Klassifikation von Stereobildern aus Mobile Mapping Daten mittels Conditional Random Fields

MAX COENEN[1] & FRANZ ROTTENSTEINER[1]

Zusammenfassung: In dieser Arbeit wird ein neues Verfahren zur kontextbasierten Klassifikation von Stereobildern mittels Conditional Random Fields (CRF) vorgestellt. Die Klassifikation setzt auf Segmenten als Knoten für das CRF auf. Die Segmentierung erfolgt im Bildraum und wird mittels einer 3D-Rekonstruktion der Szene auf die 3D-Punktwolke übertragen, was die Extraktion von 3D-Merkmalen zusätzlich zu den Bildmerkmalen sowie die Definition von realistischen Nachbarschaftsbeziehungen zwischen den Segmenten im Objektraum ermöglicht. Die Evaluierung der Methode erfolgt anhand von im urbanen Raum aufgenommenen Stereosequenzen eines Benchmark Datensatzes und liefert Ergebnisse mit einer Gesamtgenauigkeit von bis zu über 90%. Außerdem wird gezeigt, dass die Berücksichtigung von Kontext in der Klassifikation zu einer Erhöhung der Gesamtgenauigkeiten führt.

1 Einleitung

Die Nutzung von mobilen Systemen zur Erfassung von 3D-Geodaten über berührungslos aufnehmende Sensoren gewinnt v.a. im urbanen Raum zunehmend an Bedeutung. Zum Einsatz kommen dabei in der Regel hohe Punktdichten erzeugende Sensoren wie 3D-Laserscanner oder Stereokameras, wobei letztere zusätzlich zu Tiefeninformationen auch Farbinformationen liefern und deutlich kostengünstiger sind. Für viele Anwendungen, wie z.B. zur Generierung von 3D-Stadtmodellen oder in der Fahrzeug- und Roboternavigation, ist eine Klassifikation der 3D-Daten, also eine Zuweisung von semantischen Objektklassen an Teile der Szene, essentiell. So nutzen ANGULOV et al. (2005) zum Beispiel geometrische 3D-Punkt-Merkmale für die punktweise Klassifikation einer Lasercan-Punktwolke, während MATTI & NEBIKER (2014) zusätzlich zu geometrischen Merkmalen auch Farbinformation zur Klassifikation einer texturierten Laserscan-Punktwolke verwenden. SENGUPTA et al. (2013) führen eine pixelweise Klassifikation von Stereobildern aus Mobile Mapping Daten durch, wofür sie ausschließlich Bildinformationen als Merkmale nutzen und übertragen das Klassifikationsergebnis auf die aus den Stereobildern rekonstruierte 3D-Punktwolke. Um die Rechenkomplexität von punkt- bzw. pixelweise klassifizierenden Methoden zu verringern, kann eine generalisierende Datenreduktion, z.B. in Form von einer Segmentierung der Daten, sinnvoll sein, wonach die Segmente anstelle der einzelnen Punkte bzw. Pixel klassifiziert werden. So segmentieren LIM & SUTER (2009) beispielsweise terrestrisch aufgenommene Punktwolken in sogenannte Supervoxel, welche sie anhand von für jedes Supervoxel berechneten Merkmalen klassifizieren.
Insbesondere im urbanen Raum ist die Klassifikation aufgrund von komplexen und vielfältigen Strukturen und Objektklassen eine Herausforderung. Doch gerade in von Menschen gemachten Umgebungen weisen Objekte spezifische räumliche Beziehungen zueinander auf, welche als

[1] Leibniz Universität Hannover, Institut für Photogrammetrie und GeoInformation, Nienburger Str. 1, D-30167 Hannover, E-Mail: [coenen, rottensteiner]@ipi.uni-hannover.de

Kontextinformation mithilfe sogenannter Graphischer Modelle in der Klassifikation berücksichtigt werden können. Ziel dieser Arbeit ist es, einen solchen kontextbasierten Ansatz für die Klassifikation von aus einem bewegten Fahrzeug aus aufgenommenen Stereobildern bzw. daraus abgeleiteten 3D Punktwolken auf Basis von Conditional Random Fields (CRF) (KUMAR & HEBERT 2006), in welchen die statistischen Abhängigkeiten zwischen benachbarten Objekten explizit modelliert werden, zu entwickeln und zu testen.

2 Methodik

Das Ziel der Klassifikation besteht darin, jedem 3D-Punkt einer aus Stereobildern abgeleiteten 3D-Punktwolke ein Klassenlabel aus einer Menge à priori definierter Objektklassen zuzuweisen. Zu diesem Zweck wurde ein segmentbasiertes Klassifikationskonzept entwickelt, welches aus Stereobildsequenzen generierte 3D-Punktwolkensegmente klassifiziert. Hierbei werden die Stereobildpaare einzeln betrachtet und jedes Stereomodell separat klassifiziert. Der Ablauf der Klassifikation ist in Abb. 1 dargestellt und gliedert sich in mehrere Schritte, auf welche im Folgenden näher eingegangen wird.

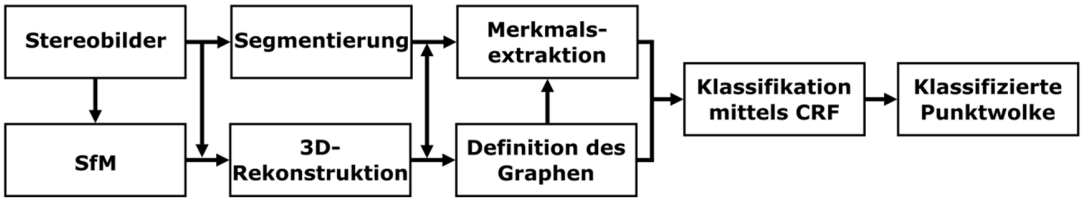

Abb. 1: Übersicht über das Konzept der Klassifikationsstrategie

2.1 3D-Rekonstruktion

Das Ziel der 3D-Rekonstruktion ist die Generierung von 3D-Punktwolken aus den Stereobildern in einem für die gesamte Stereosequenz einheitlichen Koordinatensystem, welche die Berechnung von zusätzlichen und für die Klassifikation sehr wertvollen geometrischen 3D-Merkmalen ermöglichen. Ausgehend von bekannten Parametern der inneren Orientierung und der relativen Orientierung der beiden Kameras wird zunächst mittels eines *Structure from Motion* (SfM) Verfahrens ähnlich zu (REICH et al. 2013) die äußere Orientierung aller Bilder der Stereosequenz ermittelt. Anschließend wird mithilfe des *Efficient Large-Scale Stereo Matching* (ELAS-) Verfahrens (GEIGER et al. 2011) ein dichtes Parallaxenfeld bestimmt, aus dem für jedes Pixel des linken Stereopartners mittels Triangulation unter Berücksichtigung der äußeren Orientierung ein 3D-Punkt im Objektkoordinatensystem rekonstruiert wird. Man erhält so für jedes Stereopaar eine dichte 3D-Punktwolke, die die Grundlage für die weitere Verarbeitung darstellt.

2.2 Datenreduktion durch Segmentierung

Die große Anzahl an Punkten, welche durch die dichte Bildzuordnung generiert wird, würde sich negativ auf die Rechenkomplexität der kontextbasierten Klassifikation auswirken. Um dem entgegenzuwirken und weil Segmente stabilere Merkmale als einzelne Punkte oder Pixel liefern können, wird eine Segmentierung der Daten durchgeführt. In den darauffolgenden Schritten

werden nicht die Einzelpunkte, sondern die Segmente klassifiziert. Aufgrund der geringeren Rechenkomplexität im Vergleich zu Verfahren zur Segmentierung von Punktwolken erfolgt die Segmentierung zunächst in einem der Bilder (Referenzbild, in dieser Arbeit der linke Stereopartner). Ihr Ergebnis wird auf die 3D Punktwolke übertragen, indem jeder 3D-Punkt dem Segment jenes Pixels im Referenzbild zugeordnet wird, das zur Bestimmung seiner Koordinaten beitrug. Zur Segmentierung wird das *Simple Linear Iterative Clustering* (SLIC-) Verfahren (ACHANTA et al. 2012) verwendet, welches die Pixel des Bildes in Regionen, sogenannte Superpixel, gruppiert, wobei sich die Segmentgröße als Anzahl der Pixel pro Segment *(Px/Seg*) vorgeben lässt. Abb. 2 zeigt das Beispiel einer SLIC-Segmentierung mit unterschiedlichen Segmentgrößen. Da während der Klassifikation jedem Segment nur eine Objektklasse zugewiesen werden kann, ist die Einhaltung von Objektgrenzen durch die Segmentierung wichtig. Wie man Abb. 2 entnehmen kann, passen sich die Segmentgrenzen den Objektgrenzen größtenteils sehr gut an (z.B. Gebäudedach, Vegetation, Stromkasten), mit zunehmender Größe der Segmente (z.B. 450 Px/Seg in Abb. 2) werden allerdings kleinere Objekte, wie z.B. die Stange des Straßenschilds oder auch der Bürgersteig, mit benachbarten Segmenten verschmolzen. Wählt man kleinere Segmentgrößen (z.B. 250 Px/Seg), passt sich das Segmentierungsergebnis den Objektgrenzen besser an, die daraus resultierende größere Zahl an Segmenten lässt allerdings eine etwas längere Rechenzeit erwarten.

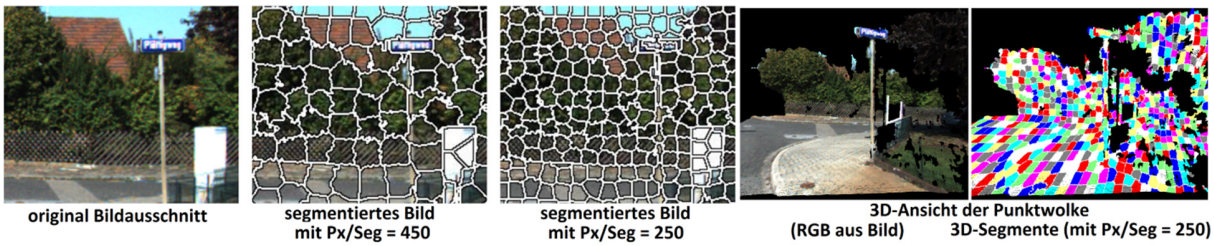

Abb. 2: Ergebnisse der SLIC Segmentierung mit unterschiedlicher Segmentgröße

2.3 Condtional Random Fields (CRF)

CRFs gehören zu den probabilistischen, diskriminativen und Kontext berücksichtigenden Klassifikationsverfahren. Die Berücksichtigung von Kontext erfolgt mithilfe eines statistischen Ansatzes, der die Abhängigkeiten zwischen benachbarten Primitiven modelliert. Dadurch beruht die Klassifizierung der Primitive nicht mehr allein auf ihren Merkmalen, sondern auch auf ihren Nachbarprimitiven. In einem CRF wird die a posteriori Verteilung $P(\boldsymbol{C}|\boldsymbol{x})$ der Klassenlabels aller Primitive (gesammelt in einem Vektor $\boldsymbol{C}$) bei gegebenen Daten $\boldsymbol{x}$ direkt modelliert und kann durch folgende Gleichung ausgedrückt werden (KUMAR & HEBERT 2006):

$$P(\boldsymbol{C}|\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x})} \exp\left(\sum_{i \in \boldsymbol{n}} A(\boldsymbol{x}, C_i) + \sum_{e_{ij} \in \boldsymbol{e}} I(\boldsymbol{x}, C_i, C_j)\right). \quad (1)$$

Dabei wird $A(\boldsymbol{x}, C_i)$ als *Assoziationspotenzial* und $I(\boldsymbol{x}, C_i C_j)$ als *Interaktionspotenzial* bezeichnet; diese Potenziale werden in Abschnitt 2.3.2 näher beschrieben. $Z(\boldsymbol{x})$ dient als Normalisierungskonstante. Das CRF wird durch einen Graphen $G = (\boldsymbol{n}, \boldsymbol{e})$ repräsentiert, der aus einer Menge von Knoten $\boldsymbol{n}$ und einer Menge von Kanten $\boldsymbol{e}$ besteht (BISHOP 2006). Die Knoten ent-

sprechen den zu klassifizierenden Punktwolkensegmenten, der Index i bezeichnet einen spezifischen Knoten in $\boldsymbol{n}$. Der Vektor $\boldsymbol{C}$ enthält die Labels C_i für jeden Knoten i und hat demnach dieselbe Anzahl an Elementen wie $\boldsymbol{n}$. Eine Kante e_{ij} aus $\boldsymbol{e}$ verbindet einen Knoten i mit seinem Nachbarn j; diese Kanten repräsentieren die statistischen Abhängigkeiten zwischen benachbarten Knoten. Ziel der Klassifikation ist es, jene Konfiguration $\boldsymbol{C}$ von Klassenlabels zu finden, für die $P(\boldsymbol{C}|\boldsymbol{x})$ maximal wird. Dabei werden die Parameter von $P(\boldsymbol{C}|\boldsymbol{x})$, d.h. die Parameter der Potenziale in Gleichung 1, aus Trainingsdaten gelernt. Die sich aus der Klassifikation der Segmente ergebenden Klassen werden anschließend auf die zugehörigen 3D-Punkte übertragen.

2.3.1 Definition des Graphen und der Merkmale

Die Knoten des CRF entsprechen in dieser Arbeit den 3D-Punktwolkensegmenten. Die Definition der Kanten des CRF erfolgt auf Grundlage von Nachbarschaften im Objektraum, weil diese nicht wie Nachbarschaftsbeziehungen im Bild durch die Abbildung der 3D-Szene auf die Bildebene beeinflusst sind. Zu diesem Zweck werden für jeden Punkt seine k-nächsten Nachbarn in der Punktwolke bestimmt. Für die k-nächste Nachbarn-Suche wird ein kd-Baum als Datenstruktur verwendet. Für Punkte an den Segmentgrenzen enthält die Menge dieser Nachbarn Punkte aus zwei oder mehr Segmenten. Die Häufigkeit, mit der Punkte aus einem Segment j in der Nachbarschaft von Punkten eines anderen Segmentes i anzutreffen sind, dient als Grundlage für die Definition der Kanten: liegt diese Häufigkeit über einem festzulegenden Schwellwert, wird zwischen den Knoten i und j eine Kante erzeugt. Da die den Knoten entsprechenden Segmente sowohl im Bild als auch in der Punktwolke vorliegen, kann zur Berechnung der Knotenmerkmalsvektoren $\boldsymbol{f}_i$, welche die Grundlage für das Assoziationspotenzial jedes Knoten i bilden, sowohl 2D-Bild- als auch 3D-Information genutzt werden. Das gleiche gilt für die Kantenmerkmalsvektoren $\boldsymbol{m}_{ij}$, welche im Interaktionspotenzial für jede Kante e_{ij} berücksichtig werden und die aus den Merkmalen der jeweiligen, durch die Kante verbundenen Knoten bestimmt werden. Für jedes Segment wird zunächst im 3D-Raum aus den zugehörigen 3D-Punkten $\boldsymbol{p}_n$ eine ausgleichende Ebene Ω der Form $\Omega\colon \mathrm{ax} + \mathrm{by} + \mathrm{cz} + \mathrm{d} = 0$ bestimmt, deren Parameter über eine Hauptkomponentenanalyse der Matrix $\boldsymbol{M}$ der Summen der Quadrate und Produkte der schwerpunktreduzierten Segmentpunkte $\boldsymbol{p}'_n = \boldsymbol{p}_n - \overline{\boldsymbol{p}} = [x'_n, y'_n, z'_n]^T$ ermittelt werden, mit $\overline{\boldsymbol{p}}$ als jeweiligem Segmentschwerpunkt und

$$\boldsymbol{M} = \begin{bmatrix} \Sigma x_n'^2 & \Sigma x'_n y'_n & \Sigma x'_n z'_n \\ \cdot & \Sigma y_n'^2 & \Sigma y'_n z'_n \\ \cdot & \cdot & \Sigma z_n'^2 \end{bmatrix}.$$

Der Normalvektor $\boldsymbol{n} = [a\ b\ c]^T$ entspricht dabei dem Eigenvektor zum kleinsten Eigenwert von $\boldsymbol{M}$ und d ergibt sich über $d = -\boldsymbol{n}^T\overline{\boldsymbol{p}}$. Aus den 3D-Punkten und den so ermittelten Ebenen kann zusätzlich zu der reinen Bildinformation eine Reihe von Merkmalen abgeleitet werden. Alle hier verwendeten radiometrischen und geometrischen Knoten- und Kantenmerkmale sind in Tab. 1 zusammengefasst.

Tab. 1: Übersicht über verwendete Knoten- und Kantenmerkmale

Knotenmerkmale f_i		**Kantenmerkmale m_{ij}**
2D-Merkmale	**3D-Merkmale**	
- RGB-Werte - Intensität, Sättigung, Farbton - Haralick Features (Energie, Kontrast, Homogenität, Entropie) (HARALICK ET AL., 1973)	- Eigenwerte und Eigenvektoren von **M**, Normalvektor der 3D-Ebenen - Segmentschwerpunkt - Anisotropie und Planarität (CHEHATA ET AL., 2009) - Höhe über Grund	- Winkel α_{ij} zw. Normalvektoren d. Seg. - Höhenunterschied Δh_{ij} und Distanz der Schwerpunkte der Segmente - Differenz der mittleren Intensität der Segmente

2.3.2 Definition der Potenziale

Für die Berechnung der Potenziale können beliebige diskriminative Klassifikatoren mit probabilistischem Output verwendet werden (KUMAR & HEBERT 2006). Das Assoziationspotenzial $A(\boldsymbol{x}, C_i)$ beschreibt die Wahrscheinlichkeit $P(C_i|\boldsymbol{x}) = P(C_i|\boldsymbol{f}_i)$ für das Auftreten einer Klasse bei gegebenem Knotenmerkmalsvektor $\boldsymbol{f}_i$, wobei als Klassifikator *Random Forests* (RF) (BREIMAN 2001) eingesetzt werden. Zur Berechnung des Interaktionspotenzials werden in dieser Arbeit zwei Modelle genutzt und evaluiert: Ein *binäres* Modell, welches eine Wahrscheinlichkeit dafür beschreibt, dass benachbarte Segmente i und j bei gegebenem Merkmalsvektor $\boldsymbol{g}_{ij} = [\boldsymbol{f}_i, \boldsymbol{f}_j, \boldsymbol{m}_{ij}]^T$ zur gleichen Klasse gehören, wobei auch hier ein RF als Klassifikator verwendet wird, sowie ein als *segmentbasiert* bezeichnetes Modell, in welchem die Stärke des Glättungseffekts mit

$$I(\boldsymbol{x}, C_i, C_j) = \begin{cases} w \cdot (\lambda_1 + \lambda_2 \cdot \cos\alpha_{ij} + \lambda_3 \cdot \exp(-\frac{(\Delta h_{ij})^2}{2\sigma^2})), & \text{für } C_i = C_j \\ 0, & sonst \end{cases} \quad (2)$$

von dem Winkel zwischen den Normalvektoren α_{ij} und dem Höhenunterschied Δh_{ij} der benachbarten Segmente i und j abhängt. Die Parameter $\lambda_{1,2,3} \geq 0$ mit $\lambda_1 + \lambda_2 + \lambda_3 = 1$ beschreiben die Gewichte der einzelnen Komponenten, wobei λ_1 den Grad der datenunabhängigen Glättung bestimmt. w ist ein Gewichtsfaktor des gesamten Interaktionspotenzials relativ zum Assoziationspotenzial, während σ^2 die Varianz der Höhendifferenz bezeichnet. Die Glättung durch dieses Modell ist somit dann am stärksten, wenn benachbarte Segmente ähnliche Normalvektoren haben, sodass $\cos\alpha_{ij} \approx 1$ und der Höhenunterschied zwischen den Segmentschwerpunkten mit $\Delta h_{ij} \approx 0$ gering ist.

2.3.3 Training und Inferenz

Im Zusammenhang mit graphischen Modellen bedeutet *Inferenz* die Bestimmung der optimalen Labelkonfiguration $\boldsymbol{C}$ durch die Maximierung von $P(\boldsymbol{C}|\boldsymbol{x})$ aus Gleichung (1). Für Graphen mit Zyklen ist die Inferenz im Mehrklassenfall nicht exakt lösbar. Daher werden approximative Lösungen verwendet, wobei in dieser Arbeit die *Loopy Belief Propagation* (LBP) (FREY & MACKAY 1998) zum Einsatz kommt. Das überwachte Training der beiden Potenziale, für welches manuell gelabelte Datensätze benötigt werden, erfolgt getrennt voneinander. Bei der Verwendung von RF als Klassifikatoren müssen die im RF verwendeten Entscheidungsbäume anhand der Trainingsdaten angelernt werden. Für das segmentbasierte Modell wird während des

Trainings der Parameter σ bestimmt, die Festlegung der Glättungskoeffizienten sowie des Gewichtsfaktors erfolgte empirisch.

3 Ergebnisse und Diskussion

Die Evaluierung der entwickelten Klassifikationsmethodik erfolgte anhand von 44 Stereobildern aus drei Stereosequenzen der „KITTI Vision Benchmark Suite“ (GEIGER et al. 2012) durch Kreuzvalidierung. Wie bei SENGUPTA et al. (2013) werden die zehn Objektklassen *Himmel, Gebäude, Vegetation, Straße, Gehweg, Zaun, Fahrzeug, Schild, Stange/Mast* und *Person* unterschieden, wobei die von SENGUPTA et al. (2013) übernommenen Referenzdaten durch manuelle Erfassung zusätzlicher Daten erweitert wurden. Die Gesamtgenauigkeit der Klassifikation mittels der unterschiedlichen Modelle ist in Tab. 2 für verschiedene Segmentgrößen dargestellt. Die Tabelle zeigt ebenso wie Abb. 3, dass, während sich für das Modell ohne Kontext bereits gute Ergebnisse ergeben, die Berücksichtigung von Kontext mittels des *binären* Modells und insbesondere bei Verwendung des *segmentbasierten* Modells zu einer Verbesserung sowie einer Glättung des Klassifikationsergebnisses führt.

Tab. 2: Punktweise ermittelte Gesamtgenauigkeiten unter Verwendung verschiedener Modelle für das Interaktionspotenzial.

Modell	100 [Px/Seg]	250 [Px/Seg]	450 [Px/Seg]
Ohne Kontext	87.9%	89.0%	89.1%
Binär	88.1%	89.6%	90.0%
segmentbasiert	91.4%	90.9%	91.0%

Genauigkeit

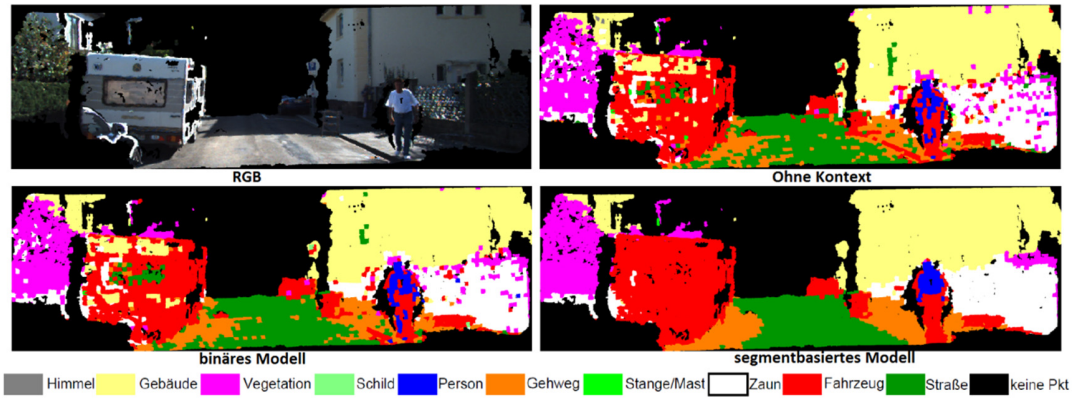

Abb. 3: Klassifikationsergebnis eines Stereomodells unter Verwendung unterschiedlicher Modelle für die Interaktionspotenziale (100 Px/Seg).

Betrachtet man die Klassifikationsqualität einzelner Klassen für das segmentbasierte Modell in Tab. 3, mit $Qualität[\%] = 100 \cdot \#TP/(\#TP + \#FN + \#FP)$, wobei

- #TP: Anzahl der richtig zugeordneten Primitive,
- #FN: Anzahl der irrtümlich einer anderen Klasse zugeordneten Primitive und
- #FP: Anzahl der irrtümlich dieser Klasse zugeordneten Primitive sind,

sowie das anschauliche Beispiel der Klassifikation einer Person in Abb. 4, so fällt auf, dass Klassen mit kleinen Objekten (*Schilder*, *Stange/Mast*, *Personen*) problematisch für die Klassifikation sind und mit zunehmender Segmentgröße schlechter erkannt werden. Ein Grund hierfür ist die bereits in Abb. 2 gezeigte Problematik der Segmentierung, mit zunehmender Segmentgröße die Objektgrenzen kleinerer Objekte zu erhalten, wodurch in die Merkmalsberechnung dieser Segmente Pixel und Punkte unterschiedlicher Objekte und/oder unterschiedlicher Objektklassen einbezogen werden, was zu instabileren Segmentmerkmalen führt. Ebenso können durch die Nichteinhaltung von Objektgrenzen durch die Segmente Nachbarschaftsbeziehungen zwischen Objekten im Vordergrund und Objekten im Hintergrund entstehen, welche durch die hier verwendete Definition der Nachbarschaften im 3D-Objektraum anstatt in der Bildebene eigentlich vermieden werden sollten. Auch die im Vergleich zu anderen Objekten relativ geringe Menge an Trainingsdaten für diese Klassen kann sich negativ auf das Ergebnis auswirken.

Tab. 3: Punktweise ermittelte Qualität einzelner Klassen bei verschiedenen Segmentgrößen

Segmentgröße	Straße	Gebäude	Vegetation	Fahrzeug	Schild	Stange/Mast	Person
100 [Px/Seg]	90.1%	86.1%	87.1%	84.7%	28.3%	21.2%	46.6%
450 [Px/Seg]	90.2%	85.8%	86.5%	82.0%	8.8%	6.1%	26.1%

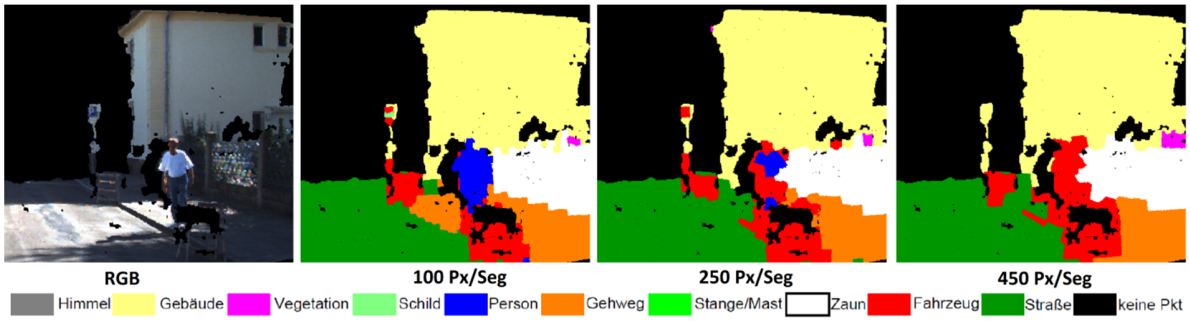

Abb. 4: Ausschnitt der Klassifikation einer Szene unter Verwendung unterschiedlicher Segmentgrößen

4 Fazit und Ausblick

Zusammenfassend lässt sich festhalten, dass sich mittels der in dieser Arbeit entwickelten Klassifikationsmethode für Stereobilder bereits ohne die Verwendung von Kontextinformation sehr gute Gesamtgenauigkeiten erreichen lassen, welche durch die Berücksichtigung von Kontext mithilfe verschiedener Modelle für das Interaktionspotenzial, insbesondere unter der Verwendung des *segmentbasierten* Modells, weiter verbessert werden können. Ferner wird damit eine deutlich glattere Labelkonfiguration erreicht. Die segmentweise Bearbeitung führt zu einer Reduktion der Rechenkomplexität, allerdings bei zu groß gewählten Segmenten auch zu Problemen für die Klassifikation kleinerer Objekte.
Weiterführende Untersuchungen der in dieser Arbeit entwickelten Methodik sollten insbesondere dahingehen, die bisher separat und unabhängig voneinander klassifizierten Stereobilder einer Sequenz gemeinsam innerhalb eines multitemporalen CRF-Modells, unter Hinzunahme einer zusätzlichen zeitlichen Komponente, zu klassifizieren.

5 Literaturverzeichnis

ACHANTA, R., SHAJI, A., SMITH, K., LUCCHI, A., FUA, P. & SUSSTRUNK, S., 2012: SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34** (11), 2274-2282.

ANGUELOV, D., TASKARF, B., CHATALBASHEV, V., KOLLER, D., GUPTA, D., HEITZ, G. & NG, A., 2005: Discriminative learning of Markov random fields for segmentation of 3D scan data. IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2**, 169-176.

BISHOP, C., 2006: Pattern Recognition and Machine Learning. New York, USA: Springer.

BREIMAN, L., 2001: Random Forests. Machine Learning **45** (1), 5-32.

CHEHATA, N., GUO, L. & MALLET, C., 2009: Airborne Lidar Feature Selection for urban Classification using Random Forests. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **38** (3/W8), 207-212.

FREY, B. & MACKAY, D., 1998: A revolution: Belief propagation in graphs with cycles. Advances in Neural Information Processing Systems **10**, 479-485.

GEIGER, A., LENZ, P. & URTASUN, R., 2012: Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. IEEE Conference on Computer Vision and Pattern Recognition, 3354-3361.

GEIGER, A., ROSER, M. & URTASUN, R., 2011: Efficient Large-Scale Stereo Matching. Computer Vision - ACCV 2010, **6492**, Springer Berlin Heidelberg, 25-38.

HARALICK, R.M., SHANMUGAM, K. & DINSTEIN, I., 1973: Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics **3** (6), 610-621.

KUMAR, S. & HEBERT, M., 2006: Discrimantive Random Fields. International Journal of Computer Vision **68** (2), 179-201.

LIM, E.H. & SUTER, D., 2009: 3D terrestrial LIDAR classifications with super-voxels and multi-scale Conditional Random Fields. Computer-Aided Design 2009 **41** (10), 701-710.

MATTI, E.K. & NEBIKER, S., 2014: Geometry and Colour Based Classification of Urban Point Cloud Scenes Using a Supervised Self-Organizing Map. Photogrammetrie Fernerkundung Geoinformation, (3), 161-173.

REICH, M., UNGER, J., ROTTENSTEINER, F. & HEIPKE, C., 2013: On-Line compatible Orientation of a Micro-UAV based on Image Triplets. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **2** (3), 37-42.

SENGUPTA, S., GREVERSON, E., SHAHROKNI, A. & TORR, P., 2013: Urban 3D semantic modelling using stereo vision. IEEE International Conference on Robotics and Automation, 580-585.