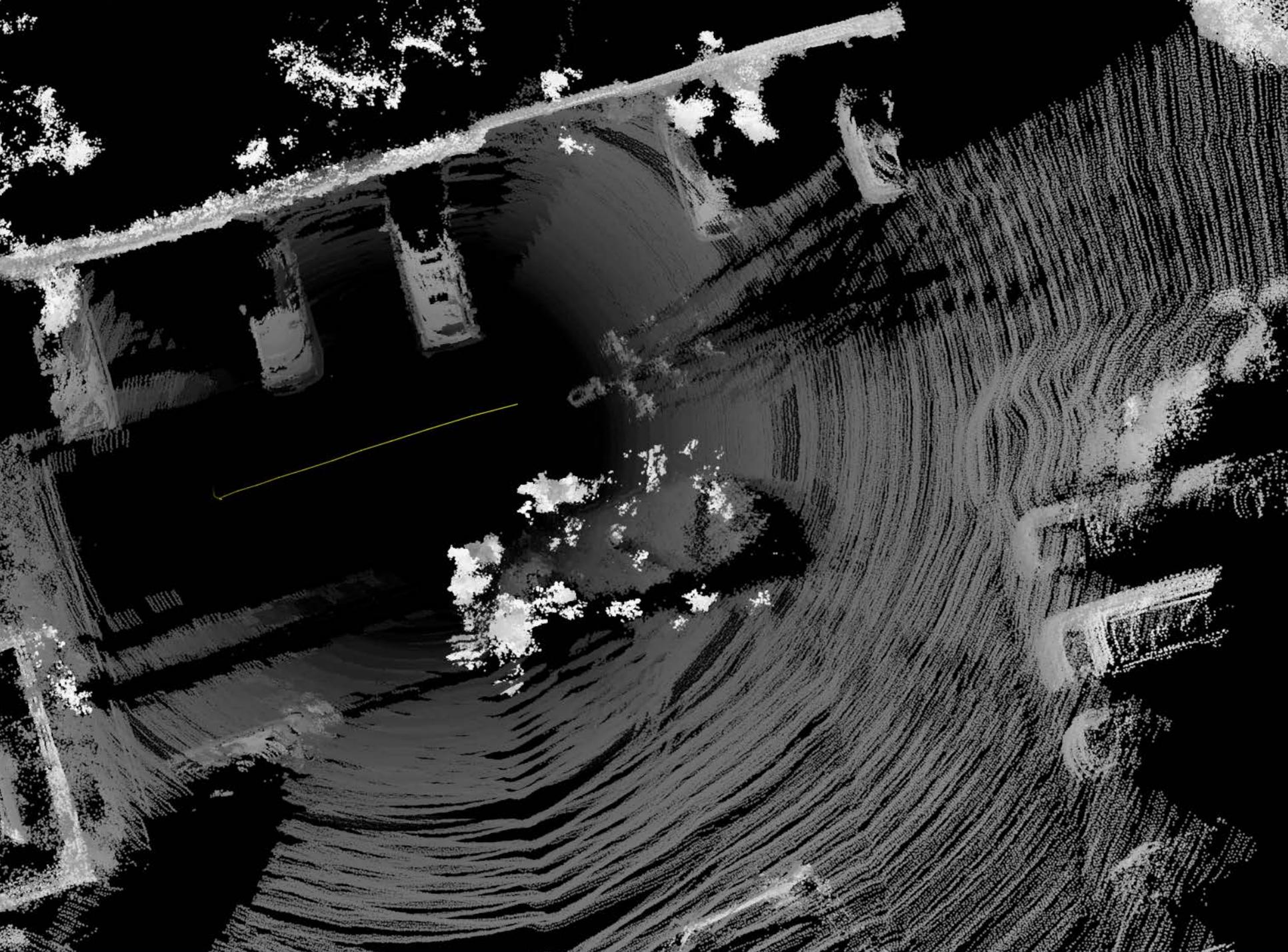
- ▶ heterogene Datenlage, Multi-Sensor-System aus mehreren mobilen Plattformen, zeitliche Synchronisierung
 - LiDAR-Punktwolken, Stereo-Kamera-Bildpaare, GNSS
 - Messkampagne („Mapathon“)
- ▶ Stereo-Video 25Hz, verlustfrei komprimiert (Full-HD)
 - insgesamt >500.000 Einzelbilder, insgesamt 1.75 TB (3 Systeme, 2h netto-Fahrtzeit)
 - datenschutzrechtlich relevant
- ▶ LiDAR
 - insgesamt >1 Mrd. Einzelpunkte, insgesamt >60 GB (1 System, 2h netto-Fahrtzeit)
 - Industriekameras zum Einfärben der Punktwolke, datenschutzrechtlich relevant
- ▶ GPS
 - proprietäre Formate und Auswertesoftware
 - Auswertungsergebnisse in Klartext-Formaten exportiert



11
102
1004

Leibniz
Universität
Hannover





IT/FDM-Infrastruktur

▶ Server Housing

- Fileserver ~16 TB (+16 TB) Kapazität
 - langfristige Speicherung gemeinsamer Daten (Rohdaten und gemeinsam verwendete abgeleitete Daten)
- Hadoop-Rechencluster → Big Data

▶ zukünftig Publikation von Daten per LUH-Forschungsdaten-Repository

▶ FDM im GRK

- interne Ablage, Datenaustausch (v.a. über Förderperioden hinweg)
- praktikable technische Realisierung von FDM zur Handhabung stark heterogener Daten über 3 Kohorten von Promovenden
- Minimierung der FDM-Tätigkeiten für Doktoranden
- geplante spätere Publikation der Daten

Realisierung FDM

- ▶ hierarchische Organisation der Daten per Ordnerstruktur
 - Ablage aller Daten in vereinheitlichten Formaten, einheitliche räumliche/zeitliche Referenzsysteme
- ▶ XML-Metadaten-File pro Datensatz*, eigenes Schema
 - **manuell** aus Vorlage zu erstellen, sobald ein Datensatz auf dem Server abgelegt wird
 - Informationen zu Verantwortlichkeiten, Experiment, Sensor-Details, räumliche/zeitliche Ausdehnung usw. → Nutzung zur Filterung
- ▶ eigene Software zur Transformation von Daten und Metadaten
 - Transformation von sensorspezifischen Formaten, zeitlichen/räumlichen Referenzsystemen
 - Traversieren des Verzeichnisbaums → Generierung einer aktuellen Metadaten-DB → Filterung der Metadaten mit SQL
- ▶ Web-Interface/Anbindung an Hadoop-Cluster
 - automatisierte Generierung von Daten-spezifischen Vorschau (Vektor- oder Rastergrafiken als Overlay über WebMap)

Veröffentlichung der Daten

- ▶ geplante Veröffentlichung von Teilen der Daten nach Auswertung durch Promovenden
 - Eignung als Benchmark-Datensatz? → erheblicher Mehraufwand
 - Selektion/Bereinigung eines Teils der Daten, Reduktion des Datenvolumens
 - Aufbereitung abgeleiteter (Standard-)Zwischenergebnisse, Dokumentation
- ▶ zentrales Forschungsdaten-Repository LUH im Aufbau (Ende 2018)
 - automatisierte DOI-Vergabe, Backend für große Datenmengen mit Backup, Autorensystem zur Verwaltung von Metadaten
 - zentrale Infrastruktur notwendig für Publikation großer Datenmengen

Bedarf an FDM-Infrastruktur

- ▶ zentrale Publikationsplattform für große Datenmengen
 - Garantien zu Persistenz und Langfristigkeit der Speicherung
- ▶ projektinterne IT-Infrastruktur zur zentralen, persistenten Ablage großer Datenmengen
 - „Content-Management“-Software
 - definierte FDM-Prozesse (z.B. auf Hochschul-Ebene), realisiert in Software/IT-Infrastruktur > geschriebene FDM-Guidelines
 - „feste“ Assoziation von Metadaten (+Publikationen, Dokumentation, ...) mit Daten
 - Verzeichnisdienst für Metadaten/Daten zur effizienten Suche
 - generisch → Unabhängigkeit von Datenformaten

Vielen Dank für Ihre
Aufmerksamkeit!