



Doctoral Thesis

## Semantic 3D Modeling from Images with Geometric Priors

**Author(s):**

Häne, Christian

**Publication Date:**

2016

**Permanent Link:**

<https://doi.org/10.3929/ethz-a-010657817> →

**Rights / License:**

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 23450

# Semantic 3D Modeling from Images with Geometric Priors

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

presented by

**Christian Häne**

MSc ETH CS, ETH Zürich

born on 22.11.1984

citizen of  
Kirchberg SG

accepted on the recommendation of

Prof. Dr. Marc Pollefeys

Prof. Dr. Konrad Schindler

Prof. Dr. Daniel Cremers

Prof. Dr. Jitendra Malik

2016

# Abstract

Acquiring dense 3D models automatically from a collection of images has been one of the core research topics in computer vision over the last two decades and has reached a level of maturity where it starts to be used commercially. However, for many practical use cases having a purely geometric 3D reconstruction is not sufficient. A segmentation into individual objects and their semantic classes plays an important role for example in robotics or architectural applications. While for some applications the lack of semantic information can be compensated by user interaction this is not a viable option for others. Hence, there is a need for obtaining semantic 3D models fully automatically from images.

This thesis introduces several algorithms which exploit semantic information in the form of geometric priors or facilitate a joint reconstruction and semantic segmentation. By reasoning about semantics and geometry jointly, better results are achieved than for each task individually.

For self-driving cars a method is presented which extracts obstacles out of fisheye images in real-time and eventually computes a semantic segmentation of the 2D plane into obstacles and drivable area. For reconstructing datasets of buildings taken from the ground a joint reconstruction and semantic segmentation approach is introduced. The main idea is to segment the 3D volume into different semantic classes plus a free space class. This makes it possible for the reconstruction and segmentation task to benefit from each other. Further, 3D object shape priors are included into the formulation to tackle objects where the geometry can not otherwise be accurately reconstructed due to missing data in textureless and other difficult areas in the images. For some applications, the final goal is the reconstruction from only a single viewpoint. To this end, an approach for man-made environments which segments the scene into piece-wise planar patches and an approach which

uses the information of a surface normal classifier are introduced. The latter method achieves plausible 3D reconstructions from a single color image when combined with a single view depth classifier.

# Zusammenfassung

Digitale 3D Modelle vollautomatisch aus Bildern zu erfassen ist seit mehr als zwei Jahrzehnten eines der Kerngebiete in der Computer Vision. Dadurch wurde ein Niveau erreicht das eine kommerzielle Verwendung ermöglicht. Allerdings ist für viele praktische Anwendungen eine rein geometrische 3D Rekonstruktion nicht genügend. Eine Segmentierung in einzelne Objekte und deren semantische Klassen spielt eine wichtige Rolle, zum Beispiel in der Robotik und Architektur. Während für einige Anwendungen der Mangel von semantischer Information durch manuelle Benutzerinteraktion kompensiert werden kann ist dies für andere Anwendungen keine gangbare Option. Deshalb herrscht ein Bedarf für Methoden die vollautomatisch semantische 3D Modelle aus Bildern generieren können.

Diese Arbeit führt mehrere Algorithmen ein die semantische Information in Form von geometrischen Vorkenntnissen oder gemeinsamer Rekonstruktion und semantischer Segmentierung ausnützen. Durch das gemeinsame begründen von Geometrie und Semantik können bessere Resultate erzielt werden als wenn beide Aufgaben einzeln gelöst werden.

Für selbst fahrende Autos werden Methoden präsentiert die Hindernisse aus Fischaugenbildern in Echtzeit extrahieren und schlussendlich eine semantische Segmentierung der 2D Ebene in Hindernisse und fahrbaren Bereich berechnen. Für die Rekonstruktion von Datensätzen von Gebäuden die vom Boden aufgenommen wurden, wird ein Ansatz eingeführt der gemeinsam rekonstruiert und segmentiert. Die Hauptidee ist es, das 3D Volumen in verschiedene semantische Klassen und eine Klasse für den freien Raum zu segmentieren. Dies ermöglicht es dass beide Aufgaben voneinander profitieren. Des weiteren werden Vorkenntnisse über die Form von 3D Objekten bestimmter Klassen, die sonst nicht genau rekonstruiert werden können wegen fehlender Daten in nicht texturi-

erten oder sonst schwierigen Bildbereichen, in die Formulierung eingebunden. Für einige Applikationen ist es das Ziel eine Rekonstruktion von einer einzelnen Ansicht aus zu generieren. Dazu wird ein Ansatz für vom Mensch gefertigte Umgebungen eingeführt der die Szene in stückweise planare Flächen aufteilt und ein Ansatz der einen Klassifizierer verwendet der die Oberflächenrichtung schätzt. Der zweite Ansatz ermöglicht plausible 3D Rekonstruktionen von nur einem einzelnen Farbbild wenn er zusammen mit einem Algorithmus der die Tiefe von einem einzelnen Farbbild schätzt verwendet wird.