Automation in Construction 86 (2018) 44–52

Contents lists available at ScienceDirect

Automation in Construction

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

Performance evaluation of 3D descriptors for object recognition in construction applications

Jingdao Chen[a], Yihai Fang[b], Yong K. Cho[c,*]

[a] Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, 801 Atlantic Dr. N.W., Atlanta, GA 30332, United States
[b] Department of Civil Engineering, Monash University, 23 College Walk, VIC 3800, Australia
[c] School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr. N.W., Atlanta, GA 30332-0355, United States

ARTICLE INFO

Keywords:
3D object recognition
Point cloud
Machine learning
Descriptor

ABSTRACT

3D object recognition from field-acquired point cloud data is important for modeling, manipulation, visualization and other post-processing tasks in the construction domain. However, building semantically-rich models from raw point cloud data is a difficult task due to the high volume of unstructured information as well as confounding factors such as noise and occlusion. Although there exist several computational recognition methods available, their performance robustness for construction applications are not well known. Therefore, this research aims to review and evaluate state-of-the-art descriptors for 3D object recognition from raw point clouds for construction applications such as workspace modeling, asset management and worker tracking. The evaluation was carried out using 3D CAD models with known labels as training data and laser-scanned point clouds from construction sites as testing data. The recognition performance was evaluated with respect to varying level of detail, noise level, degree of occlusion, and computation time. Experimental results show that for all evaluated descriptors, increasing the level of detail and decreasing the noise level results in a moderate increase in recognition accuracy whereas reducing occlusion results in a significant increase in recognition accuracy. In addition, experimental results suggest that the key features that distinguish an object can be derived around the 10 mm level and any further increase in the level of detail do not significantly increase the recognition accuracy.

1. Introduction

3D point clouds from laser scanning or photogrammetry are widely used to capture the as-built status of a construction jobsite. Point clouds usually consist of millions of points stored in an unstructured format; it is difficult for human agents and 3D modeling software to interpret and analyze the data. Automated object recognition techniques are important to recover contextual information and high-level semantics from raw point cloud data. The result of recognition tasks is relevant for applications such as safety monitoring [1,2], energy analysis [3], defect identification [4], inventory tracking [5], Building Information Modeling [6,7], and workspace modeling for equipment operation [8].

Using 3D CAD models as priors, Bosche and Haas [9] presented a semi-automated approach to recognize building objects (e.g., slabs, columns) from laser scanned point clouds. When the CAD models are extracted from a Building Information Model (BIM), the methods generally referred to "Scan-to-BIM" or "Scan-vs-BIM" [10,11]. By comparing the as-built models recognized from point cloud to the as-design BIM model, dimension discrepancy can be calculated and referred to the corresponding tolerances for construction quality control purpose [4,12]. Combining the 3D object recognition approach with schedule information, Turkan et al. [13] proposed a 4D object oriented construction progress tracking system for both permanent structural objects and secondary and temporary structures [14]. With a focus on energy simulation, Wang et al. [3] proposed an automated approach for extracting building geometries as individual objects and visualize the object as polygons from unorganized point cloud. Kim et al. used curvature information to perform automatic segmentation and 3D modeling of as-built pipelines [15–19]. In addition to building structural objects, recognizing other construction assets such as equipment, material, and foliage are of great value. To facilitate road safety inspection, Pu et al. [5] presented a recognition method that recognizes critical objects (e.g., traffic signs, barriers, trees) from mobile laser point clouds. To assist equipment operation in real-time, Cho and Gai [20] proposed a Projection–Recognition–Projection (PRP) method for rapid 3D modeling of construction equipment workspace using a pre-defined library of target objects.

Despite of the pressing demands of recognizing construction objects

* Corresponding author.
E-mail addresses: jchen490@gatech.edu (J. Chen), yihai.fang@monash.edu (Y. Fang), yong.cho@ce.gatech.edu (Y.K. Cho).

https://doi.org/10.1016/j.autcon.2017.10.033
Received 1 May 2017; Received in revised form 23 September 2017; Accepted 31 October 2017
Available online 09 November 2017

from point cloud data, existing approaches and tool sets cannot fully address the challenges such as low recognition rate with incomplete or noisy point cloud data and the need for supervision in the recognition process. One possible approach for object recognition from point clouds is to use descriptor-based recognition methods. A descriptor is a vector of features that allow shape retrieval or correspondence finding algorithms to uniquely identify a keypoint or an object. Using a machine learning framework, a novel object can be classified into a set of learned models based on its descriptors. 2D descriptors have been researched extensively in the computer vision field [21,22], yet the use of 3D descriptors for object recognition from point clouds has not been widely studied, especially in the civil engineering domain. The datasets used for studying 3D descriptors usually involve small objects obtained in a lab setting and there are few existing comparison studies that quantify the effectiveness of different descriptor-based recognition methods in complex, real-world environments. Therefore, this research aims to present a review and evaluation of state-of-the-art descriptors for 3D object recognition for construction applications. In this study, five different local and global 3D descriptors were evaluated on the object recognition task for laser-scanned point clouds obtained from construction sites. The five selected descriptors represent popular approaches to 3D object recognition available in the literature encompassing gradient-based and shape distribution-based approaches. The recognition performance was compared with respect to varying level of detail, noise level, and degree of occlusion. The rest of this paper will present, in order, the literature review, methodology, results, discussion, and conclusions.

2. Literature review

The use of laser scan data for as-built jobsite modeling and planning has been well studied in the literature [5,23]. Despite the importance of automatically extracting CAD objects from point clouds, there is still a lack of research into descriptor-based recognition that can adequately handle problems in raw data acquired from a construction environment such as outliers, noise, and missing data [6]. This study aims to address this knowledge gap by conducting a performance evaluation of 3D descriptors for object recognition. 3D descriptors can be generally divided into two families, namely local descriptors and global descriptors. Local descriptors are more commonly used for instance recognition tasks whereas global descriptors are more commonly used for classification tasks [23].

Local descriptors characterize the neighborhood of a keypoint in terms of local features such as curvature and gradients. Similar to many 2D descriptors, the technique of histogram binning is used to discretize the feature space. A standard descriptor used in the literature is the spin image [24], where the set of points in the neighborhood of the keypoint of interest is mapped to an image-like grid. The spin image encapsulates the surface features for each keypoint of an object, thus a novel object can be recognized if its spin images can be matched to spin images in an existing model database. This idea is further expanded by the Signature of Histograms of Orientations (SHOT) [25], which builds a histogram of the gradients in a region of support around a keypoint instead of the point distances. A separate idea is to construct a spherical reference frame around the keypoint and accumulate neighboring points in logarithmically-spaced subdivision. This technique is used in descriptors such as 3D Shape Context [26] and Unique Shape Context [27].

On the other hand, global descriptors summarize the geometry of an entire object in a single feature vector. Global descriptors are able to capture features relevant to the complete object geometry instead of a specific area, but require precise segmentation techniques to extract the object from the raw scan data. One approach to constructing a global descriptor is to calculate the distribution of geometric properties of points sampled throughout the object surface. For example, the Ensemble of Shape Functions (ESF) [28] uses the distributions of

Table 1
Comparison of properties between different 3D descriptors in the literature.

Descriptor	SPIN [24]	SHOT [25]	USC [27]	ESF [28]	PAD [30]
Region of support	Local	Local	Local	Global	Global
Feature dimension	153	361	1969	640	10
Uses histogram binning	Yes	Yes	Yes	Yes	No
Uses point normals	Yes	Yes	Yes	No	No
Reference frame	Cylindrical	Spherical	Spherical	None	Cartesian

distances, areas, and angles to distinguish between different objects. A different approach is to measure the relative pan, tilt, and yaw angles between the point normals and a selected viewpoint direction, used by the Viewpoint Feature Histogram [29]. Alternatively, the Principal Axes Descriptor could be used, where an occupancy grid over the object points can be constructed based on principal component analysis, from which occupancy ratios can be derived and used as features [30].

For this study, a set of five descriptors were selected which represent popular approaches to 3D object recognition available in the literature: (i) spin images (SPIN), (ii) Signature of Histograms of Orientations (SHOT), (iii) Unique Shape Context (USC), (iv) Ensemble of Shape Functions (ESF), and (v) Principal Axes Descriptor (PAD). The five descriptors were selected to compare between a mix of different methods for deriving features from a 3D point cloud (e.g. local vs. global, spherical vs. Cartesian reference frame). A summary of the properties of the different local and global 3D descriptors is given in Table 1.

Several works in the literature have attempted to identify the advantages and disadvantages of different 3D descriptors. Campbell and Flynn [31] surveyed different recognition systems for free-form objects in relation to the 3D model building process. Bronstein and Bronstein [32] also examined different feature descriptors for the task of 3D shape retrieval and correspondence finding. Guo et al. [33] presented a survey of local surface features and identified weaknesses such as sensitivity to occlusion, deformation, and point density variation. In terms of experimental work, Salti et al. [34] performed an experimental evaluation for 3D object recognition in terms of robustness to noise, clutter, occlusion, and viewpoint variation. Similarly, Arbeiter et al. [35] evaluated the performance of three local descriptors for different primitive surfaces such as cylinders, edges, and corners. Most of these works concentrate on local descriptors and did not compare their performance to that of global descriptors for object recognition. In addition, there has not yet been any comparison studies using laser-scanned data collected from the field that is subject to occlusion and sensor noise in a real-world setting.

3. Methodology

On overview of the overall experimental design is shown in Fig. 1. Detailed explanations of the steps for data acquisition and pre-processing, object recognition framework, and varying the experimental parameters will be presented in the following subsections.

3.1. Data acquisition and pre-processing

The 3D object dataset for this study consists of a training dataset, which contains objects with known labels and used to train machine learning classifiers, and a testing dataset, which is used to evaluate the recognition performance of descriptors. The dataset consists of five classes of objects commonly encountered on a construction site, namely (i) trailer, (ii) truck, (iii) worker, (iv) crane, and (v) excavator. The training dataset, shown in Fig. 2, is acquired from 3D CAD models downloaded from online repositories. The 3D models in mesh format are converted to a point cloud format using a ray tracing algorithm which works by placing virtual laser scanners around the object. The