ISPRS Journal of Photogrammetry and Remote Sensing 90 (2014) 68–82

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Indoor scene reconstruction using feature sensitive primitive extraction and graph-cut

Sven Oesau [*], Florent Lafarge, Pierre Alliez

Inria Sophia Antipolis – Méditerranée, Sophia Antipolis, France

ARTICLE INFO

Article history:
Received 16 October 2013
Received in revised form 6 February 2014
Accepted 7 February 2014

Keywords:
Indoor scenes
3D reconstruction
Laser scanning
Multi-scale line extraction
Graph cut
Energy minimization

ABSTRACT

We present a method for automatic reconstruction of permanent structures, such as walls, floors and ceilings, given a raw point cloud of an indoor scene. The main idea behind our approach is a graph-cut formulation to solve an inside/outside labeling of a space partitioning. We first partition the space in order to align the reconstructed models with permanent structures. The horizontal structures are located through analysis of the vertical point distribution, while vertical wall structures are detected through feature preserving multi-scale line fitting, followed by clustering in a Hough transform space. The final surface is extracted through a graph-cut formulation that trades faithfulness to measurement data for geometric complexity. A series of experiments show watertight surface meshes reconstructed from point clouds measured on multi-level buildings.

1. Introduction

In recent years the reconstruction of 3D architectural scenes has received more and more attention. Blueprints and precise models are often needed for the architecture, engineering and construction application domains. As the physical geometry of buildings often differs from the original plans, the reconstruction of a precise 3D model of the interior is a common need. Such reconstruction is often a manual or semi-automatic time-consuming process.

Compared to the reconstruction of outdoor environments, indoor scene reconstruction is still in an early stage. Applying outdoor reconstruction methods to indoor scenes is not practical as indoor impose different challenges than outdoor scenes. The outside of a building can often be described by a single or few cuboids, and the amount of clutter hiding part of the geometry is rather low. In contrast, interior space often has a more complex geometry and a high quantity of clutter. Further hurdles come from the varying point density and anisotropic sampling.

1.1. Related work

Surface reconstruction has been an active research topic for decades. Despite the wide variety of methods, they often perform unsatisfactorily for extracting a surface representation from indoor scenes. They commonly result in a single surface representation approximating the entire point cloud, which is unsatisfactory for further specialized processing such as semantic labeling into floors, ceilings and walls. A favored approach would provide a classification of the point cloud into permanent structures and clutter. General surface reconstruction methods instead assume that the point cloud is created from a single surface whereas indoor scenes are usually composed of planar parts and arbitrarily shaped clutter.

The RANSAC-based algorithm proposed by Schnabel et al. (2007) provides a valid solution for extracting several types of shapes from an oriented point cloud, i.e., from points with oriented normals. While missing normals can be estimated by robust feature preserving methods (Boulc'h and Marlet, 2012), adjusting the parameters of Schnabel's algorithm to detect primitives in indoor scenes is often a trial-and-error process. Another drawback for indoor scene reconstruction is that Schnabel's algorithm is not robust to highly variable point density and strong anisotropy. For the reconstruction of permanent structures from indoor scans, we advocate for considering a domain-specific knowledge. Common knowledge assumptions are piecewise planar permanent structures and Manhattan-World scenes, i.e., exactly three orthogonal directions: two for the walls and one for floors and ceilings. We classify the existing approaches into three categories: *Planar primitive detection*, *Volumetric primitive fitting* and *Mesh-based reconstruction*.

---

[*] Corresponding author. Tel.: +33 4 97 15 50 43.
E-mail address: sven.oesau@inria.fr (S. Oesau).

http://dx.doi.org/10.1016/j.isprsjprs.2014.02.004
0924-2716/© 2014 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.

1.1.1. Planar primitive detection

Sanchez and Zakhor (2012) recently proposed a method for modeling building interiors as an arrangement of planar polygons. They classify the points based on estimated normals by assuming a Manhattan-World scene. The planar primitives are reconstructed by a region growing segmentation, least-squares fitting by employing RANSAC and alpha shapes (Edelsbrunner et al., 2006) to reconstruct polygons. Additionally, they employ model fitting of a staircase model to unsegmented points. The results presented in form of a polygon soup show that they can deal with multi-level buildings. There is however no structure information such as adjacency of primitives.

Another primitive detection approach has been proposed by Budroni and Boehm (2010, 2009). The authors apply sweeping techniques to identify Manhattan-World directions and to locate wall segments. The floor plane is decomposed by the so-detected segments and marked as inside or outside based on the point density. The output is a watertight model extracted from the ground plan. However, resilience to missing data is not addressed.

Another approach proposed by Okorn et al. (2010) uses the Hough transform to extract wall segments from the point cloud. Although the majority of walls are detected, they are often not captured in their full extent. The output of the algorithm is a set of unconnected wall segments that are used to classify the points into permanent structures and clutter. However, neither structural relations nor volumes of the indoor space are provided.

Adan and Huber (2011) build on top of this approach and focus on the modeling of wall surfaces. A learning mechanism is proposed for segmenting and labeling wall surfaces into parts either occluded by clutter, or occupied (solid), or empty (e.g. windows).

1.1.2. Volumetric primitive fitting

Xiao and Furukawa (2012) introduce a *constructive solid geometry* (CSG) based method to reconstruct large-scale indoor environments for visualization purposes. The authors propose a greedy algorithm for creating a CSG model, guided by an objective function devised to both measure quality and control detail. In order to reconstruct different floor and ceiling heights they first decompose the point cloud into horizontal slices and apply their method first in 2D, then in 3D to merge the 2D models. Although this approach can deal with more than two wall directions, it performs well on scenes mostly consisting of orthogonal or parallel structures. This limitation is mostly due to the primitive generation for the CSG model extraction: They detect linear wall segments and combine parallel or orthogonal ones for generating candidate primitives to be used by the CSG model.

Jenke et al. (2009) proposed a greedy algorithm reconstructing the free space volume. This approach is based on Schnabel's algorithm to detect planar primitives. The primitives are first structured in a graph, and a graph matching is applied to locate the cuboid candidates for the greedy reconstruction. Due to the fitting of cuboids their approach is limited to rectangular geometry. In presence of clutter or missing data the graph matching may fail as it requires at least five planar primitives to fit a cuboid.

1.1.3. Mesh-based reconstruction

Newcombe et al. (2011) and Izadi et al. (2011) propose a surface reconstruction method from point clouds acquired by a low cost consumer-grade hand held scanner. By designing their approach for parallel execution on GPUs they achieve real-time performance. Reconstructing an indoor scene using a general surface reconstruction method and presenting the result as a surface mesh allows for a general and detailed representation. However, no semantic classification of the scene into wall, floor, ceiling or clutter is provided. Other work devised to semantize indoor scenes (Kim et al., 2012; Shao et al., 2012), focus on understanding clutter in indoor environments.

1.2. Contributions and overview

Our approach reconstructs volumetric models of indoor spaces, by extending a previous work (Oesau et al., 2013). Oesau et al. (2013) partition the bounding box into cells using walls detected by a Hough transform. An empty/solid space labeling is formulated as an energy minimization and solved using graph-cut. The final model is extracted as the union of all cells labeled empty. In this work the level of detail of the reconstructed model is limited by the construction of the space partitioning. By introducing a multi-scale line fitting we enhance the alignment of the space partitioning with the permanent structures, which is particularly relevant for highly detailed structures. Experiments from defect-laden data and different sensors, including a scene recorded by a Kinect handheld sensor, are provided. Our new approach substantially increases the amount of details and overcomes some of the above mentioned limitations:

- *Arbitrary wall directions:* The model is not restricted to the Manhattan-World geometry and deals with planar wall detection for arbitrary vertical directions. The only assumption is that floors and ceilings are horizontal.
- *Multi-level buildings:* Our approach reconstructs an entire building with multiple levels in a single optimization step without requiring *a priori* knowledge about the levels.
- *Missing and outlier data:* 3D space partitioning into volumetric cells and labeling of the cells by global energy minimization provides resilience to missing data and outliers.
- *Raw data:* To be as general and applicable as possible, only dense raw point sets and knowledge about vertical direction are required. Nevertheless, when oriented normals or knowledge about the scanning device position are provided, they can be used to further improve robustness.

Our method takes as input a point cloud $P = \{p_1, \ldots, p_n\} \in R^3$ and consists of two main steps depicted in Fig. 1:

1. *Space partitioning:* The bounding box of P is partitioned into volumetric cells by using detected permanent structures as splitting planes. The vertical distribution of the point cloud is first analyzed in order to separate the points into slices containing horizontal/vertical structures and clutter. Each horizontal slice containing wall structures is then analyzed in 2D after vertical projection, in order to extract the wall line segments. Line extraction is performed through a multi-scale, feature-preserving line fitting method, followed by global clustering in a Hough transform space in order to favor alignment of the walls. Such alignment reduces the complexity of the volume partitioning.
2. *Surface extraction:* The volumetric cells created in previous step are labeled into either solid or empty space, respectively for permanent structures (walls, floors, ceilings) or outside. The final reconstructed surface is then deduced from the labeled cells.

2. Space partitioning

The goal of this step is to provide a 3D space partitioning of the bounding box of P into volumetric cells that well align to the empty and solid space. Using the permanent structures for partitioning the space provides such alignment. The following sub-steps describe the extraction of directions of horizontal and vertical permanent structures to create a 3D space partitioning.