

DISS. ETH NO. 23459

Modeling Local and Global Context Information for Aerial Image Labeling

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Javier Alexander Montoya Zegarra

Master of Science in Computer Science (M.Sc. CS)
University of Campinas (Unicamp)

born on 30.01.1982

citizen of Arequipa, Peru

accepted on the recommendation of

Prof. Dr. Konrad Schindler, examiner
ETH Zurich, Switzerland

Prof. Dr. Pascal Fua, co-examiner
EPF Lausanne, Switzerland

Dr. Jan Dirk Wegner, co-examiner
ETH Zurich, Switzerland

2016

Abstract

In this thesis we explore methods to semantically analyse aerial images, with special emphasis on the automatic extraction of road networks and buildings. By semantic interpretation we mean analysing the image content and dividing it into a set of coherent regions each of which is associated with a semantic class label.

In spite of great progress, automatic aerial image labeling remains, to a large part, an unsolved problem. In our work we concentrate on urban scenes. Built-up areas exhibit a large variety of reflectance patterns, with large intra-class variations and often also low inter-class variation. Local object appearance thus is ambiguous, and one must exploit knowledge about the objects' structure. To that end, this thesis presents priors that go beyond standard smoothness or co-occurrence assumptions. They specifically model the structure of roads and buildings, by adding high-level shape representations for both classes. Buildings are represented by sets of compact polygons, while roads are modeled as a combination of long, narrow segments and larger, compact blobs.

Along with new models and inference methods, we also introduce two new aerial datasets for semantic aerial segmentation. We show marked improvements with respect to the state-of-art, highlighting thus that class-specific priors allow for more accurate semantic image segmentation.

Keywords: aerial image labeling, semantic segmentation, road network and building extraction, graphical models.

Zusammenfassung

Diese Arbeit untersucht Methoden zur semantischen Analyse von Luftbildern, mit besonderem Fokus auf der Extraktion von Strassennetzen und Gebäuden in städtischen Gebieten. Die hier vorgestellten Algorithmen interpretieren die Luftbilder indem Sie jedem Bildpunkt einer Klasse zuweisen und somit das Luftbild in koherente, semantische Einheiten aufteilen.

Angetrieben von grossem wirtschaftlichem aber auch staatlichem Interesse konnten auf diesem Gebiet bereits in der Vergangenheit bedeutende Fortschritte erzielt werden. Dennoch gilt die automatische Extraktion praktisch verwertbarer semantischer Informationen aus Luftbildern auch bis zuletzt als ein weitgehend ungelöstes Problem. Die meisst homogen gewachsenen städtischen Gebiete weisen in der Regel nur eine geringe visuelle Vielfalt zwischen den semantischen Klassen auf. Gleichzeitig führt der Anspruch an eine gewisse Robustheit der Verfahren, gerade unter realistischen Bedingungen wie z.B. inkonsistenten Wetterbedingungen sowie unkontrollierter Beleuchtung, zu einer erhöhten Varianz der Daten innerhalb der Klassen.

Um den gehobenen Anforderungen unter realistischen äusseren Bedingungen gerecht zu werden, werden innerhalb dieser Arbeit verschiedene a-priori Annahmen eingeführt, die weit über den Standard bekannter Modelle, wie dem Potts-Modell oder simpler räumlicher Ko-okkurrenz, hinausgehen. Das entwickelte Modell basiert darauf statistische Regelmässigkeiten in der räumlicher Struktur von, z.B., Strassen und Gebäuden auszunutzen. Um solche statistischen Eigenschaften besser beschreiben zu können, werden die komplexen Eigenschaften der verschiedenen Entitäten mittels einer vereinfachenden Darstellung repräsentiert: Strassen als Kombination von länglichen Segmenten verbunden durch kreisförmige Knotenpunkte. Gebäude hingegen durch eine Menge kompakter Polygone. Die Art der Repräsentation erlaubt zudem die Entwicklung praktischer Algorithmen zur Inferenz in den vorgeschlagenen Modellen.

Des weiteren werden im Verlauf dieser Arbeit zwei neue Datensätze zum untersuchten Thema vorgestellt. Eine ausführliche Auswertung der entwickelten Methoden, zeigt die Vorteile der hier vorgestellten Modelle, die an der Spitze der Forschung in diesem Bereich liegen. Dies erlaubt in einer abschliessenden Bewertung den Schluss, dass eine spezifische Modellierung verschiedener Klassen anhand geeigneter a-priori Modelle die Qualität einer semantischen Segmentierung in Luftbildern deutlich verbessert.

Stichwörter: Luftbild Kennzeichnung, Semantische Segmentierung, Extraktion von Strassenkarten und Gebäuden, Probabilistische Graphische Modelle.