

Modelle der realen Welt für die Bildinterpretation

Zusammenfassung

Dieser Beitrag beschäftigt sich mit Modellen der realen Welt für die Bildinterpretation, die der Autor seit Anfang der 90er Jahre zuerst bei Siemens und später an der Universität der Bundeswehr München (UniBwM) z. T. in Interaktion mit dem Kollegen WOLFGANG REINHARDT entwickelt hat. Am Anfang stand die Extraktion von semantischer Information aus gescannten Karten, die WOLFGANG REINHARDT bei Siemens bzw. Siemens Nixdorf Informationssysteme (SNI) mit betreut hat. In der Folge entwickelte der Autor Modelle für photographische Bilder der Objekte, beginnend mit Straßen in Luftbildern. Im Laufe der Zeit kamen Vegetation und Gebäude hinzu, wobei die Aufnahmen von Satelliten und UAVs (Unmanned Aerial Vehicles), z.T. zusammen mit höchst aufgelösten Bodenaufnahmen, stammen. Der Beitrag zeigt Entwicklungslinien auf, analysiert diese und gibt einen Ausblick auf die Zukunft.

Summary

Models of the Real World for Image Interpretation

This contribution is concerned with models of the real world for image interpretation, which I have developed since the 90ies at first at Siemens and later at Bundeswehr University Munich (UniBwM) partly in interaction with my colleague WOLFGANG REINHARDT. It started with the extraction of semantic information from scanned maps, which WOLFGANG co-supervised at Siemens and Siemens Nixdorf Informationssysteme (SNI). Later, I developed models for photographic images of objects, starting with roads in aerial images. Over the years vegetation and buildings were added and images stem from satellites and UAVs (Unmanned Aerial Vehicles) possibly in combination with very high resolution terrestrial pictures. The contribution presents directions of development, analyzes them and gives an outlook to the future.

1. Einleitung

Die Interpretation von Bilddaten impliziert ein dahinter stehendes Modell oder eine Modellierungstechnik. Dieser Beitrag beschäftigt sich mit den vom Autor seit Anfang der 90er Jahre entwickelten Modellen für Objekte der realen Welt und deren Abbildung. Hierbei wird die Entwicklung von einer expliziten symbolischen, strukturellen Modellbildung Anfang der 90er Jahre über den Fokus auf statistische und 3D-Modellierung seit ungefähr dem Jahr 2000 hin zur impliziten Modellierung mittels Convolutional (Neural) Networks an Hand verschiedener Typen von Objekten aufgezeigt. Abschließend wird diese Entwicklung analysiert und es wird ein Ausblick auf die Zukunft gegeben.

2. Modellierung für die semantische Interpretation von gescannten Karten

Am Anfang standen Arbeiten zur Interpretation von gescannten Karten [MAYER, 1994a]. Die grundlegende Idee bestand in einer expliziten Modellierung von Objekten und Relationen. Diese erfolgte auf Grundlage eines semantischen Netzes [NIEMANN et al., 1990], wobei die Objekte als Knoten und die Re-

lationen als Kanten modelliert wurden. Um automatische Schlussfolgerung auf Grundlage des Netzes zu ermöglichen, fand eine Beschränkung auf wenige Arten von Relationen, speziell auf die Teil- und die Spezialisierungsrelation statt. Damit war der Ablauf für die Schlussfolgerung aus dem semantischen Netz ableitbar.

Um allein auf Grundlage von Teil- und Spezialisierungsrelationen modellieren zu können, waren spezielle Objekte wie „*pavement_cycle-track_road-side_parallel_to_road*“ (siehe Abb. 1) notwendig. Dieses drückt aus, dass Straßenrand und Gehweg meist parallel zur Straße verlaufen. Darüber hinaus umfassten die Objekte spezielle Funktionen, die Information über Geometrie, Topologie und Semantik kodierten, deren Semantik aber außerhalb der automatischen Schlussfolgerung stand. Dies vergrößerte die Ausdrucksfähigkeit stark, macht zugleich aber eine logische Konsistenzprüfung unmöglich.

Bereits in MAYER [1994b] wurde analysiert, wieweit das in Objektartenkatalogen wie z.B. ATKIS vorliegende Wissen über Objekte ausreicht, um diese aus Bildern zu extrahieren. Das Ergebnis der Analyse war, dass zwar die Angaben im ATKIS Objektartenkatalog für menschliche Operateure ausreichend sind, aber weil insbesondere weder das Wissen über das Aussehen noch die Relationen der Objekte explizit gegeben sind, sind die Angaben für

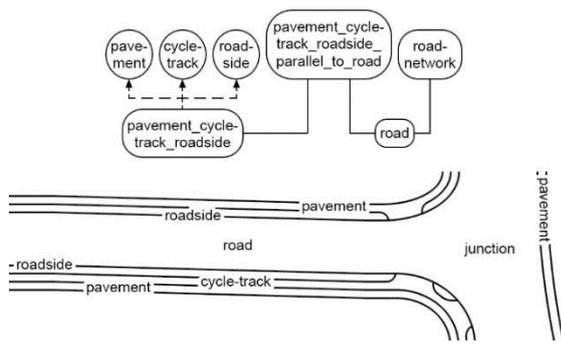


Abb. 1: Modellierung von Straße und parallelen Objekten in Form eines semantischen Netzes nach MAYER [1994b]. Pfeile implizieren Spezialisierung und für die durchgezogenen Linien sind die unteren Objekte Teile der oberen.

eine automatische Interpretation von Bildern absolut nicht ausreichend. Im Endeffekt wird angenommen, dass der menschliche Operateur im Laufe seines Lebens und seiner Tätigkeit gelernt hat, was die Objekte ausmacht. Der Objektartenkatalog spezifiziert, basierend auf diesem Wissen, nur Details der Erfassung.

3. Modellierung von Straßen in Luft- und Satellitenbildern

Zu Beginn erfolgte die Modellierung für die Straßenextraktion aus Luftbildern wiederum auf Grundlage von semantischen Netzen, allerdings nurmehr auf „konzeptueller“ Ebene, d.h. zur Strukturierung des Wissens. Die Implementierung erfolgte hingegen vor allem in C++.

Der Schwerpunkt der Modellierung für Straßen in Luftbildern [MAYER et al., 1997] lag auf deren Maßstabsverhalten und Kontext (siehe Abb. 2). Bezüglich Maßstabsverhalten führten Experimente auf Grundlage der Maßstabsraum-(Scale-Space)-Theorie [LINDBERG, 1994] zu der Erkenntnis, dass sich Straßen in unterschiedlichen Maßstäben/Auflösungen verschieden verhalten. Während bei ca. 2 m Auflösung eine Straße einer Linie entspricht, ist sie bei einer Auflösung von 25 cm eine langgestreckte Fläche. Bei der feineren Auflösung werden Markierungen sichtbar, die wiederum Linien entsprechen. In der in Abb. 2 oben gezeigten Modellierung wird neben dem Maßstabsverhalten noch explizit sichtbar gemacht, dass die Erscheinung eines Objektes im Bild von dessen physikalischer Beschaffenheit (Geometrie und Material) abhängt. Damit wird es möglich, Modelle zu erstellen, die sich (zumindest

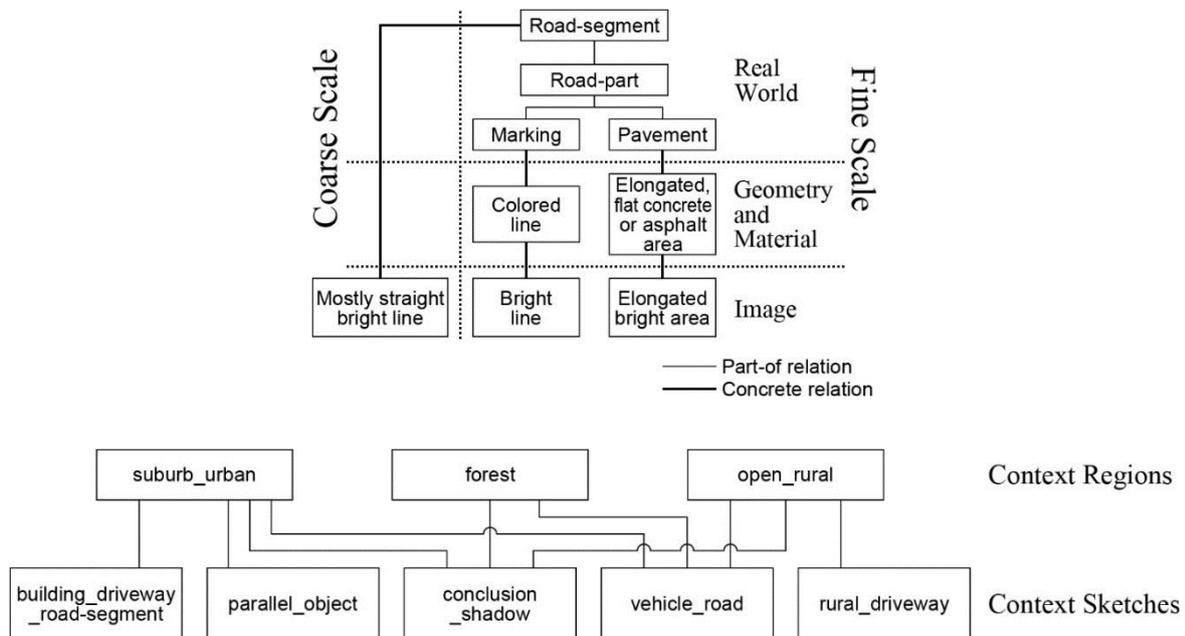


Abb. 2: Modellierung von Straßen in Bildern nach MAYER et al. [1997].
 Oben: Maßstabsverhalten und Unterscheidung von Abbildung im Bild und physikalischen Eigenschaften (Geometrie und Material).
 Unten: Kontext von Straßen in Bildern in verschiedenen Kontextregionen.

konzeptionell) für sehr verschiedene Wellenlängen und Abbildungsprinzipien, wie z.B. perspektive Abbildung von sichtbarem Licht und Synthetisches Apertur Radar (SAR), eignen.

Ein weiterer wichtiger Aspekt ist in Abb. 2 unten dargestellt: Objekte sind in vielen Fällen für sich alleine nicht eindeutig erkennbar – erst ihre Umgebung (Kontext) ermöglicht eine eindeutige Erkennung. Typische Fälle werden mit Kontextskizzen beschrieben. In vielen Fällen ist allerdings der Kontext eines Objektes wiederum von der Umgebung der Kontextskizze abhängig. So gibt es im Wald und in der offenen Landschaft meist keine parallel ausgerichteten Objekte wie Gebäude und Einfahrten zu Feldern gibt es nur in der offenen Landschaft. All diese Informationen können für die Interpretation genutzt werden und werden in den Kontextregionen zusammengefasst.

In der Straßenextraktion in hoch aufgelösten Satellitenbildern [BACHER/MAYER, 2005; BACHER, 2006] werden zur Modellierung vor allem zwei Eigenschaften von Straßen verwendet:

- Straßen verbinden bedeutende Örtlichkeiten (wenn möglich) effizient miteinander,
- Straßen besitzen parallele Ränder.

Die erste Eigenschaft wird mittels des im das Straßennetz abbildenden Graphen definierten Umwegfaktors [BACHER, 2006] formalisiert. Dieser drückt das Verhältnis zwischen der im Moment realen

Netzwerkentfernung und der durch Schließen einer Lücke möglichen (optimalen) Entfernung aus. Für Verbindungen mit großen Umwegfaktoren wird zuerst verifiziert, ob die Lücken geschlossen werden können:

$$\text{Umwegfaktor} = \frac{\text{Netzwerkentfernung}}{\text{optimale Entfernung}}$$

Die zweite Eigenschaft wird gewinnbringend für die automatische Generierung von Trainingsgebieten für die multispektrale Klassifikation genutzt. Diese werden aus parallelen Kanten mit einer hellen Linie in der Mitte gebildet (siehe Abb. 3). Die hohe Spezifität dieser Eigenschaft ermöglicht es, dass auf eine aufwändige manuelle Definition der Trainingsgebiete weitestgehend verzichtet werden kann.

4. Abstraktion durch Maßstabsräume

Die grundlegenden Erkenntnisse zum Maßstabsverhalten bei der Straßenextraktion inspirierten eine Nutzung in anderen Bereichen. Insbesondere erschienen Ereignisse im Maßstabsraum geeignet, um abstrakte(re) Repräsentationen von Objekten zu generieren, wie sie für die Generalisierung von Karten und GIS-Objekten notwendig sind [MAYER, 1998]. Abb. 4 zeigt, wie mittels morphologischem Opening und Closing die Topologie von Objekten bzw. Ob-

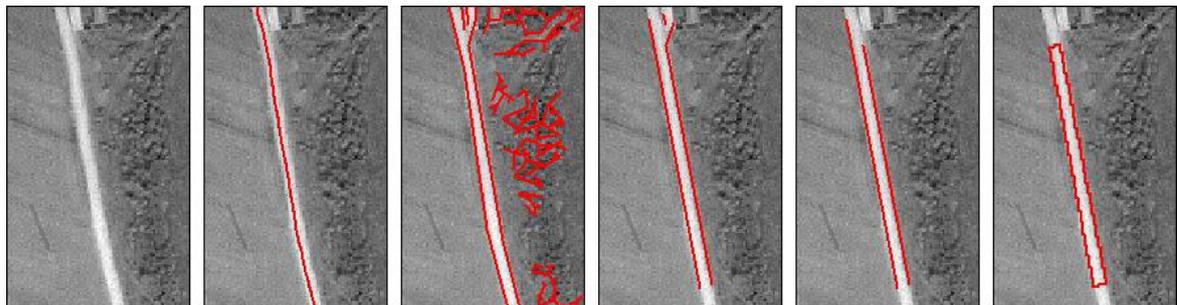


Abb. 3: Extraktion von Trainingsgebieten für die Straßenextraktion nach BACHER [2006]. Von links nach rechts: Bild, helle Linie, Kanten, Kanten in Puffer um Linie, parallele Kanten, Trainingsgebiet in Form von überlappenden parallelen Kanten auf beiden Seiten einer hellen Linie.



Abb. 4: Ereignisse bei morphologischem Opening und Closing aus MAYER [1998]. Während beim Opening eine schmale Verbindung zwischen zwei Objekten eliminiert wird, verschmelzen beim Closing zwei bisher getrennte Objekte miteinander bzw. es wird ein Ring geschlossen.

jektgruppen verändert werden kann. Beim Opening werden einzelne Objekte in mehrere aufgetrennt und beim Closing mehrere Objekte zu einem Objekt verschmolzen oder es wird ein Ring geschlossen. Die entwickelten Prinzipien wurden auch in 3D übertragen [FORBERG/MAYER, 2006].

5. Aussehensbasierte Modellierung

Die in MAYER/REZNIK [2006] und REZNIK/MAYER [2008] beschriebene 3D-Fassadeninterpretation basiert auf dem teilebasierten Ansatz zur aussehensbasierten Objektextraktion „Implicit Shape Model“ von LEIBE et al. [2004]. Aussehensbasierte Ansätze beruhen auf der Erkenntnis, dass Merkmale in Form von Kanten, Linien oder Punkten feinere Strukturen von Objekten nicht detailliert genug beschreiben. Auf der anderen Seite ist eine Repräsentation eines Objektes in Form eines großen Bildausschnittes zu unflexibel. Aussehensbasierte Ansätze verbinden daher kleine Bildausschnitte mit deren geometrischer Anordnung. Spezifisch bedeutet das beim Implicit Shape Model im Falle einer horizontal ausgerichteten Fassade mit normierter Auflösung, dass Punkte automatisch extrahiert werden (siehe Abb. 5). Kleine Bildausschnitte um die Punkte herum wer-

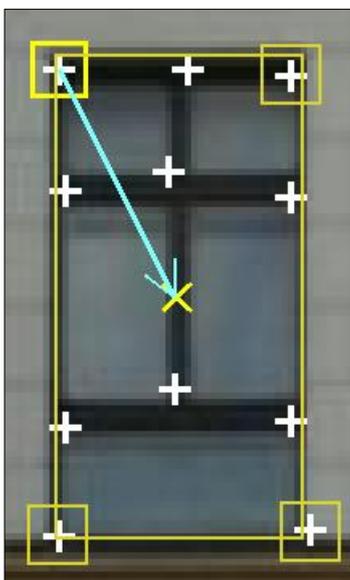


Abb. 5: Modellierung von Objekt (Fenster) in MAYER/REZNIK [2006] auf Grundlage eines Implicit Shape Models [LEIBE et al., 2004], bei dem die Umgebungen von automatisch detektierten Punktmerkmalen (gelbe Quadrate) in geometrischen Bezug zum vorgegebenen Mittelpunkt des Objektes gesetzt werden.

den zusammen mit dem relativen Vektor zwischen Punkt und Mittelpunkt des Objektes (z.B. Fenster oder Tür) abgespeichert. Hiermit wird das detaillierte Aussehen von punktförmigen Strukturen für die Objekte inklusive ihrem geometrischen Verhältnis zum Objektmittelpunkt modelliert.

6. Statistisch generative Modellierung

Generative Modellierung impliziert, dass eine Simulation des Objektes möglich ist. Da reale Objekte dreidimensional (3D) sind, folgt daraus auch, dass die Objekte räumlich modelliert werden müssen. Statistisch bedeutet, dass die Modellierung nicht deterministisch erfolgt, sondern im Rahmen einer vorgegebenen statistischen Verteilung variiert.

Für die statistisch generative Extraktion von unbelaubten Laubbäumen [HUANG/MAYER, 2009] wird ein Lindenmayer (L-) System [MĚCH/PRUSINKIEWICZ, 1996] aus der Computergraphik verwendet. Mit diesem wird die generative Verzweigungsstruktur von Bäumen modelliert (siehe Abb. 6). In den hierbei verwendeten Produktionsregeln steht die Variable F für Wachstum, d.h. für ein neues Zylinderstück. Die Konstanten beschreiben die Rotation um spezifische Achsen („+“ und „-“ bedeuten Neigung nach oben oder unten, „<“ und „>“ Azimut nach links und rechts) und die Erzeugung von neuen Ästen eingeschlossen in „[“ und „]“. Eine Produktionsregel gibt vor, wie F in der nächsten Iteration zu ersetzen ist. Abb. 6 zeigt, dass unterschiedliche Produktionsregeln zu verschiedenen Verzweigungsstrukturen führen.

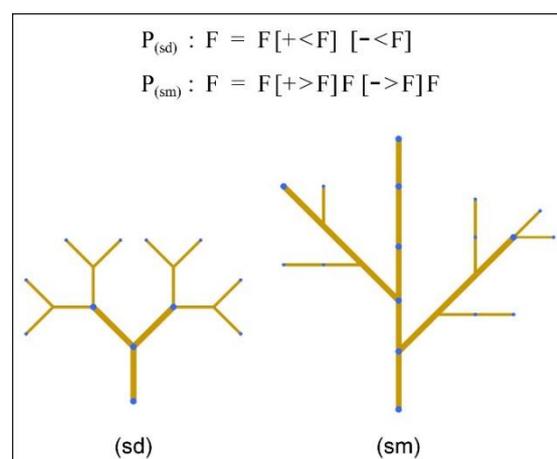


Abb. 6: Verzweigungsstrukturen von Bäumen aus HUANG [2008] – (sd) = sympodial-dichasium, (sm) = sympodial-monochasium. Die Produktionsregeln sind oben, die graphischen Darstellungen unten wiedergegeben.

Der statistische Anteil besteht grundlegend aus Prior-Verteilungen für die Parameter, wie z. B. Neigung, Ausrichtung, Dicke oder Länge eines Astes. Aufbauend auf dem gegebenen Stamm werden auf Basis von L-System und Prior-Verteilungen virtuelle 3D-Modelle für den Baum erzeugt.

Die grundlegende Idee der Interpretation von Bildern auf dieser Grundlage besteht darin, virtuelles und reales Bild zu vergleichen und damit eine (statistische) Likelihood zu schätzen. Dies ist umso erfolgversprechender, je realistischer das virtuelle Bild ist. Während in *HUANG [2008]* statistisches Rauschen verwendet wurde, was zu einer sehr beschränkten Diskriminativität führte, wurde in *HUANG/MAYER [2009]* der Hintergrund mit Bildverarbeitungsverfahren geschätzt, wodurch gute und schlechte Hypothesen deutlich besser trennbar wurden.

Die geschätzte Likelihood führt zusammen mit den Prior-Verteilungen für die Parameter zu einer Maximum-a-Posteriori (MAP) Schätzung. Im vorliegenden Ansatz wird zur statistischen Variation des Modells eine Kombination von Monte Carlo und Markov Chain Monte Carlo (MCMC) Stichprobengenerierung verwendet. Das heißt, neben einer rein zufälligen, den Prior-Verteilungen folgenden Parametergenerierung (Monte Carlo) wird bei Vorliegen eines relativ zu anderen guten Ergebnisses (d. h. MAP Schätzung) diese lokal in einer Nachbarschaft des Ergebnisses verfeinert (MCMC). Darüber hinaus

werden nicht einfach zufällig Bäume überall im Raum generiert, sondern die Hypothesen basieren auf diskriminativ aus mehreren Bildern bestimmten 3D-Stämmen in Form von Zylindern. Dies kann als vereinfachte Variante der in *TU et al. [2005]* vorgeschlagenen Kombination von generativer und diskriminativer statistischer Modellierung gesehen werden.

Durch die zufällige Erzeugung von Stichproben variieren die Ergebnisse. Solange diese allerdings innerhalb der Variabilität der menschlichen Erfassung liegen, ist dies im Normalfall akzeptabel.

7. 3D-Modellierung von Gebäuden

In *MAYER et al. [2012]* wurde ein System zur Orientierung von Bildern bei bekannter gegenseitiger Überlappung vorgestellt, das in *MICHELINI/MAYER [2016]* auf die Nutzung von unsortierten Bildmenüen erweitert wurde. Dies bildet die Grundlage für den Ansatz zur dichten 3D-Rekonstruktion [*KUHN et al., 2014; 2016*], der auf dem semiglobalen Matching von *HIRSCHMÜLLER [2008]* beruht. Die sich ergebenden hoch detaillierten 3D-Modelle bildeten die Eingabedaten für *NGUATEM/MAYER [2016; 2017]* zur 3D-Rekonstruktion von Gebäuden.

Der Kern der Modellierung in *NGUATEM/MAYER [2016]* besteht in zusammenhängenden (contiguous) Patches (siehe Abb. 7). Hierbei werden

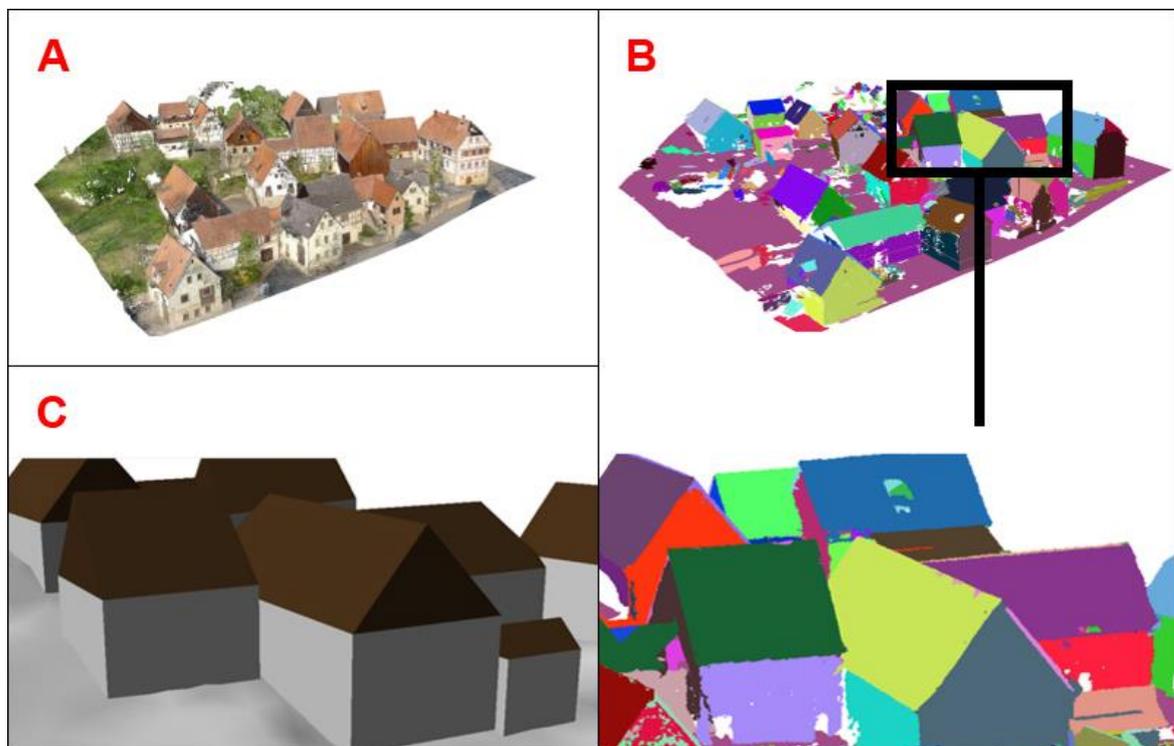


Abb. 7: Modellierung von Gebäuden in Punktwolken (A) auf Grundlage von zusammenhängenden Patches (B), die zu symmetrisch geneigten Dachstrukturen (C) führen.

in Voxeln vorgegebener Größe mittels RANSAC [FISCHLER/BOLLES, 1981] Ebenen angepasst. Aus den Normalen der Ebenen werden mittels statistischer Inferenz auf Grundlage von Dirichlet Process Mixture (DPM) und Zusammenhangskomponentenanalyse die zusammenhängenden (und ähnlich orientierten) Patches abgeleitet. Letztere sind in NGUATEM/MAYER [2017] die Grundlage, um bei gegebener Vertikale die als symmetrisch geneigte Flächen modellierten Dachflächen zu finden.

In NGUATEM et al. [2014] werden aufbauend auf als ungefähr bekannt angenommenen Fassadenebenen die Fenster und Türen in dichten 3D-Punktwolken aus Bildzuordnung als „Löcher“ in der Fassade modelliert (s. Abb. 8). Weil die Fassadenebene nur ungefähr bekannt ist, wird diese variiert, und für die verschiedenen Varianten werden jeweils Fenster und Türen bestimmt, die dann fusioniert werden. Als Modell wird ein Rechteck verwendet, wobei der obere Teil durch verschiedene Typen von Bögen (z. B. romanisch oder gotisch) ersetzt werden kann. Die Modelle werden an Bereiche mit starken Gradienten in der Tiefe bezüglich der Fassade angepasst. Hierzu wird ein Monte Carlo Ansatz inkl. Modellauswahl verwendet. Über die bekannte 3D-Lage der Fassade sind am Ende auch die 3D-Positionen von Fenstern und Türen bekannt.

8. Deep Learning

Neuronale Netze waren lange Zeit nur in speziellen Anwendungen verbreitet, da für das Training der häufig voll verbundenen Schichten für größere Probleme extreme Rechenleistungen und vor allem riesige Mengen an Trainingsdaten notwendig sind. Diese Situation hat sich mit der Einführung der Convolutional (Neural) Networks (CN) geändert. Hierbei werden die Verbindungen zwischen den Schichten in Analogie zu einer Faltung (Convolution) wiederholt. Hiermit reduziert sich die Zahl der notwen-

digen Parameter drastisch. Weitere Entwicklungen, die das Training großer CN möglich machen, sind die ideal für CN geeignete massive Rechenleistung hochgradig parallel arbeitender Graphikkarten (GPUs) und die „Augmentation“, bei der zusätzliche Trainingsdaten durch Modifikation der gegebenen Trainingsdaten, z. B. in Form von Spiegelung oder Homographien, erzeugt werden. Dies alles führt zu extrem leistungsfähigen Systemen [KRIZHEVSKY et al., 2012], die in einigen Bereichen den Stand der Technik mit einer deutlich besseren Leistungsfähigkeit als alle anderen Modellierungstechniken bilden. Anzumerken ist auch, dass es oft nicht notwendig ist, das ganze Netz zu trainieren, sondern vor allem für die ersten Schichten auf Standarddaten, wie z. B. Imagenet, trainierte Schichten [KRIZHEVSKY et al., 2012] übernommen werden können, was den Rechenaufwand und die Menge der notwendigen Trainingsdaten stark reduziert.

Das CN in SCHMITZ/MAYER [2016] wird zur Interpretation von Fassadenbildern genutzt, aus denen die Fassadenfläche, Fenster, Türen und Sonstiges extrahiert werden. Die verwendete Architektur (s. Abb. 9) reduziert zuerst die gegebenen Bildaten mit Convolutions in der Größe. Hiermit wird eine Interaktion in einer größeren Umgebung ermöglicht. Die damit erzielte Klassifikation wird anschließend mit einer inversen Convolution (Deconvolution) in Kombination mit Information aus den in der Größe unterschiedlich stark reduzierten ersten Schichten wieder in der Bildauflösung verbessert. In dem so trainierten Netzwerk werden das Aussehen der Objekte und ihre gegenseitige Anordnung implizit durch die Gewichte des Netzes modelliert.

9. Zusammenfassung, Analyse und Ausblick

Seit der gemeinsamen Zeit mit WOLFGANG REINHARDT bei Siemens bzw. Siemens Nixdorf hat sich

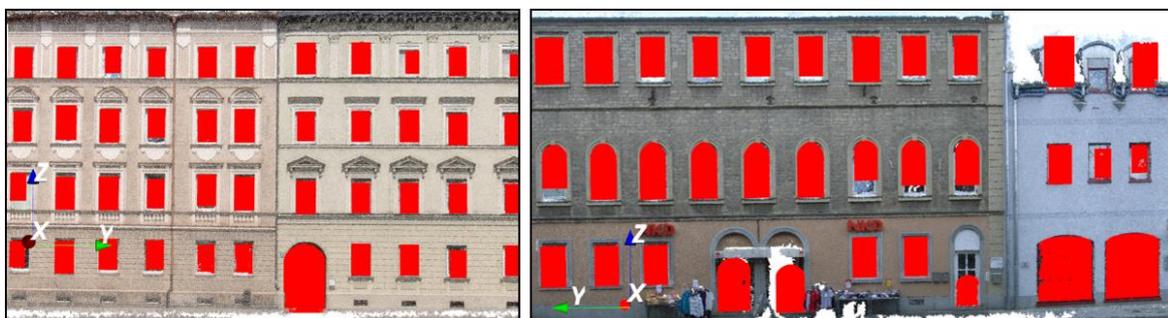


Abb. 8: Fenster und Türen als „Löcher“ in der Fassade [NGUATEM et al, 2014], die grundlegend als Rechteck modelliert werden. Der obere Teil des Rechtecks kann durch verschiedene Typen von Bögen (z. B. romanisch oder gotisch) ersetzt werden.

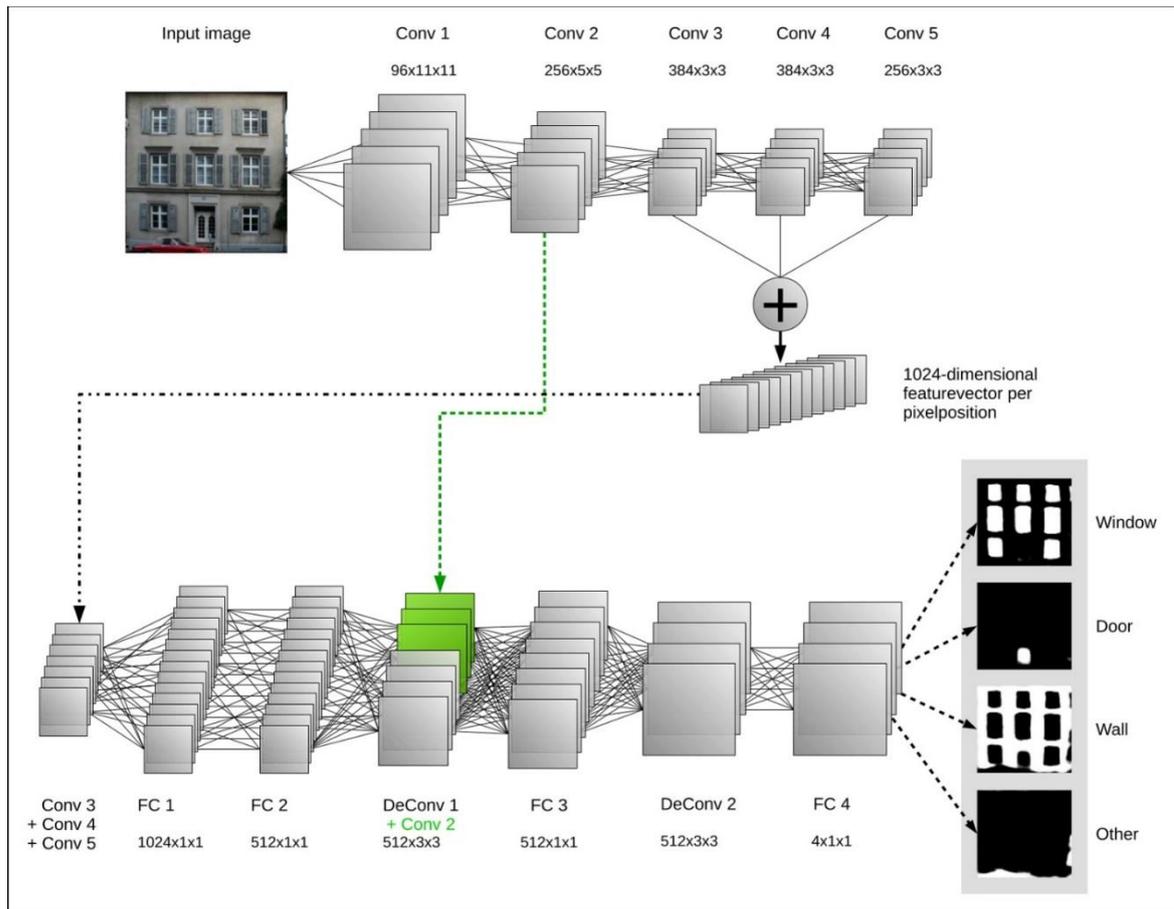


Abb. 9: Das Convolutional Network aus SCHMITZ/MAYER [2016] verwendet eine implizite Modellierung von Fenstern, Türen, Wänden und anderen Objekten über die Gewichte, die die einzelnen Schichten entweder komplett (Fully Connected – FC) oder in Form von Convolutions (Conv) bzw. Deconvolutions (DeConv) verbinden.

in der Modellierung vieles geändert. Am Anfang standen symbolische/strukturelle Methoden. Seit Mitte der 90er Jahre lag ein Schwerpunkt der Arbeiten auf dem Maßstabsverhalten von Objekten inkl. Maßstabsereignissen. Ab der Jahrtausendwende begann die Beschäftigung mit statistischen Verfahren. Durch die parallel entwickelten Methoden zur Orientierung von Bildern und zur dichten 3D-Rekonstruktion wurde die Modellierung von 3D-Objekten wie Gebäuden und Bäumen möglich. Die neuesten Arbeiten beschäftigen sich mit impliziten Convolutional Networks zur Interpretation von Fassaden.

Insgesamt sind verschiedene Entwicklungen feststellbar:

- Die Modellierung war anfangs deterministisch und wurde dann mehr und mehr statistisch. Insbesondere spielen bei der Automatisierung, bei der große Mengen grob falscher Daten (Ausreißer) entstehen, extrem robuste Verfahren wie RANSAC [FISCHLER/BOLLES, 1981] eine große Rolle. Daneben sind aber auch die Modelle un-

sicher und somit statistische Samplingverfahren wie Markov Chain Monte Carlo (MCMC) von großer Bedeutung. Auch wenn Convolutional Networks meist explizit nicht statistisch sind, so bilden sie dennoch implizit die Statistik der Daten und Modelle ab.

- Am Anfang erfolgte die Modellierung sehr explizit in Form von für die Schlussfolgerung implementierten semantischen Netzen. Die später verwendeten konzeptionellen Modelle sind bereits deutlich impliziter und im letzten Jahrzehnt wurden zwar noch Konzepte/Objekte wie Gebäude, Straßen, Fenster, Türen und deren Relationen angesprochen, aber die Modellierung erfolgte mehr oder weniger implizit geometrisch und in großen Teilen statistisch.
- Ein Fokus der Arbeiten des Autors lag am Anfang auf dem Maßstabsverhalten der Objekte und später auf der 3D-Modellierung. Beide ermöglichen eine deutlich realistischere Modellierung der Welt und damit bessere Ergebnisse.

- Während Bildinformation am Anfang in die symbolische Form von Punkten, Kanten oder Segmenten abstrahiert wurde, begann mit der aussehensbasierten Modellierung die direkte Nutzung der detaillierten Bildinformation, die sich mit den Convolutional Networks fortsetzt.
- Die Modelle wurden am Anfang in Form von Parametern und Schwellwerten vorgegeben. Mit den Arbeiten zur aussehensbasierten Modellierung begann das Lernen von Teilen der Modelle. Bei Deep Learning liegt jetzt der Schwerpunkt auf dem Lernen.

Deep Learning ist der momentane Trend im Bereich der Bildinterpretation. Die Verbesserung der Ergebnisse ist z. T. beachtlich. Trotzdem bleibt eine Reihe von Fragen offen. Spezifische Fragen sind z. B., wie eine gemeinsame Analyse mehrerer Bilder mit 3D-Ergebnissen auf ein Neuronales Netz abgebildet werden kann? Eine allgemeine Frage besteht darin, wie weit ein vollkommen generisches Netz in der Lage ist, komplexe Probleme zu lösen? Die neuronalen Strukturen bei Tieren und Menschen sind vorgegeben und das Lernen erfolgt aufbauend auf diesen. Fragen bestehen auch dahingehend, ob die im Vergleich zum biologischen Vorbild sehr einfachen Verknüpfungen ausreichend sind, um eine ähnliche Leistungsfähigkeit zu erzielen? Und zuletzt spielt bei praktisch, z. B. in Fahrzeugen zur automatischen Steuerung, eingesetzter Software auch die Frage nach der Nachvollziehbarkeit von Entscheidungen eine u. U. ausschlaggebende Rolle. Die Begründung, dass das Convolutional Network sich nach dem Training einfach so verhält, wird einen Geschädigten eines durch ein autonomes System verursachten Unfalls wahrscheinlich nicht befriedigen.

Literatur / Quellen

- BACHER, Uwe [2006]:* Automatische Straßenextraktion aus optischen Satellitenbildern hoher und höchster Auflösung. Dissertation. Universität der Bundeswehr München, Neubiberg, 109 S.
<http://athene-forschung.unibw.de/node?id=86111> – letzter Aufruf: 06.11.2017
- BACHER, Uwe / MAYER, Helmut [2005]:* Automatic Road Extraction from Multispectral High Resolution Satellite Images. – In: *STILLA, Uwe / ROTTENSTEINER, Franz / HINZ, Stefan (Hrsg.):* Proceedings of the ISPRS Workshop CMRT 2005: Object Extraction for 3D City Models, Road Databases and Traffic Monitoring – Concepts, Algorithms and Evaluation, 29-30 Aug. 2005, Vienna, Austria. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI-3/W24, S. 29-34.
ISSN: 1682-1750 (Druck)
ISSN: 1682-1777 (CD-ROM)
http://www.isprs.org/proceedings/XXXVI/3-W24/papers/CMRT05_Bacher_Mayer.pdf – letzter Aufruf: 06.11.2017
- FISCHLER, Martin A. / BOLLES, Robert C. [1981]:* Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. – In: *Communications of the ACM*, Vol. 24, No. 6, S. 381-395.
ISSN: 0001-0782
<https://www.sri.com/sites/default/files/publications/ransac-publication.pdf> – letzter Aufruf: 06.11.2017
- FORBERG, Andrea / MAYER, Helmut [2006]:* Simplification of 3D Building Data. – In: *Zeitschrift für Vermessungswesen (zfv)*, 131. Jhrg., Heft 3, S. 148-152.
ISSN: 0044-3689
- HIRSCHMÜLLER, Heiko [2008]:* Stereo Processing by Semiglobal Matching and Mutual Information. – In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, Issue 2, S. 328-341.
ISSN: 0162-8828
- HUANG, Hai [2008]:* Terrestrial Image Based 3D Extraction of Urban Unfoliated Trees of Different Branching Types. – In: *CHEN, Jun / JIANG, Jie / FÖRSTNER, Wolfgang (Hrsg.):* Proceedings of the XXII ISPRS Congress, Technical Commission III, July 3-11, 2008, Beijing, China. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII-B3a, S. 253-258.
ISSN: 1682-1750 (Druck)
ISSN: 1682-1777 (CD-ROM)
http://www.isprs.org/proceedings/XXXVII/congress/3_pdf/39.pdf – letzter Aufruf: 07.11.2017
- HUANG, Hai / MAYER, Helmut [2009]:* Generative statistical 3D reconstruction of unfoliated trees from terrestrial images. – In: *Annals of GIS*, Vol. 15, Issue 2, S. 97-105.
ISSN: 1947-5683
- KRIZHEVSKY, Alex / SUTSKEVER, Ilya / HINTON, Geoffrey E. [2012]:* ImageNet Classification with Deep Convolutional Neural Networks. – In: *Advances in Neural Information Processing Systems (NIPS 2012)*, Vol. 25, S. 1097-1105.
<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> – letzter Aufruf: 07.11.2017
- KUHN, Andreas / MAYER, Helmut / HIRSCHMÜLLER, Heiko / SCHARSTEIN, Daniel [2014]:* A TV Prior for High-Quality Local Multi-View Stereo Reconstruction. – In: *Proceedings, 2014 2nd International Conference on 3D Vision (3DV 2014)*, The University of Tokyo, Tokyo, Japan, 8-11 Dezember 2014, Vol. 1. IEEE Computer Society, Los Alamitos, S. 65-72.
ISSN: 1550-6185
eISBN: 978-1-4799-7001-8
- KUHN, Andreas / HIRSCHMÜLLER, Heiko / SCHARSTEIN, Daniel / MAYER, Helmut [2016]:* A TV Prior for High-Quality Scalable Multi-View Stereo Reconstruction. – In: *International Journal of Computer Vision*, Vol. 124, Issue 1, S. 2-17.
ISSN: 0920-5691
eISSN: 1573-1405
- LEIBE, Bastian / LEONARDIS, Ales / SCHIELE, Bernt [2004]:* Combined Object Categorization and Segmentation with an Implicit Shape Model. – In: *ECCV'04 Workshop on Statistical Learning in Computer Vision*, May 15, 2004, Czech Technical University, Prague, S. 1-16.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.6272&rep=rep1&type=pdf> – letzter Aufruf: 07.11.2017
- LINDBERG, Tony [1994]:* Scale-Space Theory in Computer Vision. The Kluwer International Series in Engineering and

- Computer Science (SECS), Vol. 256: Robotics: Vision, manipulation and sensors. Kluwer Academic Publishers, Boston/London/Dordrecht, XII, 423 S.
 ISBN: 0-7923-9418-6
 ISSN: 0893-3405
- MAYER, Helmut [1994a]: Automatische wissensbasierte Extraktion von semantischer Information aus gescannten Karten. Deutsche Geodätische Kommission, Reihe C, Nr. 417. Bayerische Akademie der Wissenschaften München / Beck in Komm., München, 113 S.
 ISBN: 3-7696-9461-9
- MAYER, Helmut [1994b]: Is the Knowledge in Map-legends and GIS-models Suitable for Image Understanding? – In: WELCH, Roy (Hrsg.): Proceedings of the Symposium Mapping and Geographic Information Systems, May 31 - June 3, 1994, Athens, Georgia, USA. International Archives of Photogrammetry and Remote Sensing, Vol. 30, Part 4, S. 52-59.
 ISSN: 0256-1840
- MAYER, Helmut [1998]: Model-Generalization of Building Outlines Based on Scale-Spaces and Scale-Space Events. – In: SCHENK, Toni / HABIB, Ayman (Hrsg.): ISPRS Commission III Symposium Object Recognition and Scene Classification from Multispectral and Multisensor Pixels, July 6-10, 1998, Columbus, Ohio. RICS Books, Coventry. International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 3/1, S. 530-536.
 ISSN: 0256-1840
- MAYER, Helmut / BALSAVIAS, Emmanuel / BACHER, Uwe [2006]: Automated Extraction, Refinement, and Update of Road Databases from Imagery and Other Data. – In: EuroSDR Projects. European Spatial Data Research (Euro SDR) Official Publication, No. 50, S. 217-280.
 ISSN: 0257-0505
- MAYER, Helmut / BARTELTSEN, Jan / HIRSCHMÜLLER, Heiko / KUHN, Andreas [2012]: Dense 3D Reconstruction from Wide Baseline Image Sets. – In: DELLAERT, Frank / FRAHM, Jan-Michael / POLLEFEYS, Marc / LEAL-TAIXÉ, Laura / ROSENHAHN, Bodo [Hrsg.]: Outdoor and Large-Scale Real-World Scene Analysis. 15th International Workshop on Theoretical Foundations of Computer Vision, Dagstuhl Castle, Germany, June 26 - July 1, 2011. Lectures Notes in Computer Science (LNCS), Band 7474. Springer, Heidelberg/Dordrecht/London/New York, S. 285-304.
 ISSN: 0302-9743
 eISSN: 1611-3349
 ISBN: 978-3-642-34090-1
 eISBN: 978-3-642-34091-8
- MAYER, Helmut / LAPTEV, Ivan / BAUMGARTNER, Albert / STEGER, Carsten [1997]: Automatic Road Extraction Based on Multi-scale Modeling, Context, and Snakes. – In: KRUPNIK, Amnon (Hrsg.): ISPRS WG III/2, III/3 and II/8 Joint Workshop on Theoretical and Practical Aspects of Surface Reconstruction and 3D Object Extraction, September 9-11, 1997, Haifa, Israel. International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 3/2-W3, S. 106-113.
 ISSN: 0256-1840
- MAYER, Helmut / REZNIK, Sergiy [2006]: MCMC Linked with Implicit Shape Models and Plane Sweeping for 3D Building Facade Interpretation in Image Sequences. – In: FÖRSTNER, Wolfgang (Hrsg.): Symposium on ISPRS Commission III Photogrammetric Computer Vision (PCV'06), Bonn, Germany, 20-22 September 2006. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36, Part 3, S. 130-135.
 ISSN: 1682-1750
- MĚCH, Radomír / PRUSINKIEWICZ, Przemyslaw [1996]: Visual Models of Plants Interacting with Their Environment. – In: FUJII, John (Hrsg.): SIGGRAPH'96, Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, USA, August 4-9, 1996. Association for Computer Machinery (ACM), New York, S. 397-410.
 ISBN: 0-89791-746-4
http://delivery.acm.org/10.1145/240000/237279/p397-mech.pdf?ip=137.193.150.146&id=237279&acc=ACTIVE%20SERVICE&key=2BA2C432AB83DA15%2E2A25F324ACF6B22F%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=826651855&CFTOKEN=14603738&__acm__=1510045341_a47b84f77c261f92884be650224b3d91 – letzter Aufruf: 07.11.2017
- MICHELINI, Mario / MAYER, Helmut [2016]: Efficient Wide Baseline Structure from Motion. – In: HALOUNOVÁ, Lena / SCHINDLER, Konrad / LIMPOUCH, Ales / PAJDLA, Tomas / ŠAFÁŘ, Václav / MAYER, Helmut / OUDE ELBERINK, Sander / MALLET, Clément / ROTTENSTEINER, Franz / BRÉDIF, Mathieu / SKALOUD, Jan / STILLA, Uwe (Hrsg.): XXIII ISPRS Congress, Commission III, 12-19 July 2016, Prague, Czech Republic. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. III-3, S. 99-106.
 ISSN: 2194-9042 (Druck)
 ISSN: 2196-6346 (CD-ROM)
 eISSN: 2194-9050 (Internet/USB)
<https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/III-3/99/2016/isprs-annals-III-3-99-2016.pdf> – letzter Aufruf: 07.11.2017
- NGUATEM, William / DRAUSCHKE, Martin / MAYER, Helmut [2014]: Localization of Windows and Doors in 3d Point Clouds of Facades. – In: SCHINDLER, Konrad / PAPANODITIS, Nicolas (Hrsg.): ISPRS Technical Commission III Symposium, 5-7 September 2014, Zurich, Switzerland. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-3, S. 87-94.
 ISSN: 2194-9042 (Druck)
 ISSN: 2196-6346 (CD-ROM)
 eISSN: 2194-9050 (Internet/USB)
<https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-3/87/2014/isprsannals-II-3-87-2014.pdf> – letzter Aufruf: 07.11.2017
- NGUATEM, William / MAYER, Helmut [2016]: Contiguous Patch Segmentation in Pointclouds. – In: ROSENHAHN, Bodo / ANDRES, Bjoern (Hrsg.): Pattern Recognition. 38th German Conference, GCPR 2016, Hannover, Germany, September 12-15, 2016, Proceedings. Lecture Notes in Computer Science (LNCS), Vol. 9796. Springer International Publishing, Cham, S. 131-142.
 ISSN: 0302-9743
 eISSN: 1611-3349
 ISBN: 978-3-319-45885-4
 eISBN: 978-3-319-45886-1
- NGUATEM, William / MAYER, Helmut [2017]: Modeling Urban Scenes From Pointclouds. – In: 2017 IEEE International Conference on Computer Vision (ICCV 2017), Venice, Italy, October 22-29, 2017. S. 3837-3846.
http://openaccess.thecvf.com/content_iccv_2017/html/Nguatem_Modeling_Urban_Scenes_ICCV_2017_paper.html – letzter Aufruf: 22.11.2017

NIEMANN, Heinrich / SAGERER, Gerhard / SCHRÖDER, Stefan / KUMMERT, Franz [1990]: ERNEST: A Semantic Network System for Pattern Understanding. – In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, Issue 9, S. 883-905.

ISSN: 0162-8828

REZNIK, Sergiy / MAYER, Helmut [2008]: Implicit Shape Models, Self-Diagnosis, and Model Selection for 3D Facade Interpretation. – In: Photogrammetrie – Fernerkundung – Geoinformation (PFG), Heft 3/2008, S. 187-196.

ISSN: 1432-8364

SCHMITZ, Matthias / MAYER, Helmut [2016]: A Convolutional Network for Semantic Facade Segmentation and Interpretation. – In: HALOUNOVÁ, Lena / SCHINDLER, Konrad / LIMPOUCH, Ales / PAJDLA, Tomas / ŠAFÁŘ, Václav / MAYER, Helmut / OUDE ELBERINK, Sander / MALLET, Clément / ROTTENSTEINER, Franz / BRÉDIF, Mathieu / SKALOUD, Jan / STILLA, Uwe (Hrsg.): XXIII ISPRS Congress, Commission III, 12-19 July 2016, Prague, Czech Republic. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLI-B3, S. 709-715.

ISSN: 1682-1750 (Druck)

ISSN: 2194-9034 (Internet / USB)

<https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B3/709/2016/isprs-archives-XLI-B3-709-2016.pdf> – letzter Aufruf: 07.11.2017

TU, Zhuowen / CHEN, Xiangrong / YUILLE, Allen L. / ZHU, Song-Chun [2005]: Image Parsing: Unifying Segmentation, Detection and Recognition. – In: International Journal of Computer Vision, Vol. 63, Issue 2, S. 113-140.

ISSN: 0920-5691

eISSN: 1573-1405

Anschrift des Verfassers:

Univ.-Prof. Dr.-Ing. habil. HELMUT MAYER

Universität der Bundeswehr München

Fakultät für Informatik

Institut für Angewandte Informatik

Professur für Visual Computing

D-85577 Neubiberg

Tel.: +49 / (0)89 / 6004 - 3429

E-Mail: Helmut.Mayer@unibw.de