Semantische Segmentierung

Björn Fröhlich

Lehrstuhl für Digitale Bildverarbeitung
Friedrich-Schiller-Universität Jena
bjoern.froehlich@uni-jena.de
http://www.inf-cv.uni-jena.de

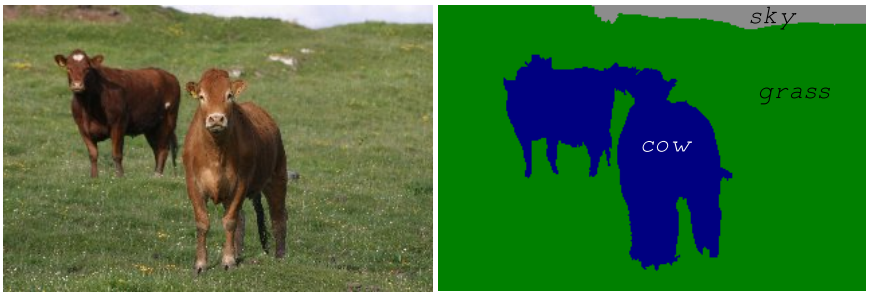
Abstract: Das automatische Erlernen und Erkennen von Objektkategorien und deren Instanzen gehören zu den wichtigsten Aufgaben der digitalen Bildverarbeitung. Aufgrund der aktuell sehr weit fortgeschrittenen Verfahren, die diese Aufgaben beinahe so gut wie ein Mensch erfüllen können, hat sich der Schwerpunkt von einer groben auf eine genaue Lokalisierung der Objekte verlagert. In der vorliegenden Arbeit werden verschiedene Techniken für die Pixel genaue Klassifikation von Bildern, auch *semantischen Segmentierung* genannt, analysiert. Dieses relativ neue Gebiet erweitert die grobe Lokalisierung von Objekten in Bildern um eine punktgenaue Klassifikation. Der Schwerpunkt dieser Arbeit ist es, aktuelle Verfahren der *semantischen Segmentierung* zu vergleichen. Dabei werden verschiedene Methoden zur Ermittlung von Merkmalen, Merkmalstransformationen, Klassifikation und zur globalen Optimierung, wie zum Beispiel durch die Betrachtung von formbasierten Eigenschaften, vorgestellt. Abschließend werden die präsentierten Verfahren in umfangreichen Experimenten auf verschiedenen, frei zugänglichen Datensätzen verglichen und analysiert.

Art der Arbeit:	Diplomarbeit
Betreuer:	Dipl.-Inf. Erik Rodner
	Prof. Dr.-Ing. Joachim Denzler

1 Einleitung

Die Aufteilung von Bildern in mehrere Bereiche und die Zuordnung dieser zu Objektklassen hat in den letzten Jahren stark an Bedeutung gewonnen. Noch vor zehn Jahren war es nahezu unmöglich, die Frage zu beantworten, ob ein Bild eine Instanz einer bestimmten Objektklasse enthält (Bildkategorisierung). Seit neue Methoden dieses lösen können, stellt sich die Frage, wo im Bild sich diese Objekte befinden (Objektlokalisation). Das Ziel dieser Diplomarbeit war es, den nächsten Schritt in dieser Entwicklung anzugehen: die punktgenaue Lokalisation von Objekten in Bildern. Hierbei wird für jeden Pixel die Zugehörigkeit zu einer bekannten Objektklasse bestimmt. Dieses Verfahren ist auch bekannt als *semantischen Segmentierung*. Ein Beispiel hierfür wird in Abbildung 1 dargestellt.

Den Inhalt von Bildern automatisch zu erkennen, ist ein weitgehend ungelöstes Problem. Nützlich könnten die Ergebnisse zum Beispiel für eine Online Bildersuche sein, um eine Datenbank von Bildern gezielt auf Objekte und Kombinationen von Objekten zu prüfen.



(a) Eingabebild (b) Ergebnis der *semantischen Segmentierung*

Abbildung 1: Beispielergebnis der *semantischen Segmentierung*; jede Farbe steht für eine andere Klasse

Zur Zeit wird hierfür meist der Dateiname des Bildes bzw. der textuelle Inhalt der dazugehörenden Internetseite analysiert. Beides spiegelt nicht zwangsläufig den Inhalt eines Bildes wieder.

Ein weiteres Anwendungsgebiet für die *semantischen Segmentierung* ist die automatische Analyse und Auswertung von Satellitenaufnahmen. Für das Projekt OpenStreetMap[1] wäre es zum Beispiel sehr nützlich, wenn Flächen wie Wald, Straßen und Gebäude automatisch punktgenau lokalisiert und diese in eine digitale Karte umgewandelt werden. Zur Zeit wird dies durch die Mitglieder des Projekts manuell umgesetzt.

Weiterhin ist es sinnvoll, für 3-D-Modelle von Städten vorhandene Aufnahmen der Fassaden platzsparend zu speichern. Dies kann nur erreicht werden, wenn ein Modell der Gebäude erzeugt wird, welches wiederum das Wissen über die punktgenaue Lage von Fenstern, Türen und Wand in den Eingabebildern voraussetzt.

In der Diplomarbeit „Semantische Segmentierung“ werden verschiedene Verfahren analysiert. Es werden vorhandene Verfahren vorgestellt und mit neuen Ideen und Algorithmen kombiniert. Abschließend werden diese Methoden getestet, genau ausgewertet und verglichen.

2 Allgemeiner Ablauf

In diesem Abschnitt soll der abstrakte Ablauf eines Algorithmus zur *semantischen Segmentierung* mit lokalen Merkmalen dargestellt werden. Abbildung 2 zeigt die drei erforderlichen und weitere optionale Schritte. Als erstes werden für jedes Bild an verschiedenen Stellen lokale Merkmale bestimmt. Hierfür werden die sogenannten Opponent-SIFT-Merkmale von [vdSGS10] verwendet. Diese Merkmale zeichnen sich durch Invarianz gegenüber Skalierung, Rotation und Beleuchtungsänderungen aus. Anschließend können diese Merkmale in sogenannte High-Level-Merkmale transformiert werden. Csurka et al.

[1] http://www.openstreetmap.org/

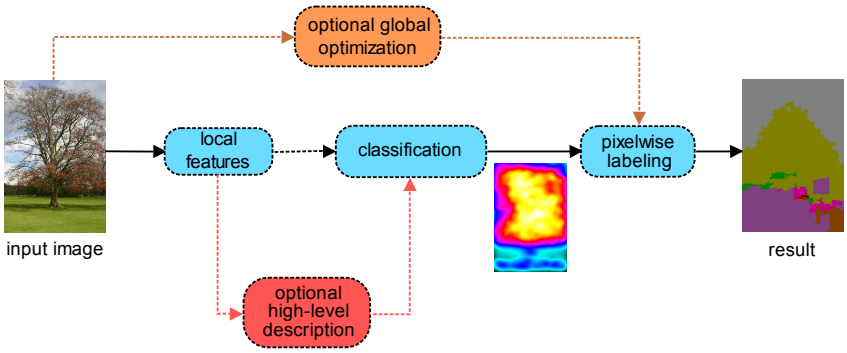


Abbildung 2: Ablaufschema für die *semantischen Segmentierung* mit lokalen Merkmalen ähnlich zu [CP08]

[CP08] schlägt hierfür zwei Verfahren vor: ein Bag-of-Words-Ansatz basierend auf einen Gaußschen Mischmodell (GMM) und dem Fisher-Kernel. Ein Ziel der Diplomarbeit war das Ermitteln eines robust und schnell geschätzten GMM. Zusätzlich wird noch ein alternativer Bag-of-Words Ansatz basierend auf k-Means vorgestellt und mit den beiden anderen Methoden verglichen. Diese High-Level-Merkmale bzw. die lokalen Merkmale werden anschließend klassifiziert. Dies geschieht entweder mit einem Wald aus Entscheidungsbäumen, die randomisiert angelernt werden [Bre01], oder einem Sparse-Logistic-Regression Klassifikator [KH05]. Somit wird für jede Klasse und für jeden Pixel eine Wahrscheinlichkeit ermittelt, welche in sogenannten Wahrscheinlichkeitskarten für jede Klasse dargestellt wird.

Ein weiterer Schwerpunkt der Diplomarbeit ist die Optimierung dieser Ergebnisse durch verschiedene Nachverarbeitungsschritte. Eine zusätzliche Verwendung einer klassischen Segmentierung [CM02] führt zum Beispiel zu signifikant besseren Erkennungsraten. Außerdem wird ein neues Verfahren vorgestellt, welches sich am klassischen Regionen-Wachstum [Ros98] orientiert und anhand von Momenten [Hu62] die Form von Objekten ausnutzt, um das Ergebnis zu verbessern. Um fehlerhafte Regionen zu beseitigen, wird zusätzlich ein auf Graph-Cut basierendes Verfahren [BK04, RVG$^+$07] präsentiert. Weiterhin wird gezeigt, wie die typische relative Lage zweier Klassen zueinander (zum Beispiel ist Klasse „Himmel“ immer über Klasse „Gras“) zur Optimierung verwendet werden kann [GRC$^+$08].

3 Zusammenfassung

Die verschiedenen Verfahren wurden unter anderem auf den eTRIMS Datensatz [KF09] getestet. Es konnte gezeigt werden, dass die Zufälligen Wälder bessere Erkennungsraten liefern, wenn die Merkmale nicht mit dem Bag-of-Words Ansatz transformiert werden und dass dies beim Sparse Logistic Regression Klassifikator genau anders herum ist. Insgesamt ist die Leistung der Kombination von Bag-of-Words mit Sparse Logistic Regression etwas

besser (ca. 65% auf eTRIMs), als die zufälligen Wälder ohne Bag-of-Words (64%). Dafür sind die zufälligen Wälder signifikant schneller beim Training und beim Testen.

Es konnte gezeigt werden, dass das Graph-Cut basierende globale Optimierungsverfahren Verbesserungen bringt (ca. 2%) in dem es übliche Nachbarschaften ausnutzt. In der Diplomarbeit konnten verschiedene Ansätze zur *semantischen Segmentierung* vorgestellt und weiterentwickelt werden. Es wurde gezeigt, dass unterschiedliche Methoden ähnliche Ergebnisse liefern und dass diese Ergebnisse durch weitere Optimierungsschritte verbessert werden können.

Literatur

[BK04] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[Bre01] Leo Breiman. Random Forests. *Mach. Learn.*, 45(1):5–32, 2001.

[CM02] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[CP08] G. Csurka and F. Perronnin. A Simple High Performance Approach to Semantic Segmentation. In *British Machine Vision Conference*, pages 213–222, 2008.

[GRC+08] Stephen Gould, Jim Rodgers, David Cohen, Gal Elidan, and Daphne Koller. Multi-Class Segmentation with Relative Location Prior. *Int. J. Comput. Vision*, 80(3):300–316, 2008.

[Hu62] Ming K. Hu. Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, IT-8:179–187, February 1962.

[KF09] Filip Korč and Wolfgang Förstner. eTRIMS Image Database for Interpreting Images of Man-Made Scenes. Technical Report TR-IGG-P-2009-01, Dept. of Photogrammetry, University of Bonn, March 2009.

[KH05] Balaji Krishnapuram and Alexander J. Hartemink. Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):957–968, June 2005.

[Ros98] Paul L. Rosin. Refining Region Estimates. *International Journal of Pattern Recognition and Artificial Intelligence*, 12(6):841–866, 1998.

[RVG+07] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in Context. In *Proc. IEEE 11th International Conference on Computer Vision ICCV 2007*, pages 1–8, 14–21 Oct. 2007.

[vdSGS10] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(in press), 2010.