

# Bilder vom Boden und vom UAV: Orientierung, dichte 3D-Rekonstruktion und funktionale Modellierung

HELMUT MAYER<sup>1</sup>, HAI HUANG<sup>1</sup>, ANDREAS KUHN<sup>1</sup>, MARIO MICHELINI<sup>1</sup>, WILLIAM NGUATEM<sup>1</sup>, LUKAS ROTH<sup>1</sup>, MATTHIAS SCHMITZ<sup>1</sup> & MARTIN DRAUSCHKE<sup>2</sup>

*Zusammenfassung: Für eine detaillierte Modellierung von Gebäuden und ihrer Umgebung bietet sich die gemeinsame Auswertung von Bildern vom Boden und von kleinen Unmanned Aerial Vehicles (UAVs) an. Die Automatisierung der Auswertung führt zu einer Vielzahl von technischen und wissenschaftlichen Fragestellungen, von denen eine Reihe in diesem Beitrag angesprochen und für die an der Universität der Bundeswehr München und am Institut für Robotik und Mechatronik des Deutschen Zentrums für Luft- und Raumfahrt (DLR) entwickelte Lösungen dargestellt werden. Hierbei wird davon ausgegangen, dass nur Bilder mit ungefähre Kalibrierung, aber keine Navigationsinformation zur Verfügung steht. Die Leistungsfähigkeit der vorgestellten Ansätze für Orientierung, dichte 3D-Rekonstruktion und funktionale Modellierung werden anhand eines aussagekräftigen Beispiels aufgezeigt.*

## 1 Einleitung

Für die detaillierte dreidimensionale (3D) Modellierung von Gebäuden und ihrer Umgebung eignet sich insbesondere eine Kombination von Bildern vom Boden und von kleinen (< 5 kg) Unmanned Aerial Vehicles (UAVs). Diese erlaubt eine Abbildung der Gebäude von allen Seiten (siehe Abb. 1) auch aus unterschiedlichen Höhen sowie von oben und von schräg oben. Sie führt damit zu hoch aufgelöster Bildinformation aus frontaler Perspektive sowohl für Wände als auch für Dächer. Aus diesen sind extrem detaillierte 3D-Modelle ableitbar, die für Stadtplanung, Architektur, Navigation, Tourismus, Filme und auch für Einsatzplanung im urbanen Raum geeignet sind.

Der vorgestellte Ansatz ist sehr allgemein gehalten und geht für die Bilder nur davon aus, dass eine ungefähre Kalibrierung, wie sie aus den EXIF Daten für fast alle Kameras verfügbar ist, aber keine Navigationsinformation zur Verfügung steht. Damit kann für die Orientierung der 5-Punkt-Algorithmus von NISTÉR (2004) verwendet werden. In Kombination mit RANdom SAMple Consensus – RANSAC (FISCHLER & BOLLES 1981), kleinster Quadrate Bildzuordnung und robuster Bündelausgleichung erhält man einen Algorithmus, der hoch genaue Orientierungen auch für große Basen bestimmen kann. Die hierfür notwendige Information über die Verknüpfung der Bilder in Form von Bildtriplets wird mittels des Ansatzes von (MICHELINI & MAYER 2016) bestimmt.

Auf Grundlage der hoch genauen Orientierung wird eine dichte 3D-Rekonstruktion mit dem Ansatz von KUHN et al. (2016a) durchgeführt. Aus den mittels SemiGlobalem Matching – SGM (HIRSCHMÜLLER 2008) berechneten Tiefenkarten wird die Genauigkeit der Zuordnung für Paare abgeschätzt und in den 3D-Raum projiziert. Die Information für verschiedene Paare wird fusioniert und daraus hoch detaillierte 3D-Oberflächen abgeleitet. Da die Fusion lokal begrenzt ist, kann sie

---

<sup>1</sup> Universität der Bundeswehr München, Institut für Angewandte Informatik, Werner-Heisenberg-Weg 39, D-85577 Neubiberg, E-Mail: [Helmut.Mayer, Hai.Huang, Andreas.Kuhn, Mario.Michelini, Lukas.Roth, Matthias.Schmitz]@unibw.de

<sup>2</sup> Deutsches Zentrum für Luft- und Raumfahrt, Institut für Robotik und Mechatronik, Münchener Straße 20 D-82234 Oberpfaffenhofen-Weßling, E-Mail: Martin.Drauschke@dlr.de

parallel durchgeführt werden, was die Skalierbarkeit der 3D-Rekonstruktion auf große Mengen von hoch aufgelösten Bildern ermöglicht. Durch die Abschätzung der Zuordnungsgenauigkeit wird trotz der lokalen Begrenzung eine hohe Qualität erzielt.

Die bei der 3D-Rekonstruktion entstehenden 3D-Modelle sind extrem detailliert, damit allerdings auch bezüglich des Speicherplatzes sehr umfangreich und besitzen keine Semantik. Daher wird in der funktionalen Modellierung nach Objekten und deren Teilen, hier Gebäude mit Fassaden, Dächern, Fenstern und Türen gesucht. Grundlage ist eine effiziente Klassifikation, der bei der dichten 3D-Rekonstruktion entstehenden farbigen Punktwolke (KUHNER et al. 2016b) sowie die Berechnung zusammenhängender Flächenelemente (NGUATEM & MAYER 2016), die zu Dach-, Fassaden- und Bodenflächen führen. Die Dachflächen werden mit statistischen Verfahren klassifiziert (Giebeldach, Walmdach, Mansarddach etc.). Türen und Fenster werden als Strukturen hinter der Fassade bestimmt (NGUATEM et al. 2016). Parallel dazu werden in entzerrten Bildern der Fassade mittels eines Convolutional Networks Fenster und Türen klassifiziert und segmentiert (SCHMITZ & MAYER 2016).

Im Weiteren werden die einzelnen Ansätze für Orientierung, dichte 3D-Rekonstruktion und funktionale Modellierung im Detail vorgestellt.

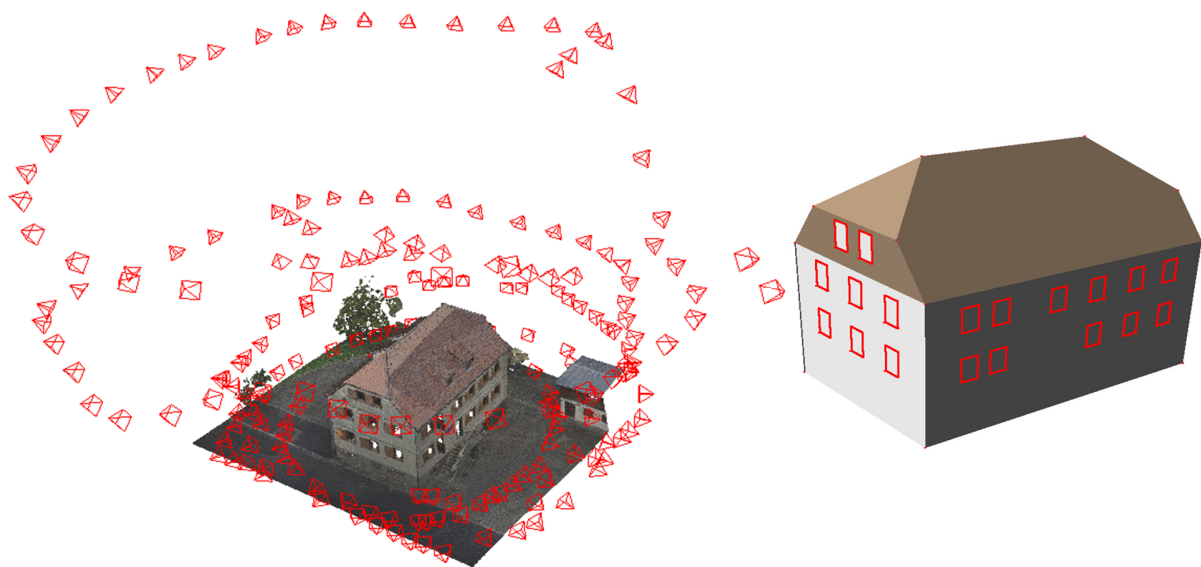


Abb. 1: Dichte 3D-Rekonstruktion eines Gebäudes inkl. der als Pyramiden dargestellten Orientierungen der Kameras (links) und funktionale Modellierung des Gebäudes mit Wänden (grau), Dach (braun) und Fenstern (rot umrandet) (rechts)

## 2 Orientierung

Die (relative) Orientierung der genähert kalibrierten Bilder erfolgt auf Grundlage des 5-Punkt-Algorithmus sowie RANSAC. Für große Basen ergeben sich wenig Inlier für RANSAC. Insbesondere sind die mittels 5-Punkt-Algorithmus bestimmten Lösungen auch für fünf Inlier oft nur lokal, d.h. für Teile der maximal möglichen Inlier repräsentativ. Deswegen wird RANSAC mit robuster Bündelausgleichung in Form eines an Expectation Maximization (EM) angelehnten Ansatzes (MAYER et al. 2012) kombiniert, bei dem lokale Lösungen schrittweise erweitert werden

und womit eine stabile Lösung auch für große Basen möglich wird. Diese Vorgehensweise wird für Paare und für Triplets verwendet. Triplets werden hierarchisch zu Bildmengen vereinigt (MAYER 2014).

Die effiziente automatisierte Bestimmung der Verknüpfung für größere Bildblöcke erfolgt auf Grundlage von Merkmalsextraktion auf der Graphikkarte. Das zentrale Problem für größere Mengen von (ungeordneten) Bildern ist der exponentielle Anstieg der Paare. Um mit diesem umzugehen, findet eine Reduktion des bei der Bestimmung der Verknüpfung entstehenden Graphen statt. Da eine Übertragung der Skalierung der Orientierung nur für Triplets möglich ist und auch nur diese für große Basen eine stabile Orientierung ermöglichen, wurden zudem Baumstrukturen (Steinerbäume) für Triplets verwendet, die eine (näherungsweise) minimale Verbindung garantieren. Nach Verifikation von Paaren und Triplets mittels des Ansatzes von (MAYER et al. 2012) und einer evtl. notwendigen Neuberechnung der Verknüpfung erfolgt zuletzt eine gezielte Verdichtung der Netzstruktur zur Stabilisierung der Geometrie (MICHELINI & MAYER 2016).

Die in (MICHELINI & MAYER 2016) beschriebene Vorgehensweise eignet sich für große Basen. Die Zuordnung von Paaren mit extrem großen Basen ist wünschenswert, im Fall unbekannter Verknüpfung aber unrealistisch, da erst nach der rechenaufwändigsten Variante der Zuordnung klar wird, ob ein Paar eine extreme Basis besitzt oder einfach nicht denselben Ausschnitt der Welt zeigt. Je größer die Bildmenge ist, desto größer wird die Wahrscheinlichkeit für Letzteres. Für kleine Mengen (z.B. < 20) oder bei manueller Vorgabe ist allerdings eine Zuordnung von Bildern mit extremer Basis sinnvoll. Hierzu wird in (ROTH et al. 2017) ein Ansatz vorgestellt, bei dem projektiv verzerrte Bilder simuliert werden. In den verzerrten Bildern werden verschiedenen aufwändige Merkmalsdetektoren angewandt, die Merkmale werden zugeordnet und zuletzt die Zuordnungen mittels 5-Punkt-Algorithmus überprüft. Abb. 2 demonstriert, dass mit der vorgestellten Vorgehensweise auch perspektiv extrem verzerrte Bilder mit kleinem Überlappungsbereich erfolgreich zugeordnet und orientiert werden können.

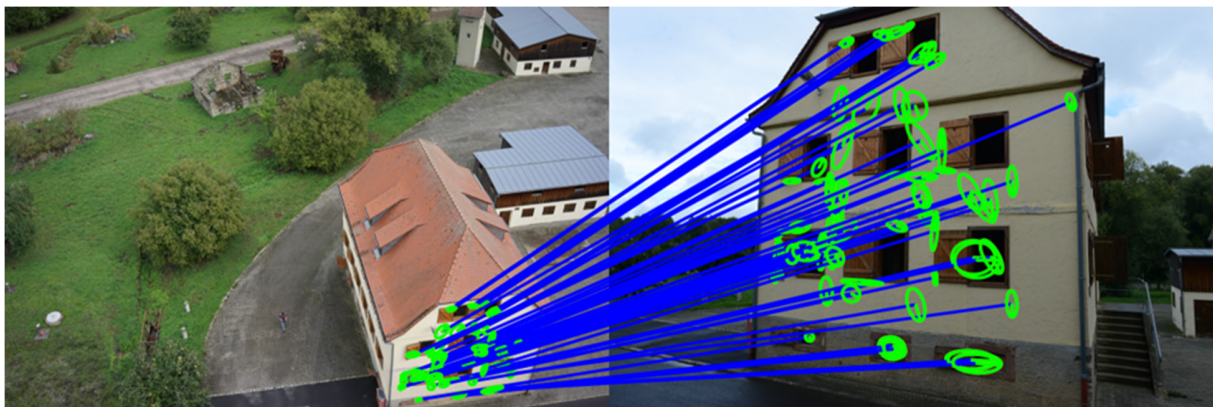


Abb. 2: Zuordnung von Punkten zwischen Bildern mit extrem großer Basis. Grün: Punkte mit Größe und Ausrichtung; blau: Zuordnung.

### 3 Dichte 3D-Rekonstruktion

Die hoch genaue Orientierung inkl. Genauigkeitsinformation aus der Bündelausgleichung ist die Grundlage für die dichte 3D-Rekonstruktion (KUHN et al. 2016a). Hierfür werden mit SGM für Bildpaare Disparitätskarten berechnet. Für die Disparität wird die Unsicherheit auf Grundlage eines auf Total Variation (TV) basierenden Maßes geschätzt. Die Zuordnungsfunktion wurde aus Ground Truth Daten mittels eines Lernansatzes unter Nutzung von EM bestimmt. Die Unsicherheit wird in den 3D-Raum propagiert und führt zu Wahrscheinlichkeiten für Voxel, dass diese vor bzw. hinter einer Oberfläche liegen. Die Ergebnisse verschiedener Paare werden mittels binärer Bayes Fusion kombiniert und die Oberfläche als der wahrscheinlichste Ort zwischen den Voxel vor und hinter der Oberfläche bestimmt.

Der Ansatz wird skalierbar, indem der 3D-Raum mittels eines Octrees dynamisch unterteilt wird (KUHN & MAYER 2015). Da die Fusion lokal begrenzt durchgeführt wird, reicht eine vorgegebene Überlappung aus, um ein global konsistentes Ergebnis zu erzielen. Weil die Fusion in den unterschiedlichen Zweigen eines Octrees unabhängig durchgeführt werden kann, ist der Ansatz parallelisierbar. Nach Filterung bezüglich Sichtbarkeit wird die Oberfläche zuletzt trianguliert.

Der Ansatz ist durch die Parallelisierbarkeit gut für Rechner mit vielen Kernen und für Rechnercluster geeignet. Durch Nutzung der Genauigkeitsinformation der Zuordnung führt er zudem zu Ergebnissen mit hoher Qualität. Dies gilt im Gegensatz zu vielen anderen Ansätzen auch für große Mengen von Bildern (hunderte oder tausende) mit hoher Auflösung (50 Megapixel und mehr).

Abb. 3 stellt zwei schattierte Ansichten des Gebäudes aus Abb. 1 dar. Insbesondere zeigt die rechte Ansicht eine Treppe, bei dem die dünnen Stangen des Geländers 3D rekonstruiert wurden. Dies zeigt die Detaillierung der 3D-Rekonstruktion.

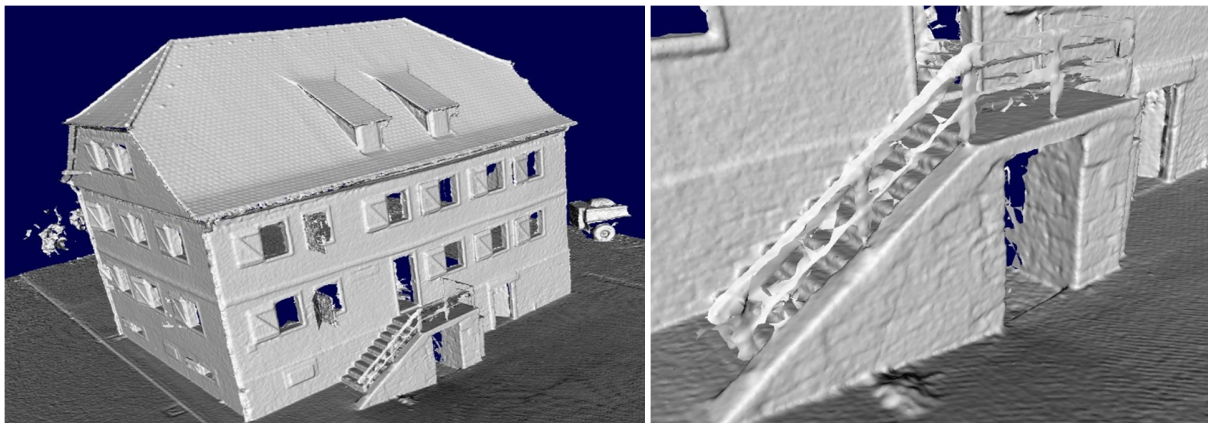


Abb. 3: Schattierte Ansichten der 3D-Rekonstruktion des Gebäudes aus Abb. 1. Der Ausschnitt rechts zeigt die 3D-Rekonstruktion der dünnen Stangen des Geländers.

### 4 Funktionale Modellierung

Die Modelle der 3D-Rekonstruktion zeigen kleinste Details, benötigen aber auch große Mengen an Speicherplatz (für das in Abb. 1 und 3 dargestellte Modell sind es für 23,6 Millionen Dreiecke bei binärer Speicherung ein halbes Gigabyte). Darüber hinaus besitzen die Dreiecke keine Semantik. In der funktionalen Modellierung werden daher Objekte und ihre Teile, hier Gebäude mit Fassaden, Dächern, Fenstern und Türen bestimmt.

Der erste Schritt besteht in einer Klassifikation der bei der dichten 3D-Rekonstruktion entstehenden farbigen Punktwolke in die Klassen Boden, Gebäude, Gras und Bäume (KUHNS et al. 2016b). Der Ansatz ist effizient, weil er ausnutzt, dass für die Klassifikation keine hoch aufgelöste 3D-Oberfläche notwendig ist. Mittels Occupancy Grids wird eine sehr schnelle Fusion von 3D-Punkten inkl. Wahrscheinlichkeiten und Farbwerten möglich. Eine weitere Beschleunigung wird durch die Nutzung von (erweiterten) Superpixeln erzielt, die mittels Segmentierung bezüglich Farbe, Höhe, und Neigung abgeleitet werden (HUANG & MAYER 2015). Durch die Nutzung relativer Merkmale ist der Ansatz nicht nur robust bezüglich Änderungen in den Umgebungsbedingungen, sondern auch parallelisierbar.

Die Klassifikation führt zu Bereichen mit einer hohen Wahrscheinlichkeit für Gebäude. In solchen Bereichen werden zusammenhängende Flächenelemente mittels eines Clustering Verfahrens berechnet (NGUATEM & MAYER 2016). Die Schätzung der Flächennormalen für Voxel auf Grundlage von RANSAC macht den Ansatz robust und effizient. Eine Dirichlet-Prozess Mixtur (DPM) Modell von Gauß-Funktionen vermeidet die Vorgabe der Zahl der Flächen. Zuletzt werden zusammenhängende Flächen bestimmt.

Aus den zusammenhängenden Flächen werden mittels statistischer Verfahren Dach, Fassaden und Boden abgeleitet. Für die Dachflächen führen Sequentielles Monte Carlo Sampling und Modellselektion zur Unterscheidung von Giebeldach, Walmdach und Mansarddach (NGUATEM et al. 2013). Das Dach wird um vertikale Fassaden ergänzt und Fenster sowie Türen werden als Strukturen hinter den Fassaden bestimmt. Hierzu werden verschiedene Hypothesen für die Fassadenebene gebildet, die zu unterschiedlichen Segmentierungen der 3D-Punkte führen. Auf deren Grundlage erfolgt eine probabilistische Musterzuordnung inkl. Modellselektion (NGUATEM et al. 2014). Einen Überblick über alle Schritte der funktionalen Modellierung von Gebäuden gibt (NGUATEM et al. 2016).

In einem parallel zu den statistischen Verfahren für 3D-Punkte entwickelten Ansatz (SCHMITZ & MAYER 2016) werden in entzerrten Bildern der Fassade mittels eines Convolutional Networks Fenster und Türen klassifiziert und segmentiert. Für Letzteres werden neben „Convolutional“ auch „Deconvolutional“ Verknüpfungen verwendet, die die Auflösung nach der Klassifikation vergrößern. Die Nutzung der Bildinformation macht den Ansatz unabhängig von Löchern und unscharfen Kanten in der Punktwolke.

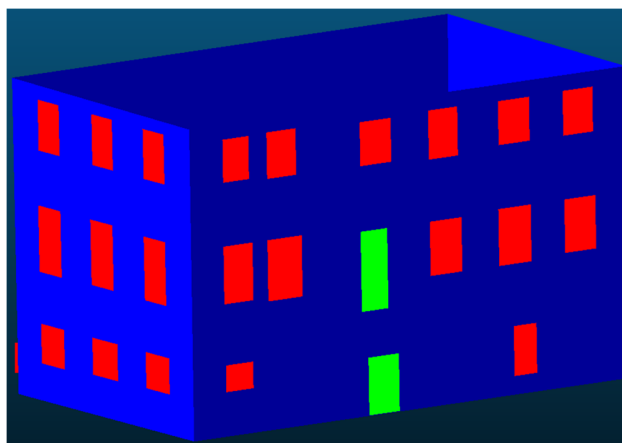


Abb. 4: 3D-Projektion der Ergebnisse der Interpretation von entzerrten Fassadenbildern mittels eines Convolutional Networks mit Fenstern (rot) und Türen (blau) für das Gebäude aus Abb. 1. Die Fassadenebenen wurden mit Hilfe des Ansatzes von NGUATEM & MAYER (2016) bestimmt.

## 5 Fazit & Ausblick

Der Beitrag gibt einen Überblick über verschiedene an der Universität der Bundeswehr München und dem Institut für Robotik und Mechatronik des Deutschen Zentrums für Luft- und Raumfahrt (DLR) entwickelte Ansätze, die von der Orientierung über die dichte 3D-Rekonstruktion bis hin zur funktionalen Modellierung von Gebäuden reichen.

In der Zukunft sollen die Ansätze für die funktionale Modellierung in Richtung komplexerer Gebäude sowie weiterer Objektklassen erweitert werden. Für eine Verbesserung der Zuverlässigkeit sollen die Ergebnisse für die statistischen Verfahren mit denen der Convolutional Networks kombiniert werden.

## 6 Literaturverzeichnis

- FISCHLER, M. & BOLLES, R., 1981: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM* **24**(6), 381-395.
- HIRSCHMÜLLER, H., 2008: Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(2), 328-341.
- HUANG, H. & MAYER, H., 2015: Robust and Efficient Urban Scene Classification Using Relative Features. *23rd SIGSPATIAL International Conference on Advances in Geographic Information System*, 81:1-81:4.
- KUHN, A., HIRSCHMÜLLER, H. & MAYER, H., 2013: Multi-Resolution Range Data Fusion for Multi-View Stereo Reconstruction. *German Conference on Pattern Recognition*, 41-50.
- KUHN, A., HIRSCHMÜLLER, H., SCHARSTEIN, D. & MAYER, H. 2016a: A TV Prior for High-Quality Scalable Multi-View Stereo Reconstruction. *International Journal of Computer Vision*, Online First.
- KUHN, A., HUANG, H., DRAUSCHKE, M. & MAYER, H., 2016b: Fast Probabilistic Fusion of 3D Point Clouds via Occupancy Grids for Scene Classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (III-3)*, 325-332.
- KUHN, A. & MAYER, H., 2015: Incremental Division of Very Large Point Clouds for Scalable 3D Surface Reconstruction. *International Conference on Computer Vision Workshop (ICCVW)*, 4321-4329.
- KUHN, A., MAYER, H., HIRSCHMÜLLER, H. & SCHARSTEIN, D., 2014: A TV Prior for High Quality Local Multi-View Stereo Reconstruction. *2nd International Conference on 3D Vision (3DV)*, 65-72.
- MAYER, H., 2014: Efficient Hierarchical Triplet Merging for Camera Pose Estimation. *German Conference on Pattern Recognition*, 399-409.
- MAYER, H., BARTELTSEN, J., HIRSCHMÜLLER, H. & KUHN, A., 2012: Dense 3D Reconstruction from Wide Baseline Image Sets. *Outdoor and Large-Scale Real-World Scene Analysis – 15<sup>th</sup> International Workshop on Theoretical Foundations of Computer Vision, Lecture Notes in Computer Science 7474*, Springer, Berlin, 285-304.
- MAYER, H. & MICHELINI, M., 2016: Orientierung großer Bildverbände. *Handbuch der Geodäsie, Band Photogrammetrie und Fernerkundung*, 32 Seiten.

- MICHELINI, M. & MAYER, H., 2016: Efficient Wide Baseline Structure from Motion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (III-3)*, 99-106.
- NGUATEM, W., DRAUSCHKE, M. & MAYER, H., 2013: Roof Reconstruction from Point Clouds Using Importance Sampling, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II (3/W3)*, 73-78.
- NGUATEM, W., DRAUSCHKE, M. & MAYER, H., 2014: Localization of Windows and Doors in 3D Point Clouds of Facades. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II (3)*, 87-94.
- NGUATEM, W., DRAUSCHKE, M. & MAYER, H., 2016: Automatic Generation of Building Models with Levels of Detail 1-3. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (XLI-B3)*, 649-654.
- NGUATEM, W. & MAYER, H., 2016: Contiguous Patch Segmentation in Pointclouds. *German Conference on Pattern Recognition*, 131-142.
- NISTÉR, D., 2004: An Efficient Solution to the Five-Point Relative Pose Problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(6), 756-770.
- ROTH, L., KUHN, A. & MAYER, H., 2017: Wide-Baseline Image Matching with Projective View Synthesis and Calibrated Geometric Verification. *Photogrammetrie – Fernerkundung – Geoinformation*, angenommen.
- SCHMITZ, M. & MAYER, H., 2016: A Convolutional Network for Semantic Facade Segmentation and Interpretation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (XLI-B3)*, 649-654.