ISPRS Journal of Photogrammetry and Remote Sensing 105 (2015) 286–304

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers

Martin Weinmann [a,*], Boris Jutzi [a], Stefan Hinz [a], Clément Mallet [b]

[a] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Englerstraße 7, 76131 Karlsruhe, Germany
[b] Université Paris-Est, IGN, SRIG, MATIS, 73 avenue de Paris, 94160 Saint-Mandé, France

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form 28 January 2015
Accepted 30 January 2015
Available online 27 February 2015

Keywords:
Point cloud
Neighborhood selection
Feature extraction
Feature selection
Classification
3D scene analysis

ABSTRACT

3D scene analysis in terms of automatically assigning 3D points a respective semantic label has become a topic of great importance in photogrammetry, remote sensing, computer vision and robotics. In this paper, we address the issue of how to increase the distinctiveness of geometric features and select the most relevant ones among these for 3D scene analysis. We present a new, fully automated and versatile framework composed of four components: (i) neighborhood selection, (ii) feature extraction, (iii) feature selection and (iv) classification. For each component, we consider a variety of approaches which allow applicability in terms of simplicity, efficiency and reproducibility, so that end-users can easily apply the different components and do not require expert knowledge in the respective domains. In a detailed evaluation involving 7 neighborhood definitions, 21 geometric features, 7 approaches for feature selection, 10 classifiers and 2 benchmark datasets, we demonstrate that the selection of optimal neighborhoods for individual 3D points significantly improves the results of 3D scene analysis. Additionally, we show that the selection of adequate feature subsets may even further increase the quality of the derived results while significantly reducing both processing time and memory consumption.

1. Introduction

Due to the increasing availability of 3D point cloud data and respective acquisition systems, the automated analysis of 3D point clouds has become a topic of great importance in photogrammetry, remote sensing, computer vision and robotics. Exploiting such data, recent investigations address a variety of different tasks such as the extraction of building structures (Vanegas et al., 2012), the recognition of power-line objects (Kim and Sohn, 2011), the extraction of roads and curbstones or road markings (Boyko and Funkhouser, 2011; Zhou and Vosselman, 2012; Guan et al., 2014), the mapping of vegetation (Wurm et al., 2014), the detection of numerous different objects (Kim and Medioni, 2011;Pu et al., 2011;Velizhev et al., 2012;Bremer et al., 2013;Serna and Marcotegui, 2014), the accessibility analysis in urban environments (Serna and Marcotegui, 2013), the creation of large-scale city models (Poullis and You, 2009; Lafarge and Mallet, 2012; Zhou and Neumann, 2013), or the semantic perception for ground robotics (Hebert et al., 2012). However, many of these tasks are based on the results of a 3D scene analysis in terms of uniquely assigning a semantic label (e.g. *ground*, *building* or *vegetation*) to each 3D point of a given point cloud.

When addressing the task of 3D scene analysis, we have to account for the general ideas shared by many respective approaches. Typically, 3D scene analysis involves (i) the recovery of a local neighborhood for each 3D point, (ii) the extraction of geometric features based on all 3D points within the local neighborhood, and (iii) the classification of all 3D points based on the respective features. Since often as many features as possible are exploited due to a lack of knowledge, recent investigations also addressed the selection of meaningful features as additional step between feature extraction and classification (Chehata et al., 2009; Mallet et al., 2011; Khoshelham and Oude Elberink, 2012; Weinmann et al., 2013; Weinmann et al., 2014). For all steps, however, a variety of challenges results from the complexity of 3D scenes caused by irregular point sampling, varying point density and very different types of objects. Furthermore, the computational burden arising from large 3D point clouds and many available features has to be taken into account.

In this paper, we focus on individual point classification, i.e. we only exploit feature vectors for assigning class labels, since respec-

* Corresponding author.
E-mail addresses: martin.weinmann@kit.edu (M. Weinmann), boris.jutzi@kit.edu (B. Jutzi), stefan.hinz@kit.edu (S. Hinz), clement.mallet@ign.fr (C. Mallet).

http://dx.doi.org/10.1016/j.isprsjprs.2015.01.016
0924-2716/

tive improvements also represent an important issue for methods involving contextual information. Besides revisiting foundations and trends in 3D scene analysis, we also provide new insights addressing all major steps in the respective data processing. These insights are based on our previous work involving feature relevance assessment (Weinmann et al., 2013), recovery of optimal 3D neighborhoods (Weinmann et al., 2014) and large-scale capability (Weinmann et al., 2015). Resulting from these investigations, we may easily derive a fully automatic, efficient and general framework for 3D scene analysis which involves

- neighborhoods of optimal size,
- low-level geometric 3D and 2D features,
- different strategies for feature selection, and
- efficient methods for supervised classification

while preserving both reproducibility and applicability of the involved methods. Hereby, we want to emphasize that neighborhood size selection and feature extraction are strongly interleaved issues, since the distinctiveness of geometric features strongly depends on the respective neighborhood encapsulating those 3D points which are taken into consideration for feature extraction. We further extend the framework by adding several approaches to different components, so that a variety of approaches is available for each component (Fig. 1). By providing a detailed evaluation involving two standard benchmark datasets, we are able to derive general statements on the suitability of the different approaches. Since only the spatial 3D geometry in terms of an appropriate representation of object surfaces as measured counterpart of the real world serves as input, our framework is generally applicable for interpreting 3D point cloud data obtained via different acquisition techniques such as terrestrial laser scanning (TLS), mobile laser scanning (MLS) or airborne laser scanning (ALS). Furthermore, the framework may be applied for point clouds captured with 3D cameras or point clouds obtained via 3D reconstruction from images.

In the following, we first reflect related work in Section 2. Subsequently, in Section 3, we explain the single components of our framework and respective methods in detail. For demonstrating the performance of our framework, we describe the involved publicly available datasets, the conducted experiments and the respective results in Section 4. Additionally, we discuss the derived results in Section 5. Finally, in Section 6, we provide concluding remarks and suggestions for future work.

2. Related work

In this section, we reflect the related work on 3D scene analysis and group respective approaches according to the single steps of 3D scene analysis.

2.1. Neighborhood selection

For being able to describe the local 3D structure around a given point $\mathbf{X}$ via geometric features, a respective neighborhood definition encapsulating all considered 3D points is required. Generally, different strategies may be applied for defining the local neighborhood $\mathcal{N}$ around a given 3D point $\mathbf{X}$. Among these, the most commonly applied neighborhood definitions are represented by

- a spherical neighborhood definition $\mathcal{N}_s$, where the neighborhood is formed by all 3D points in a sphere of fixed radius $r_s \in \mathbb{R}$ around the point $\mathbf{X}$ (Lee and Schenk, 2002),
- a cylindrical neighborhood definition $\mathcal{N}_c$, where the neighborhood is formed by all those 3D points whose 2D projections onto a plane (e.g. the ground plane) are within a circle of fixed radius $r_c \in \mathbb{R}$ around the projection of $\mathbf{X}$ (Filin and Pfeifer, 2005), and
- a neighborhood definition $\mathcal{N}_k$ based on a fixed number of the $k \in \mathbb{N}$ closest neighbors of $\mathbf{X}$ in 3D (Linsen and Prautzsch, 2001) or in 2D (Niemeyer et al., 2014).

Hereby, the third definition also results in a spherical neighborhood if 3D distances are evaluated for finding the closest neighbors, but – in contrast to the first definition – a variable absolute size is taken into account. Whereas these definitions with a constant scale parameter (i.e. either a fixed radius or a constant value k) across all 3D points provide a straightforward solution to neighborhood selection, it has to be taken into account that the scale parameter is typically selected with respect to heuristic or empiric knowledge on the scene and thus specific for each dataset. Furthermore, the scale parameter may not be identical across all considered 3D points, since it intuitively rather depends on the local 3D structure and point density. This holds particularly for MLS data, where due to the process of data acquisition dense and accurate 3D point clouds with significant variations in point density may be expected.

In order to avoid strong assumptions on local 3D neighborhoods, more recent investigations focused on introducing an optimal neighborhood size for each individual 3D point and thus increasing the distinctiveness of derived features. Most of the presented approaches exploit the idea of a neighborhood based on the k closest 3D points and optimize k for each individual 3D point. This optimization may for instance be based on iterative schemes relating neighborhood size to curvature, point density and noise of normal estimation (Mitra and Nguyen, 2003; Lalonde et al., 2005) which is particularly relevant for rather densely sampled and thus almost continuous surfaces. Other alternatives also account for more cluttered surface representations and are based on surface variation (Pauly et al., 2003; Belton and Lichti, 2006), dimensionality-based scale selection (Demantké et al., 2011) or eigenentropy-based scale selection (Weinmann et al., 2014). Even though deriving individual neighborhoods causes additional effort, the need for such concepts clearly becomes visible when considering the suitability of respective geometric features for neighborhoods of different size (Blomley et al., 2014) or the significant improvement in comparison to neighborhoods with a constant scale parameter across all 3D points (Weinmann et al., 2014).

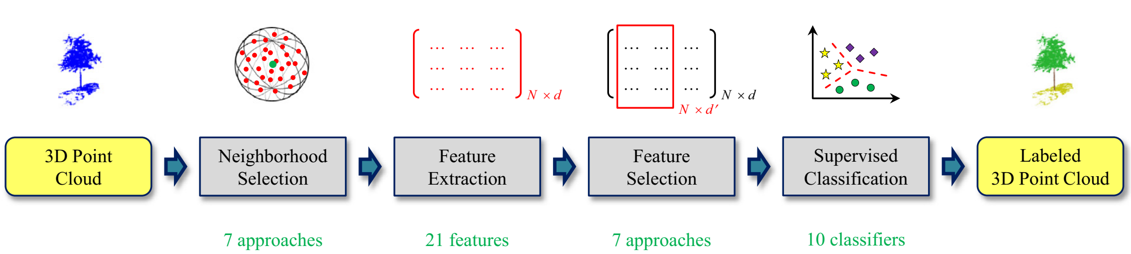

Fig. 1. The proposed framework and the quantity of attributes/approaches taken into account for evaluation.