Principal Axes Descriptor for Automated Construction-Equipment Classification from Point Clouds

Jingdao Chen[1]; Yihai Fang[2]; Yong K. Cho[3]; and Changwan Kim[4]

Abstract: Recognizing construction assets (e.g., materials, equipment, labor) from point cloud data of construction environments provides essential information for engineering and management applications including progress monitoring, safety management, supply-chain management, and quality control. This study introduces a novel principal axes descriptor (PAD) for construction-equipment classification from point cloud data. Scattered as-is point clouds are first processed with downsampling, segmentation, and clustering steps to obtain individual instances of construction equipment. A geometric descriptor consisting of dimensional variation, occupancy distribution, shape profile, and plane counting features is then calculated to encode three-dimensional (3D) characteristics of each equipment category. Using the derived features, machine learning methods such as k-nearest neighbors and support vector machine are employed to determine class membership among major construction-equipment categories such as backhoe loader, bulldozer, dump truck, excavator, and front loader. Construction-equipment classification with the proposed PAD was validated using computer-aided design (CAD)–generated point clouds as training data and laser-scanned point clouds from an equipment yard as testing data. The recognition performance was further evaluated using point clouds from a construction site as well as a pose variation data set. PAD was shown to achieve a higher recall rate and lower computation time compared to competing 3D descriptors. The results indicate that the proposed descriptor is a viable solution for construction-equipment classification from point cloud data. **DOI: 10.1061/(ASCE)CP.1943-5487.0000628.**

Author keywords: Object recognition; Object classification; Scattered point clouds; Laser scanning; Machine learning.

Introduction

In past decades, construction practices and research have been actively embracing emerging technologies to improve project performance in productivity, quality, and safety. The technical categories that have been extensively studied and tested in the construction industry include data sensing, simulation, information modeling, and visualization. From these categories, as-built data sensing and visualization are considered by many industry practitioners and academia experts as some of the most promising technologies that will greatly expand the utilization of site information (Tang et al. 2010a; Bosché 2010). For instance, three-dimensional (3D) point clouds produced by laser scanners and other data acquisition technologies such as photogrammetry have been widely used for acquiring and generating as-built site information to support applications such as construction quality assessment and control (Bu and Zhang 2008; Ahmed et al. 2012), construction progress tracking (Golparvar-Fard et al. 2009; Turkan et al. 2012), building energy analysis (Wang et al. 2015a; Ham and Golparvar-Fard 2013), construction hazard recognition (Fekete et al. 2010; Wang et al. 2015b), structural health monitoring (Huber 2014), and highway asset management (Gong et al. 2012; Pu et al. 2011). Each application has a different focus of target objects (e.g., building components, materials, equipment, labors, and traffic signs) on the as-built point clouds. Thus, it is important to recognize the objects of interest from the large and complex point cloud. This process, however, is very challenging in unstructured construction environments.

Regardless of the application, recognizing objects from point clouds is often a key step for post-data processing and analyses. Although object classification is a well-studied problem in the field of computer vision, classifying objects from point clouds of uncontrolled environments such as construction sites remains a challenging and unsolved task. Many previous algorithms for object classification were validated based on complete scans of small objects (e.g., fruit, household objects) obtained in controlled laboratory settings (Marton et al. 2010b; Rusu et al. 2010; Wohlkinger and Vincze 2011). In uncontrolled settings with the presence of occlusion and other objects that share similar features with the target object, many state-of-the-art recognition algorithms fail to maintain a good recognition rate (Wohlkinger and Vincze 2011). These problems become even more serious when acquiring as-built 3D data from construction sites where the presence of huge building structures and a large amount of equipment and materials introduce potential occlusions to the point cloud acquired regardless of what technology is used. As construction projects become larger and more complex, obtaining complete 3D data becomes even more difficult (Kim et al. 2013). In addition, similarities among different types of construction assets, especially equipment (e.g., bulldozer versus loader), present a formidable challenge to classification algorithms that aim to distinguish between them.

[1]Ph.D. Student, School of Electrical and Computer Engineering, Robotics Institute, Georgia Institute of Technology, 777 Atlanta Dr. N.W., Atlanta, GA 30332-0355. E-mail: jchen490@gatech.edu

[2]Ph.D. Candidate, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlanta Dr., Atlanta, GA 30332-0355. E-mail: yihaifang@gatech.edu

[3]Associate Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlanta Dr. N.W., Atlanta, GA 30332-0355 (corresponding author). E-mail: yong.cho@ce.gatech.edu

[4]Professor, Dept. of Architectural Engineering, Chung-Ang Univ., 221 Heuksuk-dong, Dongjak-gu, Seoul 156-756, Korea. E-mail: changwan@cau.ac.kr

Note. This manuscript was submitted on April 12, 2016; approved on July 11, 2016; published online on September 26, 2016. Discussion period open until February 26, 2017; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

This study proposes the novel principal axes descriptor (PAD) for construction-equipment classification from point cloud data. The approach proposed in this research can be used to recognize various construction assets with complex shapes (e.g., materials, temporary structures, equipment). To demonstrate the methodology and performance of the recognition method, this study choses construction equipment as the initial target objects because automatically describing and distinguishing each equipment class is technically challenging due to their complexity and similarity in shape. This paper starts with a thorough literature review on as-built data acquisition, postprocessing, and classification, followed by a detailed description of the methodology applied in the proposed approach. The validation of the proposed approach on three test data sets is introduced, and the limitations and future work on the developed method are discussed.

Literature Review

Construction asset recognition and classification include multiple steps that involve techniques in digital data acquisition, feature description, and machine learning. This section will review the state of the art in these relevant fields of study.

As-Built Data Acquisition in Construction

Three-dimensional as-built data provide additional information and flexibility for visualization, as-built modeling, and other applications that heavily depend on spatial data. The most popular technologies for obtaining as-built 3D data on construction sites are 3D range imaging camera, photogrammetry, and laser scanning. This section will briefly review the characteristics of these three technologies and their applications in construction environments.

Based on different techniques such as stereo triangulation, structured light, and time of flight, a 3D range imaging camera produces a range image with pixel values representing the intensity and range. Although the 3D range imaging camera has been widely applied in the fields of virtual reality and gaming, very few investigations of its benefits in construction have been undertaken. Such research focused on range data processing for real-time site spatial modeling (Gong and Caldas 2008). Compared to laser scanners, 3D range imaging cameras are more portable and less expensive, but they operate in a much closer range and do not provide reliable range images when direct sunlight is present, which makes it difficult to apply on large construction sites (Park et al. 2012; Remondino and Stoppa 2013).

Photogrammetry is an image-based technology that reconstructs 3D objects from two-dimensional (2D) photographs. This technology extracts 2D input data from photographs and maps them onto a 3D space. Because constructing a 3D model only requires taking images from different angles, using photogrammetry for 3D data acquisition is flexible, cost-effective, and noninvasive to the survey objects. In the field of construction engineering, photogrammetry has been applied to acquiring as-is geometry data for building component modeling (Dai and Lu 2010), construction progress monitoring and control (El-Omari and Moselhi 2008), and construction documentation especially for historical structures (Yastikli 2007). However, several limitations hinder the wide adoption of photogrammetry in construction engineering. First, photography is sensitive to lighting and weather conditions, which makes it difficult to perform reliable analysis (Cho and Gai 2014; Cho et al. 2012). Second, it will not work for the scenes that have no distinctive features such as simple plane walls (Wang et al. 2015a). Furthermore, it involves longer processing times and obtains data with reduced accuracy compared to laser scanning (Dai et al. 2012).

Laser scanning is a nondestructive technology that can rapidly and accurately capture the shape of physical objects based on time of flight or phase-shift principle. A scanned object or scene is represented by a dense point cloud comprised of millions of points, each of which contains position (XYZ), color (red/green/blue), and/or intensity data. This technology, therefore, is suitable for a wide range of measurement at high resolution. Compared to photogrammetry, laser scanning is far less sensitive to light conditions and usually generates a more accurate and dense point cloud of the scanned objects in a large scene (Cho and Gai 2014; Wang and Cho 2015). However, it suffers from high equipment cost, size, and extended time required for data collection compared with 3D range imaging camera and photogrammetry. Due to the frequent presence of occlusion on a construction site, obtaining a full scan of the desired objects or scene requires multiple scans from different locations.

While understanding the positive and negative aspects of each data acquisition technology in a construction environment is important for real-time applications, it is also vital to consider the difficulties of data processing and extracting useful semantic information out of as-built data, regardless of the data acquisition methods.

Object Recognition from a Point Cloud

The typical pipeline of point cloud processing for object recognition consists of the following steps. First, shape descriptors are computed for each object instance or class to be recognized and stored in an offline database. Second, point cloud data from a processed scene is segmented into candidate object instances and query descriptors are computed for each object instance. Finally, objects from the model database with high descriptor similarity are aligned with the objects recognized in the point cloud to produce a match (Tang et al. 2010b). A major disadvantage of instance-based object recognition is that it is not as applicable and effective when significant shape variance exists among objects (Tang et al. 2010b). The most common approach in class-based object recognition is to use global shape descriptors, which are less discriminative than semilocal descriptors but better adapted to shape variance. However, a major challenge is that this approach cannot handle a partial or incomplete point cloud caused by occlusion or clutter (Ruiz-Correa et al. 2006). To resolve this problem, one solution is to use descriptors that are more robust to shape changes (Ruiz-Correa et al. 2006).

Mobile Construction Assets Recognition

In addition to obtaining as-built condition of building components, point clouds can be used to recognize mobile construction assets such as workers and equipment. Cho and Gai (2014) proposed a projection-recognition-projection (PRP) method to automatically recognize construction equipment from a point cloud and fitting and matching a complete 3D model to the corresponding point cloud in near-real time. The 3D point cloud is projected to a 2D space where the geometric features represented by a local speeded up robust features (SURF) descriptor are compared to a prepared template database for recognition. This method is very effective and efficient for recognizing target objects that are known to be present on the construction site. For unknown objects with high shape variance, however, the performance of this method is limited.

State-of-the-Art Geometric Descriptors

A different approach to object recognition in the 3D domain is to compute point features directly from the point cloud and store

feature values in a vector representation known as a descriptor. This study will focus on global 3D descriptors because they are better suited for class-based recognition compared with local 3D descriptors, which are better suited for instance-based recognition. The following sections discuss three different global descriptors used in recent research representing different approaches to feature calculation, namely, normal-based, radius-based, and shape distributions.

Viewpoint Feature Histogram Descriptor

An example of a local point descriptor is the point feature histogram (Rusu 2009), where normal vectors are computed around the neighborhood of a selected key point and the differences between normal vectors are binned into a histogram. The resulting histogram provides a high-dimensional representation of the surface variation around the key point. For the case where a point cloud cluster representing a complete object is available, the method can be extended to a global shape descriptor. Rusu et al. (2010) proposed a viewpoint feature histogram (VFH), which consists of a viewpoint component and an extended fast-point feature histogram component computed at the object's centroid. The descriptor was designed for identification and pose estimation of kitchenware objects from stereo camera–generated point clouds. The descriptor performed well in object recognition tasks despite noisy sensor data and subtle variation in object geometry, but is not invariant to viewpoint.

Global Radius-Based Surface Descriptor

Surface categorization based on curvature is another useful tool in 3D object classification. The radius-based surface descriptor (RSD) stores the minimum and maximum radii of curvature around the local neighborhood of a point (Marton et al. 2010a). This information can be used to categorize surfaces into planes, cylinders, edges, rims, and spheres. Similiar to the VFH case, a global descriptor can be derived from this local descriptor using the same features. The global radius-based surface descriptor (GRSD) first computes a local surface categorization from RSD for each point and determines the global descriptor values by constructing a histogram of relationships between local surface labels. GRSD is shown to achieve good classication result when used together with visual features in a two layer classification scheme (Marton et al. 2010b).

Ensemble of Shape Functions Descriptor

The concept of shape distribution was introduced by Osada et al. (2001) to describe the geometric signature of an object. Shape distributions are sampled from multiple shape functions that measure global geometric properties of an object. The distance between two randomly chosen points (D2), for instance, is a robust shape function for distinguishing different object classes. The performance of this method can be compromised when the number of classes to be recognized increases or only a partial view of the object is available in the point cloud. Wohlkinger and Vincze (2011) proposed a robust global shape descriptor called the ensemble of shape function (ESF) for recognizing a variety of object classes (e.g., mugs, fruits, cars) from point clouds. This descriptor specifically addresses the challenge of using partial views in incomplete point clouds for object recognition through construction of three distinct shape functions describing distance, angle, and area distributions. Each shape distribution is represented by an ESF histogram that can be used to efficiently retrieve the k-closest matches from a predefined class database using the k-nearest neighbor algorithm (k-NN). Although the ESF descriptor shows fairly good recognition rates for small-scale objects such as apples (98.45%) and mugs (99.46%), it shows limited capability in recognizing large objects such as cars (43.64%).

Machine Learning Techniques in Object Classification

In addition to a robust description of objects, another critical component in object recognition is training a classification algorithm. Typically, the training process consists of the following four steps: (1) gather a training set that consists of a number of point clouds of the object classes that need to be recognized and label the point clouds with desired classifications; (2) determine a descriptor to represent the objects in the training set; (3) select the structure of the learned function and corresponding learning algorithm [e.g., support vector machine (SVM), k-nearest neighbor, decision tree]; and (4) determine training parameters and run the learning algorithm on the gathered training set. Thus, a new input point cloud can be recognized by mapping its descriptor with the function inferred by the previously labeled training data. Similar to other applications adopting supervised learning, the availability and collection of training data is a bottleneck due to the lack of a centralized repository for point clouds. This problem can be addressed by using synthetic point clouds generated from 3D computer-aided design (CAD) models (Wohlkinger and Vincze 2011).

Methodology

Framework

To improve on existing techniques for object recognition in point clouds, this study introduces a novel 3D descriptor, PAD, for construction-equipment recognition from scattered as-is point clouds. Fig. 1 shows the processing pipeline for the proposed construction-equipment recognition approach using PAD. The left column represents the training stage where synthetic point clouds were generated from CAD models to create a training database. The right column, which represents the application of the method to an actual laser-scanned point cloud, is in turn divided into a point cloud preprocessing stage and a classification stage. The point cloud preprocessing stage is concerned with segmentation and

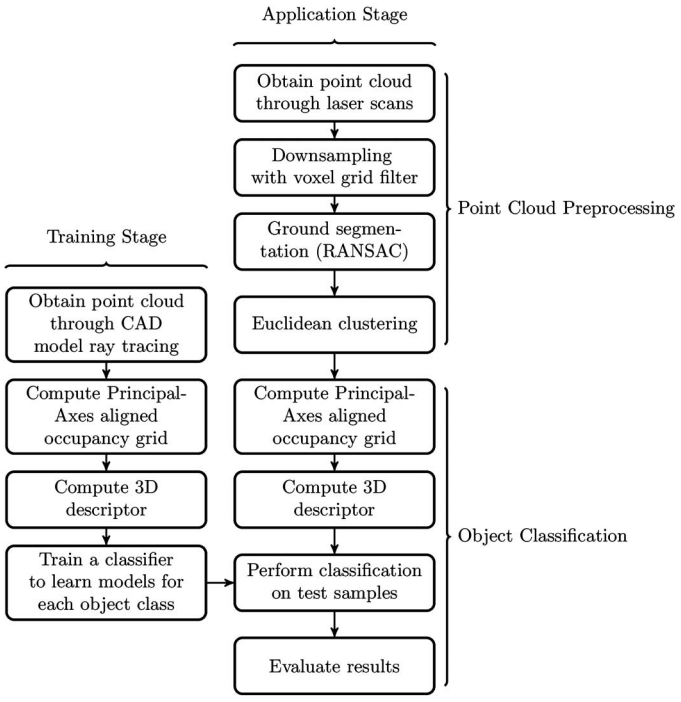

Fig. 1. Flowchart for object recognition

clustering, which enables the separation of targeted assets in the foreground from background points. The classification stage is concerned with computing 3D descriptors for each construction-equipment instance and classifying them into predefined categories such as excavator and loader based on the derived descriptor. The two stages are further discussed in detail in the following sections.

Point Cloud Segmentation and Clustering from Laser Scans

Laser-scanned point clouds from construction sites often consist of millions of points, which renders the data processing step computationally demanding. Thus, a voxel grid representation is used to store point cloud data in a condensed form in order to achieve efficient segmentation and clustering. Next, a segmentation technique is employed to filter out background points so that foreground points containing the targeted objects can be properly identified. Ground segmentation is especially pertinent because ground points usually make up a majority of the background points in an outdoor scene. The technique used in this study is an iterative ground erosion technique as illustrated in Fig. 2. The ground points, shown as dots, can either be flat, recessed, or protruded, as is typical on a construction site. To overcome the global nonuniformity in ground points, the scene is first divided into local regions, shown as vertical dashed lines. For each local region, the minimum z-coordinate of all the points in the region is identified and the corresponding point is filtered out (shown as crosses). This process is repeated over several iterations until the ground points are gradually eroded away. Finally, with the ground points filtered out, the Euclidean distance metric (Rusu 2009) can be used to incrementally combine neighboring points into point cloud clusters for each construction-equipment instance present in the acquired data. This is represented in Fig. 2(b) as bounding boxes around each cluster of object points. Because the computed bounding boxes are computed for all remaining points, which may include spurious background objects, an additional size filter is applied on the bounding boxes to remove objects that are too small or too large. The dimensions of the bounding box filter are determined based on the acquired training data.

Object Classification

The classification stage involves generating a training database of point cloud samples from construction equipment with known categorization and matching query point cloud samples to the appropriate category based on computed shape descriptors. Further details concerning the classification process are discussed in the following sections.

Synthetic Point Cloud Generation

Sample 3D CAD models of the desired classification categories were downloaded online from *3D Warehouse*. In this study, construction-equipment samples were collected from the backhoe loader, bulldozer, dump truck, excavator (backhoe), and front loader categories (Fig. 3). The downloaded CAD models were then converted into point clouds to form the training data set. A ray casting technique was used to sample points along the model surface with respect to virtual laser scanners placed at multiple view locations around the model. In this study, the scan origins are distributed evenly in 16 different locations around each model and the

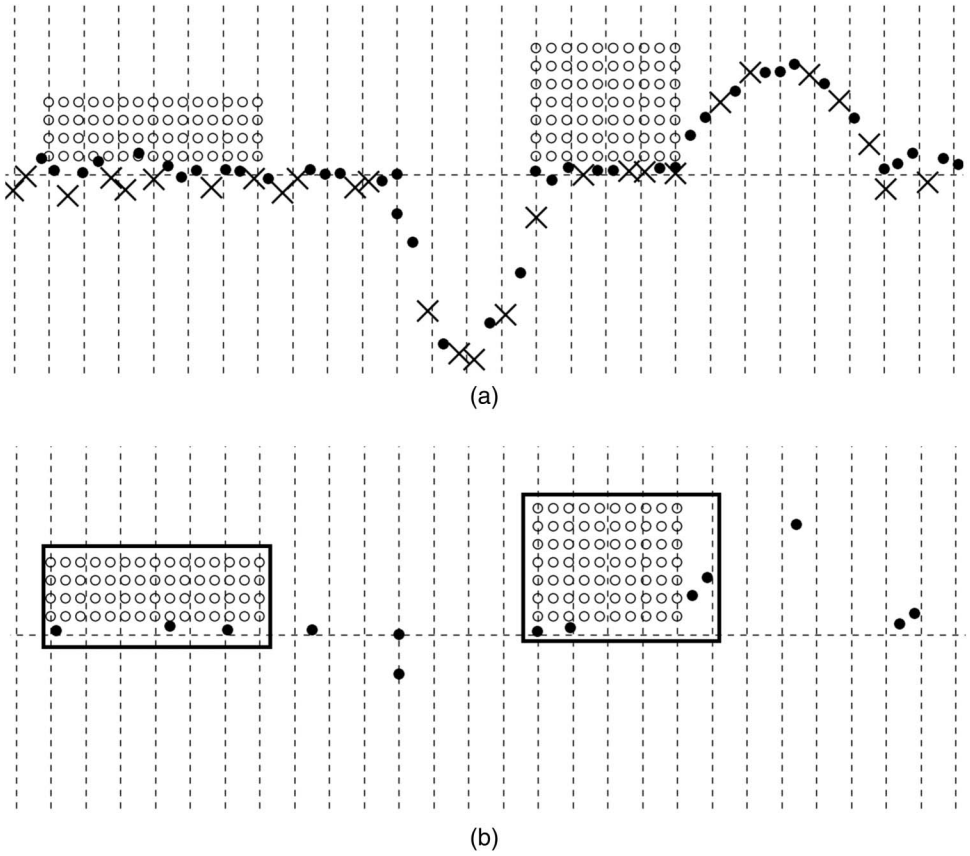

Fig. 2. Point cloud segmentation and clustering: (a) ground erosion; (b) Euclidean clustering; circles represent object points while dots represent ground points; crosses are ground points that have been filtered out

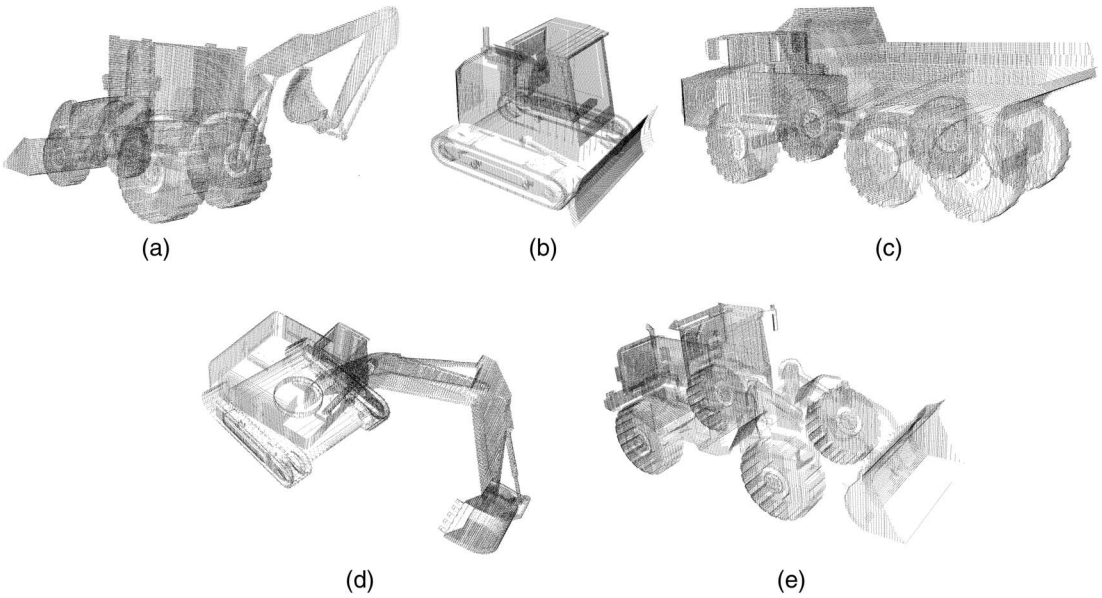

Fig. 3. Synthetic point clouds of construction equipment: (a) backhoe loader; (b) bulldozer; (c) dump truck; (d) excavator; (e) front loader

scan lines are separated by an angular resolution of 0.005 rad. To increase the amount of training data available for the classification process, common data augmentation techniques from machine learning are used. For each equipment model, the generated point clouds were randomly perturbed with Gaussian white noise to simulate the data collection process in real-world scenarios. In addition, the view angle of each virtual laser scanner was also artificially restricted by a random amount to simulate occlusion. Furthermore, CAD design tools were used to vary the joint angles for equipment with moving parts to handle cases where construction equipment may be present in different poses. This randomized process allows multiple point cloud samples to be generated from the same construction-equipment model. The artificial perturbation of training data is a form of regularization which allows the classifier to be robust to data variability within each equipment class. It is often the case that the test data do not have exactly the same dimensions and descriptor values as the training data even though they belong to the same class. Artificial perturbations are necessary to ensure that in the process of learning the features for each equipment category, the general trend of feature values are identified instead of specific values.

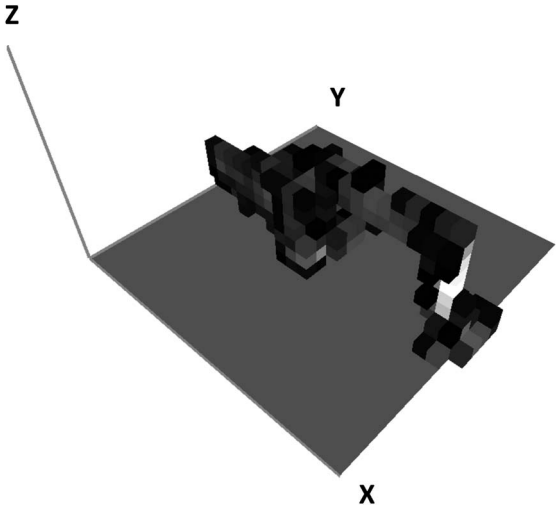

Fig. 4. Principal axes–aligned occupancy grid for an excavator

Descriptor Calculation

The key idea behind PAD is to leverage the rectangular structure and line symmetry of most construction equipment. Different classes of construction equipment can thus be identified by drawing discriminating features from these major axes of symmetry, or principal axes. The technique of principal component analysis (PCA) is commonly used to derive surface normal and curvature values for 3D objects (Rusu 2009). Here, however, PCA is used to infer direction vectors along which certain geometric features may be identified. The first step in the descriptor calculation is to form a 3D occupancy grid with grid axes aligned with the principal axes of the point cloud under consideration. The occupancy grid is generated uniformly with the shortest dimension having a size of nine grid elements. Thus, it is scale-independent with respect to the original point cloud but the relative scale between axes is maintained. Fig. 4 shows an example of the 3D occupancy grid for an excavator in which darker grid locations indicate a high density of points and lighter grid locations indicate a lower density of points.

Next, features are derived based on the occupancy grid to store information from dimensional variation, occupancy distribution, shape profile, and plane counting. Dimensional variation captures the distribution of length, width, and height between different equipment categories. Height is defined as the dimension of grid elements along the z-axis, which is the principal axis perpendicular to the ground. Of the remaining axes, length is defined as the longer principal axis, whereas width is defined as the shorter principal axis. The first two features of the descriptor are calculated as the ratio between these dimensions. This helps discriminate one from other equipment categories because some equipment, like dump trucks, has large length relative to height, whereas for bulldozers the value would be lower. To determine occupancy distribution, the occupancy grid is split along the middle for each of the length, width, and height dimensions and the ratio between occupied grid cells of each half is taken. This can be used to identify equipment such as excavators, which have less cells distributed in front due to the boom. Because PCA is not deterministic with respect to the

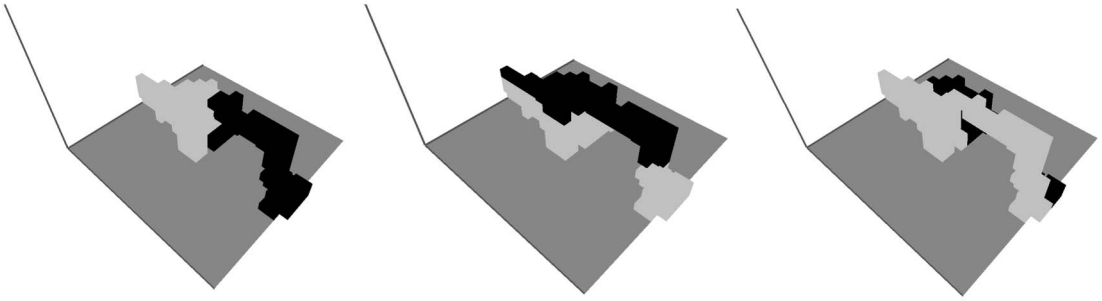

Fig. 5. Occupancy distribution features: (a) front to back; (b) top to bottom; (c) left to right

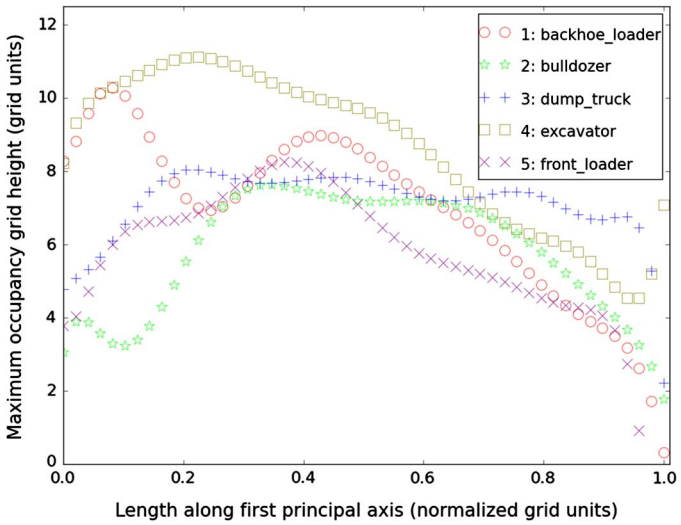

Fig. 6. Average shape profile of different equipment categories

Algorithm 1 Principal Axes Descriptor Algorithm

```
procedure GETFEATURES
                                                    ▷ Dimensional Variation
    feature(1) ← length / min(width,height)
    feature(2) ← height / width
                                                    ▷ Occupancy Distribution
    front ← occupancy[x > mid]
    back ← occupancy[x < mid]
    left ← occupancy[y < mid]
    right ← occupancy[y > mid]
    top ← occupancy[z > mid]
    bottom ← occupancy[z < mid]
    feature(3) ← front : back
    feature(4) ← mid : (front + back)
    feature(5) ← mid : (left + right)
    feature(6) ← top : bottom
                                                    ▷ Shape Profile
    profile ← z for x = 1 → length
    peak ← max(profile)
    feature(7) ← gradient(peak)
    feature(8) ← peak : length
                                                    ▷ Plane Counting
    feature(9) ← count(plane,size < min(width,height)/2)
    feature(10) ← count(plane,z < mid)
    return feature
end procedure
```

Fig. 7. PAD calculation procedure

directions of the principal axes, it is necessary to normalize the feature values by taking the ratio between minimum and maximum values of each configuration. An example of the occupancy distribution features is shown in Fig. 5, where the ratio of occupied grid cells is calculated between two subregions of the grid (bright cells versus dark cells). On the other hand, the shape profile of an equipment describes the maximum height of occupied grid cells along its length. The shape profile can be plotted as shown in Fig. 6, where the x-axis is placed along the length of the equipment and the y-axis is placed along the height of the equipment. The x-axis is calculated in normalized units (0 to 1) of the occupancy grid to account for differences in length between equipment of the same class, whereas the y-axis is calculated in unnormalized grid units to demonstrate the height variation between equipment classes. Additional features can thus be derived by calculating the gradient to the peak along the profile and the location of the peak with respect to the length.

The last two features are plane counting features designed to detect specific components of construction equipment. For example, the likelihood of an excavator instance can be measured by counting the number of planes along the occupancy grid below a certain size threshold, which would indicate a boom, whereas the likelihood of a loader instance can be measured by counting the number of planes below a certain height threshold, which would indicate a bucket. The PAD algorithm calculation procedure is summarized in Fig. 7.

For comparison purposes, VFH, GRSD, and ESF descriptors were also calculated through the implementation in Point Cloud Library (Rusu and Cousins 2011) to evaluate the classification performance. Default parameters were used in the ESF calculation, whereas for VFH a normal estimation radius of 1.0 m was used. For the GRSD calculation, the radius for normal estimation was set to 0.5 m and the radius for neighborhood search was set to 0.85 m. The radii parameters were selected to be approximately an order of magnitude of the sampled point cloud resolution. Several radii values were tested and the best option in terms of classification result was reported.

Machine Learning Classification Techniques

In the final stage of object classification, class labels are assigned to each input point cloud cluster by training and applying machine learning classifiers. The classification approaches considered in this study are the k-nearest neighbor (Altman 1992), linear discriminant analysis (LDA) (Hastie et al. 2008), logistic regression (Lin et al. 2011), and support vector machine (Chang and Lin 2011). The k-NN method involves picking k training samples closest to the test sample in terms of Euclidean distance in descriptor space. The output class label is then determined through uniform voting over class labels of the k nearest training samples. Linear

discriminant analysis fits a Gaussian density model with equal covariance to each class and separates the classes using a linear decision boundary. On the other hand, the logistic regression approach estimates the class label probability using logistic functions and learns a set of weights for descriptor elements from the training samples. The SVM method involves constructing a set of separating hyperplanes in high-dimensional space to perform classification. To achieve multiclass classification, the one-versus-all scheme was used for logistic regression and LDA, whereas the one-versus-one scheme was used for SVM.

After training and testing samples are obtained, a corresponding 3D descriptor is calculated for each point cloud cluster as described in the previous section. For both training and testing, the descriptor values are scaled uniformly to a range of 0–1 and provided as input to each machine learning classifier. The output class labels and class probabilities are computed and stored for evaluation purposes. To handle cases in which the test object does not fit one of the trained categories, a rejection threshold is set using the computed class probability and ratio between the highest and second highest class probability. In this study, the training and testing processes were implemented using a Python toolkit known as *scikit-learn* (Pedregosa et al. 2011).

Results

The test data set is divided into three cases to evaluate the recognition performance in different scenarios. The first data set is composed of laser-scanned point clouds collected from a construction-equipment yard. The construction-equipment yard is chosen because it contains a large number of equipment test samples available to fully evaluate the object recognition performance. The second data set is obtained from laser scans of a construction site with ongoing construction work. This data set is aimed at demonstrating the relevance of the proposed method in a complex and challenging environment. The third data set is composed of point clouds of an excavator with different configurations of articulated parts to measure the robustness of the proposed method with respect to pose variations.

The first data set, the construction-equipment yard, consists of two sites, one that contains mostly large equipment such as excavators and dump trucks and another that contains mostly small equipment. Fig. 8 shows a composite image collected from photographs taken from each laser scan location. For each site, seven laser scans were taken and fused together into a single point cloud using registration techniques. The point cloud consists of approximately 1 million points after downsampling. The recognition results are shown in Fig. 9. The point cloud clusters associated with recognized objects are identified with a bounding box labeled with the predicted object class. Quantitative evaluation of the classification performance is discussed in the "Validation" section.

The second data set is taken from an active construction site with different scenes to target different construction equipment. One scene contains an excavator next to an excavated hole in the ground [Fig. 10(a)], whereas another scene contains a parked backhoe loader with other background objects present such as trucks, pipes, trees, and bushes [Fig. 10(b)]. Without proper segmentation methods, the trained classifier will be unable to recognize the targeted construction equipment because the foreground points are mixed with the background points. The first scene is especially challenging for the segmentation process because the ground surrounding the excavator is not flat and the bucket is placed on top of a pile of dirt. Figs. 11(a–d) show the detailed segmentation and clustering process for the first scene to obtain bounding boxes for candidate objects to be recognized. First, ground erosion is carried out to filter the background points surrounding the target objects. Figs. 11(a and b) show the resulting point cloud after one and two iterations of ground erosion, respectively. Next, the clustering process is used to group neighboring points into individual point cloud clusters and bounding boxes were drawn around each point cloud cluster [Fig. 11(c)]. Following that, a size filter is applied to the bounding boxes to eliminate spurious background objects such as trees, bushes, and pipes [Fig. 11(d)]. The point cloud resolution is reduced at the ground erosion stage [Figs. 11(a and b)] to speed up preprocessing but is increased after the clustering stage [Figs. 11(c and d)] to improve the accuracy of the object recognition stage. Finally, the proposed descriptor is applied to the remaining candidate objects and passed to a machine learning classifier to determine the object identity. Figs. 12(a and b) show the final classification results where the excavator and backhoe loader were correctly identified. In Figs. 12(a and b), darker points indicate background elements that have been filtered out after the segmentation stage, whereas lighter points indicate foreground elements that remain after the segmentation stage. A final bounding box and corresponding class label are only drawn around objects that were classified with high confidence. For example, in Fig. 12(b), a truck (upper-left corner) is also included as a candidate object after segmentation but it does not pass the final classification stage because the classifier determines that it is a poor fit for the targeted equipment classes of backhoe loader, bulldozer, dump truck, excavator, and front loader.

The third data set is derived from a miniature excavator model that can be manually controlled to change the angle of the cab, the boom, and the bucket. The joint angles are collectively adjusted to create four different poses (Fig. 13) and point cloud data were collected for each pose using a laser scanner. This data set is used to assess the capability of the proposed method to recognize articulated construction equipment that may have different poses in

Fig. 8. Equipment yard test site image (panorama) (image by Yihai Fang)

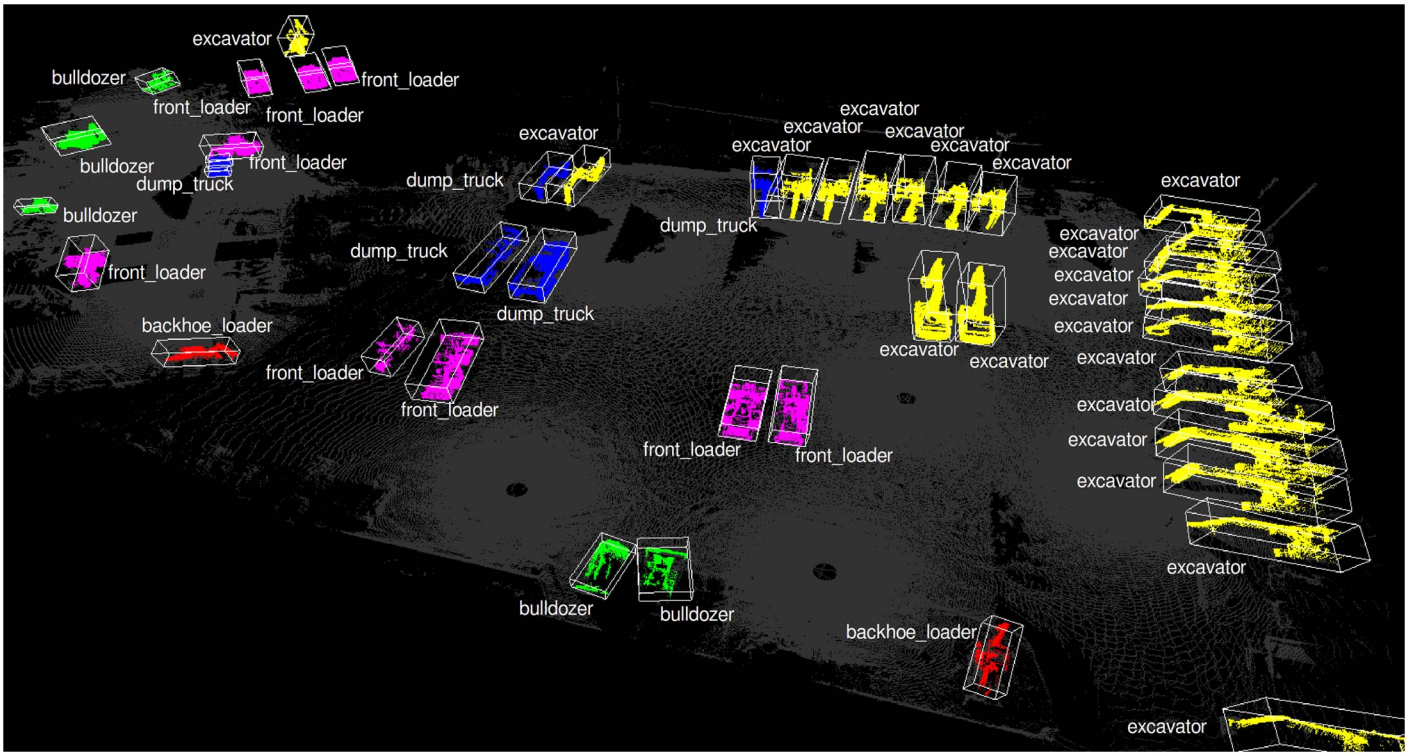

Fig. 9. Equipment yard test site recognition results

(a) (b)

Fig. 10. Point cloud from construction site test data: (a) scene with excavator; (b) scene with backhoe loader

actual working scenarios. An excavator was selected for this data set because it has the highest degrees of freedom in terms of joints that may be manipulated. Detailed evaluation of the classification performance in this data set is presented in the "Validation" section.

Validation

To evaluate the classification performance of the proposed PAD, the performance metrics of accuracy, precision, and recall were used. Accuracy is defined as the number of true positives and true negatives divided by the total overall samples. Precision is the number of true positives over the sum of true positives and false positives, while recall is the number of true positives over the sum of true positives and false negatives. Table 1 shows the accuracy, precision, and recall values for each construction-equipment category using PAD with the SVM classifier. The descriptor showed effective classification with accuracy values of more than 90% for all five equipment categories. The bulldozer and dump truck categories also demonstrated great recall rates where all the test samples in those categories were correctly identified. However, for the backhoe loader, bulldozer, and dump truck categories, some of the recognition rates may be skewed low because the number of test samples in those categories is much smaller than the total number of test samples.

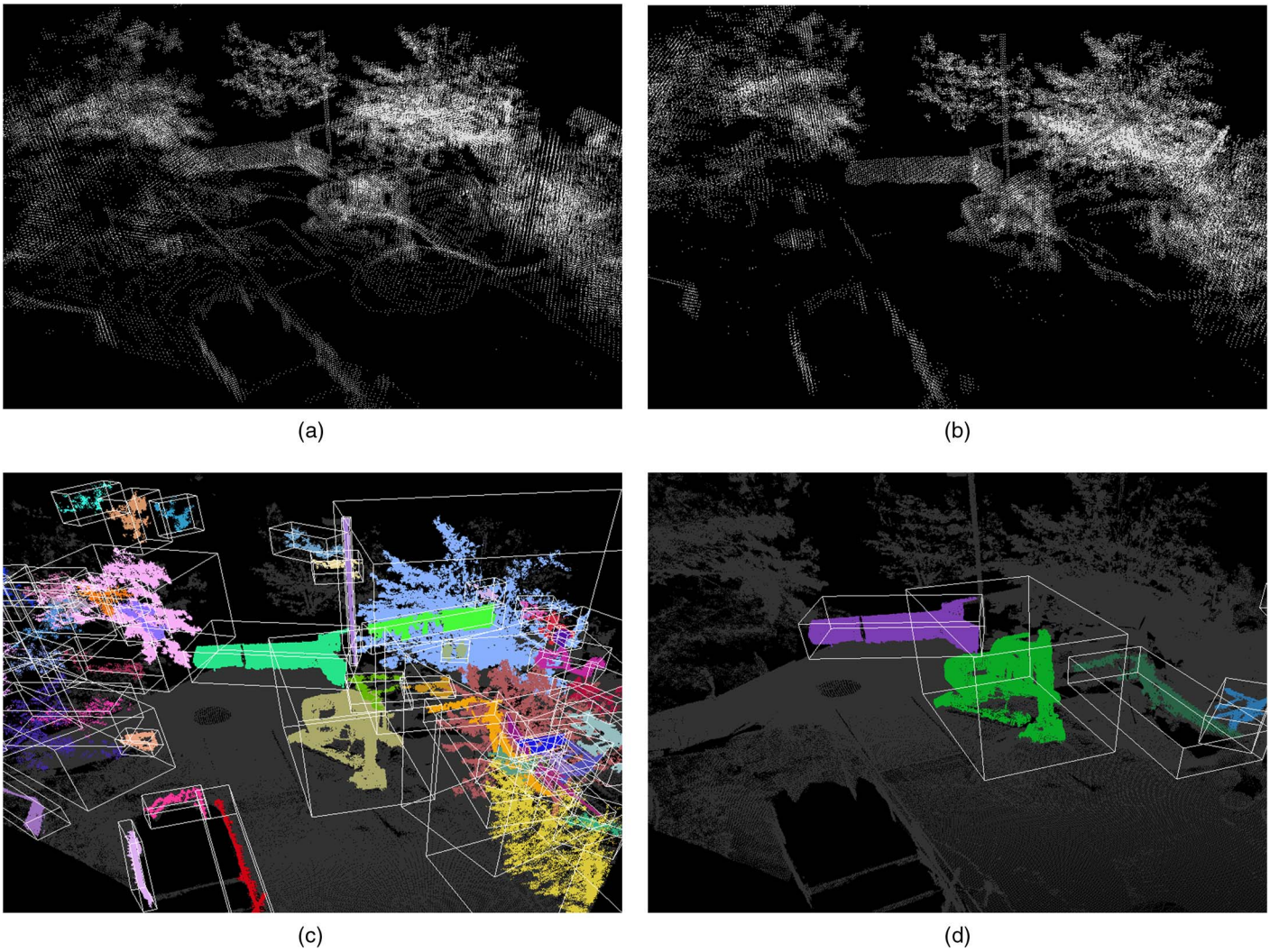

Fig. 11. Segmentation and clustering process for an excavator in a construction site point cloud: (a) point cloud after one iteration of ground erosion; (b) point cloud after two iterations of ground erosion; (c) clustering of objects and computation of bounding boxes; (d) refined object bounding boxes after size filtering

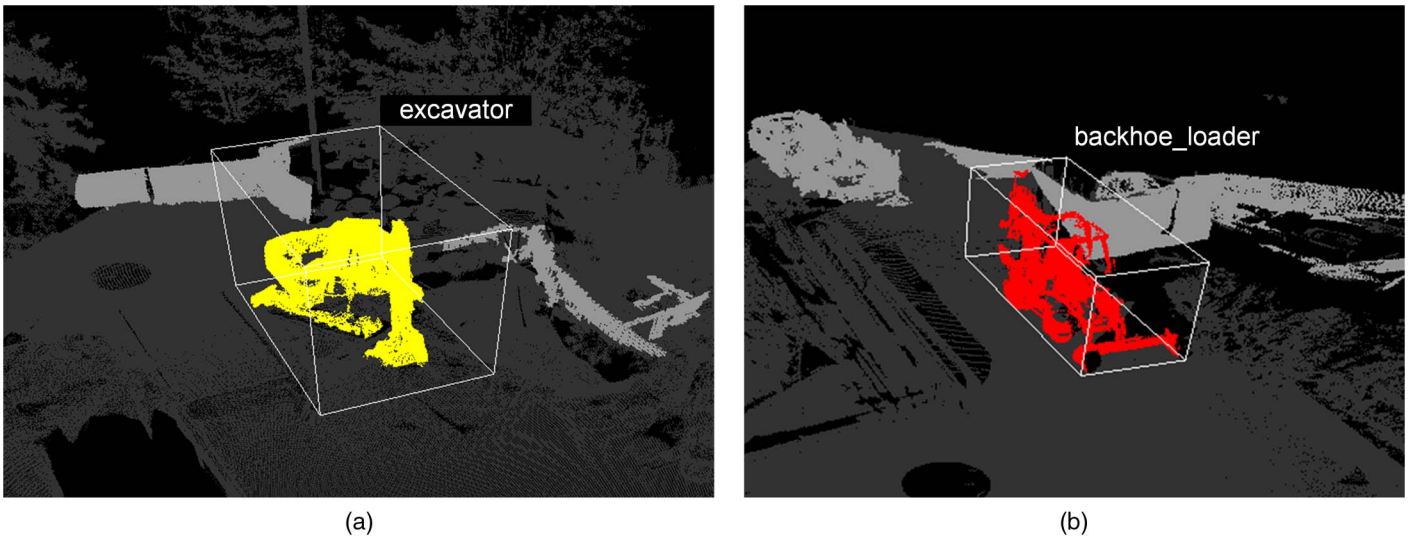

Fig. 12. Construction site test data recognition results: (a) scene with excavator; (b) scene with backhoe loader

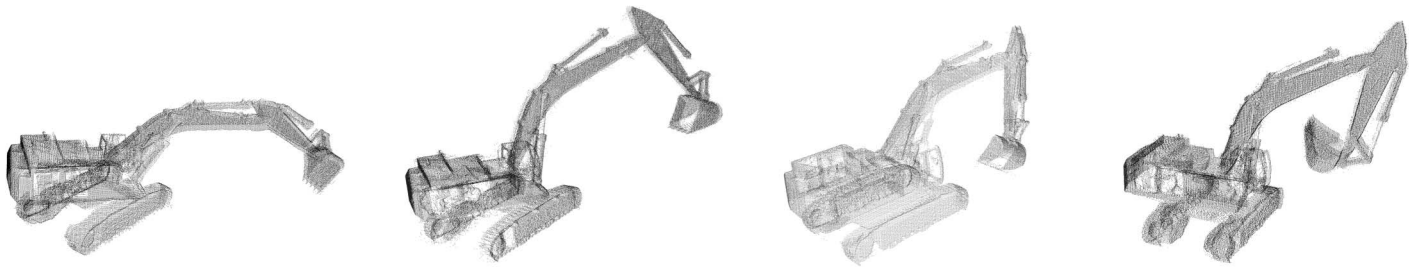

Fig. 13. Pose variation test data set

Table 1. Accuracy, Precision, and Recall Rates for Each Construction-Equipment Category

Category	Total samples	Correct samples	Accuracy (%)	Precision (%)	Recall (%)
Backhoe loader	4	2	95.45	100	50
Bulldozer	3	3	95.45	60	100
Dump truck	2	2	90.91	33.33	100
Excavator	23	21	93.18	95.45	91.3
Front loader	12	9	93.18	100	75

Table 2. Overall Recall Rate of Each Descriptor Based on Classification Algorithm

Classification algorithms	3D descriptors			
	ESF (%)	VFH (%)	GRSD (%)	PAD (%)
k-NN	64	66	14	75
SVM	77	55	11	84
LDA	34	18	20	73
Logistic regression	77	41	16	75
Average	63	45	15	76

Table 3. Recall Rate of Each Descriptor for Different Pose Configurations

Recall	ESF	VFH	GRSD	PAD
Correct samples/total samples	3/4	1/4	1/4	4/4
Recall rate (%)	75	25	25	100

Next, the overall recall rate of PAD was compared with that of other 3D descriptors such as ESF, VFH, and GRSD (Table 2). The overall recall rate was calculated for each descriptor using four different classification algorithms, which are k-NN, SVM, LDA, and logistic regression. PAD showed high recall rates for each classification algorithm used and also achieved the highest average recall rate over all classification methods.

The recall rate of PAD was also evaluated in situations where a construction equipment may be present in different poses. This is evaluated in the pose variation test data set with an excavator in different configurations of boom and bucket angles. Table 3 shows the number of poses where the sample is correctly identified as an excavator and the corresponding recall rate. PAD was able to achieve the highest recall rate compared with competing descriptors where all the test samples were correctly identified.

Finally, the performance metric of computation time is considered because it is relevant for real-time applications of object recognition. Table 4 shows the time taken in seconds for descriptor calculation in the classification training and testing phase for PAD compared with ESF, VFH, and GRSD. The number of feature dimensions is also shown because it highly factors into the computation time involved for each descriptor. PAD demonstrates significant savings in computation time due to the low number of feature dimensions.

Table 4. Computation Time of Each Descriptor

Descriptor	ESF	VFH	GRSD	PAD
Number of feature dimensions	640	308	21	10
Training time (s)	136	147	974	31
Testing time (s)	17	19	127	4

Discussion

As shown in the previous section, the proposed PAD demonstrated better performance compared to other descriptors in recognizing test samples from the five equipment categories considered (Table 2). It achieved the highest average recall rate of 75%, whereas the next best candidate is the ESF descriptor at 63%. It can be postulated that because PAD leverages known geometric structure in the domain of construction equipment, it is able to outperform more generic 3D descriptors such as ESF, VFH, and GRSD, which were originally studied in the context of recognizing small household objects.

It is also possible to analyze the classification performance with respect to the machine learning classifiers used such as k-NN, SVM, LDA, and logistic regression. Each classifier is characterized by its decision function to separate between classes and assumptions made about the input feature space. Thus, different classifiers would differ in terms of computation speed and classification performance on a given data set. For example, the k-NN classifier directly matches the query descriptor to the class with most similar descriptors without any assumptions about the input features. On the other hand, classification with LDA requires that the input features are normally distributed with identical covariances for each class, which might lead to worse classification performance, as shown in Table 2, if the assumptions do not hold. PAD is able to achieve good classification results regardless of the classification method, which indicates a strong distinction between features of different classes and makes them easily separable.

PAD is able to successfully handle object intraclass variation because, being a global 3D descriptor, it learns a global representation of the object features instead of any specific local features. In the training stage, heavy regularization in the form of artificial perturbations is employed to account for data imperfections and individual model differences as well as pose variations. The test results indicate that PAD demonstrates robustness with respect to

object intraclass variation. For example, within the excavator class of the equipment yard data set, the classifier is able to successfully recognize both wheel excavators and track excavators as well as excavators with different bucket types. In the pose variation data set, all object instances were correctly classified even with different poses.

PAD also demonstrates an advantage in computation speed when compared with competing 3D descriptors (Table 4). In this paper's implementation, the input point cloud is reduced to a simplified occupancy grid representation and a compact set of features is calculated on a global level. On the other hand, a descriptor such as GRSD requires the calculation of local RSD features for each point and so does not scale well with the size and complexity of the input point cloud.

Overall, the low dimensionality of PAD is emphasized because it represents a compact set of geometric features and significantly lowers the computation time. Low dimensionality can also be seen as an advantage for PAD in classification tasks with a small number of categories, whereas higher-dimensional descriptors might have a tendency to overfit the training data.

Limitations

The proposed PAD for construction-equipment classification is subject to a few limitations. First, the descriptor design is focused toward classification of the five selected equipment categories. The descriptor may not achieve the demonstrated classification performance if used as-is for construction equipments that are not within those categories. Thus in this study, the equipment categories selected are the ones most commonly used in a construction site scenario such as loader, excavator, and bulldozer. On a similar note, because PAD is characterized by low dimensionality, it may underfit the input data if the number of classification categories is too high. This problem can be addressed in future studies by extending the descriptor with more features and having a larger set of training data. Finally, this study faced the problem of a limited number of training and testing data. In the machine learning domain, thousands or even millions of training samples are commonly used to tune state-of-the-art classification frameworks. However, this is difficult to achieve in the experimental setup because a large amount of laser-scanned point clouds of construction equipment is challenging to acquire and process. In this study, the problem was handled by using data augmentation techniques such as training data synthesized from multiple views and artificially added noise to enable the classifier to generalize over a larger set of test cases.

Conclusion

The main contribution in this study is to develop a novel 3D descriptor named principal axes descriptor for the task of recognizing different types of construction equipment from a laser-scanned point cloud. PAD is applied in a recognition pipeline on segmented point cloud clusters of target construction equipment to identify discriminating geometric features in 3D. Test results indicate that PAD achieved faster and more accurate classification performance on construction-equipment data when compared against existing 3D descriptors such as ESF, VFH, and GRSD. The main benefits of PAD are a compact and efficient feature representation and improved recall rates from the acquired point cloud data. For future work, the authors envision expanding the training data set to take into consideration more categories (e.g., wheel types) of construction equipment as well as a larger variety of models within each category. The proposed method can also be extended to the more general case of identifying onsite assets in a construction project.

Acknowledgments

This material is based on work supported by the National Science Foundation (NSF) (Award # CMMI-1358176). Any opinions, findings, and conclusions or recommendations expressed on this material are those of the authors and do not necessarily reflect the views of the NSF. This research was also supported by the Chung-Ang University Research Grants in 2015.

References

3D Warehouse [Computer software]. Trimble Navigation, Sunnyvale, CA.

Ahmed, M., Haas, C. T., and Haas, R. (2012). "Toward low-cost 3D automatic pavement distress surveying: The close range photogrammetry approach." *Can. J. Civ. Eng.*, 38(12), 1301–1313.

Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression." *Am. Stat.*, 46(3), 175–185.

Bosché, F. (2010). "Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction." *Adv. Eng. Inf.*, 24(1), 107–118.

Bu, L., and Zhang, Z. (2008). "Application of point clouds from terrestrial 3D laser scanner for deformation measurements." *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII*, Copernicus GmbH, Göttingen, Germany, 545–548.

Chang, C., and Lin, C. (2011). "LIBSVM: A library for support vector machines." *ACM Trans. Intell. Syst. Technol.*, 2(3), 1–27.

Cho, Y. K., and Gai, M. (2014). "Projection-recognition-projection method for automatic object recognition and registration for dynamic heavy equipment operations." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000332, A4014002.

Cho, Y. K., Wang, C., Tang, P., and Haas, C. T. (2012). "Target-focused local workspace modeling for construction automation applications." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000166, 661–670.

Dai, F., and Lu, M. (2010). "Assessing the accuracy of applying photogrammetry to take geometric measurements on building products." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0000114, 242–250.

Dai, F., Rashidi, A., Brilakis, I., and Vela, P. (2012). "Comparison of image- and time-of-flight-based technologies for 3D reconstruction of infrastructure." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0000565, 69–79.

El-Omari, S., and Moselhi, O. (2008). "Integrating 3D laser scanning and photogrammetry for progress measurement of construction work." *Autom. Constr.*, 18(1), 1–9.

Fekete, S., Diederichs, M., and Lato, M. (2010). "Geotechnical and operational applications for 3-dimensional laser scanning in drill and blast tunnels." *Tunnelling Underground Space Technol.*, 25(5), 614–628.

Golparvar-Fard, M., Pena-Mora, F., and Savarese, S. (2009). "D4AR—A 4 dimensional augmented reality model for automation construction progress monitoring data collection, processing and communication." *J. Inf. Technol. Constr.*, 14(6), 129–153.

Gong, J., and Caldas, C. H. (2008). "Data processing for real-time construction site spatial modeling." *Autom. Constr.*, 17(5), 526–535.

Gong, J., Zhou, H., Gordon, C., and Jalayer, M. (2012). "Mobile terrestrial laser scanning for highway inventory data collection." *Computing in civil engineering*, ASCE, Reston, VA, 545–553.

Ham, Y., and Golparvar-Fard, M. (2013). "An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery." *Adv. Eng. Inf.*, 27(3), 395–409.

Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The elements of statistical learning*, Springer, New York.

Huber, D. (2014). "ARIA: The aerial robotic infrastructure analyst." ⟨http://spie.org/newsroom/technical-articles/5472-aria-the-aerial-robotic-infrastructure-analyst⟩ (Oct. 7, 2015).

Kim, C., Kim, C., and Son, H. (2013). "Automated construction progress measurement using a 4D building information model and 3D data." *Autom. Constr.*, 31, 75–82.

Lin, C., Yu, H., and Huang, F. (2011). "Dual coordinate descent methods for logistic regression and maximum entropy models." *Mach. Learn.*, 85(1–2), 41–75.

Marton, Z., Pangercic, D., Blodow, N., Kleinehellefort, J., and Beetz, M. (2010a). "General 3D modelling of novel objects from a single view." *Intelligent Robots and Systems (IROS)*, IEEE, Piscataway, NJ, 3700–3705.

Marton, Z., Pangercic, D., Rusu, R. B., Holzbach, A., and Beetz, M. (2010b). "Hierarchical object geometric categorization and appearance classification for mobile manipulation." *Proc., IEEE-RAS Int. Conf. on Humanoid Robots*, IEEE, Piscataway, NJ.

Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D. (2001). "Matching 3D models with shape distributions." *Proc., Int. Conf. on Shape Modeling and Applications, SMI 2001*, IEEE, Piscataway, NJ, 154–166.

Park, M.-W., Koch, C., and Brilakis, I. (2012). "Three-dimensional tracking of construction resources using an on-site camera system." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000168, 541–549.

Pedregosa, F., et al. (2011). "Scikit-learn: Machine learning in Python." *J. Mach. Learn. Res.*, 12, 2825–2830.

Pu, S., Rutzinger, M., Vosselman, G., and Oude Elberink, S. (2011). "Recognizing basic structures from mobile laser scanning data for road inventory studies." *ISPRS J. Photogramm. Remote Sens.*, 66(6), S28–S39.

Remondino, F., and Stoppa, D. (2013). *TOF range-imaging camerasfabio remondino*, Springer, Berlin.

Ruiz-Correa, S., Shapiro, L. G., Meila, M., Berson, G., Cunningham, M. L., and Sze, R. W. (2006). "Symbolic signatures for deformable shapes." *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(1), 75–90.

Rusu, R. B. (2009). "Semantic 3D object maps for everyday manipulation in human living environments." Ph.D. dissertation, Technischen Univ. Munchen, Munich, Germany.

Rusu, R. B., Bradski, G., Thibaux, R., and Hsu, J. (2010). "Fast 3D recognition and pose using the viewpoint feature histogram." *Intelligent Robots and Systems (IROS)*, IEEE, Piscataway, NJ, 2155–2162.

Rusu, R. B., and Cousins, S. (2011). "3D is here: Point cloud library (PCL)." *IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, Piscataway, NJ.

Tang, P., Akinci, B., and Huber, D. (2010a). "Semi-automated as-built modeling of light rail system guide beams." *Proc., Construction Research Congress 2010*, ASCE, Reston, VA.

Tang, P., Huber, D., Akinci, B., Lipman, R., and Lytle, A. (2010b). "Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques." *Autom. Constr.*, 19(7), 829–843.

Turkan, Y., Bosche, F., Haas, C. T., and Haas, R. (2012). "Automated progress tracking using 4D schedule and 3D sensing technologies." *Autom. Constr.*, 22, 414–421.

Wang, C., and Cho, Y. K. (2015). "Smart scanning and near real-time 3D surface modeling of dynamic construction equipment from a point cloud." *Autom. Constr.*, 49, 239–249.

Wang, C., Cho, Y. K., and Kim, C. (2015a). "Automatic BIM component extraction from point clouds of existing buildings for sustainability applications." *Autom. Constr.*, 56, 1–13.

Wang, J., Zhang, S., and Teizer, J. (2015b). "Geotechnical and safety protective equipment planning using range point cloud data and rule checking in building information modeling." *Autom. Constr.*, 49, 250–261.

Wohlkinger, W., and Vincze, M. (2011). "Ensemble of shape functions for 3D object classification." *2011 IEEE Int. Conf. on Robotics and Biomimetics*, IEEE, Piscataway, NJ, 2987–2992.

Yastikli, N. (2007). "Documentation of cultural heritage using digital photogrammetry and laser scanning." *J. Cult. Heritage*, 8(4), 423–427.