Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds

NICHOLAS ARNOLD

PLAMEN ANGELOV

TIM VINEY

PETER ATKINSON

*Author affiliations can be found in the back matter of this article

RESEARCH ARTICLE

]u[ubiquity press

ABSTRACT

This research addresses the problem of automatic extraction of memorial objects from cultural heritage sites represented as scenes of 3D point clouds. Point clouds provide a fine spatial resolution and accurate proxy of the real world. However, how to use them directly is not always obvious. This is especially true for applications where extensive training data or computational resources are not available. In this paper, we present a methodology for automatic segmentation and labelling of cultural heritage objects from 3D point cloud scenes. The proposed methodology is based on machine learning techniques and, in particular, makes use of the concept of transfer learning. Memorial objects are segmented from the scene based on their geometric shape characteristic through a conditional multi-scale partitioning scheme. Then, high-level latent feature descriptors are extracted by a convolutional neural network pre-trained on different 3D object models from a standard dataset (e.g., ModelNet). Based on these descriptors, a classification model (multilayer perceptron) is trained and applied to obtain semantic labels. Experiments demonstrated that the proposed methodology is effective for the extraction and labelling of grave marker objects from cultural heritage sites.

CORRESPONDING AUTHOR:

Nicholas Arnold

Lancaster University, GB

n.arnold@lancaster.ac.uk

KEYWORDS:

3D; Point Cloud; Transfer Learning; Cultural Heritage Management; Object Extraction

TO CITE THIS ARTICLE:

Arnold, N., Angelov, P., Viney, T and Atkinson, P. 2021. Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds. *Journal of Computer Applications in Archaeology*, 4(1), 79–93. DOI: <https://doi.org/10.5334/jcaa.66>

1 INTRODUCTION

Historic, cultural heritage and archaeological sites can be interpreted as hierarchical organisations of objects. The process of mapping and keeping an inventory of physical objects is fundamental to site conservation, management and analysis. Traditionally, objects are observed physically and recorded manually by an operator. Recent advances in remote sensing technologies, such as light detection and ranging (LiDAR) and digital photogrammetry, make it possible to instead create digital representations in the form of 3D point clouds. Indeed, 3D scanning technologies are becoming both more affordable and more versatile (Chase, Chase and Chase 2017; Favorskaya and Jain 2017; Royo and Ballesta-Garcia 2019). LiDAR based sensor hardware is appearing in both wearable systems and handheld devices. Additionally, photogrammetry software allows 2D images to be stitched together into a 3D point cloud scene; with this there is the potential to turn any camera into a proxy 3D sensor. These offer non-invasive, fine resolution alternatives to manual recording. As a result, 3D point cloud data are becoming a valuable resource for the fields of archaeological and cultural heritage. However, the question is then how to design an automated methodology for extracting, labelling and organising objects from these point clouds; especially one that is suitable for the real-world context.

Despite the adoption of digital technology, it remains a time-consuming task for an operator to find and label each object. Machine learning techniques that seek to automate object detection in point cloud data have been recently proposed; the most notable of which are built around convolutional neural networks (CNN) (Bello, Yu and Wang 2020). These supervised networks are composed of sequential layers wherein increasingly complex features are extracted. Specifically, the convolutional layers slide systematically a learnable convolution matrix, or kernel, across an input. This aggregates information from adjacent entities into features that are then passed to the next layer. Provided with a large training set of labelled data, CNNs are capable of generating discriminative high-level features (Bello, Yu and Wang 2020). The problem is that point clouds are an unusual data type. They are an unordered set of points in space and represent the external surface of the sampled object or scene. Each point is a vector denoting its x, y, z coordinates, and, depending on the specific acquisition technology, the points may contain additional observed information such as colour or intensity. Moreover, CNNs cannot easily take unstructured point clouds as input.

Many practical problems, and, in particular, cultural heritage and archaeological applications, often have limited access to labelled data, and in some cases the required data may be entirely non-existent, a necessary component for training CNNs. Training of CNNs is also computationally demanding. Moreover, the addition of

new validation data requires a complete retrain of the network. Therefore, it is not immediately clear how to take advantage of point clouds in real world applications such as these. To this end, we present a methodology suitable for the automatic extraction and identification of objects from cultural heritage sites. We highlight how point cloud data can be used directly to map and extract objects from archaeological and cultural heritage contexts without the need to first rasterise or convert to another representation (e.g., digital elevation, surface or terrain models). We apply and validate the proposed methodology for the task of locating, extracting and labelling grave markers from cultural heritage sites.

Grave marker detection is a relatively unexplored area. Notably, to the best of our knowledge, this is the first research of its kind on the extraction of grave marker objects directly from 3D point cloud data. Grave markers can be made from many different material components and take on a multitude of different sizes and shapes depending on their location, environment, age, condition and cultural background. Therefore, a highly generalisable methodology is necessary for their detection. While not related directly to the proposed methodology, LiDAR data have been used previously to aid in cemetery surveys (Weitman 2012). Additionally, point clouds have been used to represent memorial object models; for example Jakli et al. 2015 reconstructed sarcophagi from a point cloud data representation of a sunken Roman shipwreck. Zacharek et al. 2017 presented a low-cost approach to the collection of 3D grave marker models. To look below ground, Cannell et al. 2018 used ground penetrating radar and geochemical analysis to explore an unmarked graveyard at a medieval church site in Norway.

A recent focus in the literature has been on applying machine learning techniques to automate the detection of archaeological and cultural heritage objects. Initial methods were concerned mainly with 2D representations (Chase, Chase and Chase 2017). However, the growing presence of LiDAR as a digital surveying tool, as well as its integration into platforms such as geographic information system (GIS) software, has made point cloud analysis an important subject of research (Chase, Chase and Chase 2017). A natural step has been to ask how point cloud data can be paired with machine learning to benefit the archaeological and cultural heritage fields. Point cloud derived representations, such as surface and terrain models, have been used in conjunction with machine learning algorithms to automate the detection of barrows (Kramer 2015; Sevara et al. 2016) and Neolithic burial mounds (Guyot, Hubert-Moy and Lorho 2018), as well as in the detection of sub-surface archaeological structures (Fryskowska et al. 2017). While more traditional machine learning techniques are often employed, neural networks have been considered as well. Kazimi et al. 2018 demonstrated a CNN using LiDAR derived digital terrain maps to examine historic mining

regions in Germany. However, by not using point clouds directly, these approaches fail to take advantage of the innate dimensional information inherent to 3D.

A limitation of traditional supervised machine learning processes is that they are domain-specific; that is, they make predictions through learned properties determined by the data with which they are trained. In contrast, transfer learning allows a machine learning model trained in one domain to be reapplied to another domain of similar data. Transfer learning is one possible solution for areas where limited training data are available. Within the context of machine learning applied to cultural heritage and archaeology, the transfer learning concept has been applied to images from remote sensing surveys (Trier et al. 2016; Trier, Cowley and Waldeland 2019; Zingman et al. 2016). More recently, Verschoof-van der Vaart and Lambers 2019 explored transfer learning with region-based CNNs to detect barrows, Celtic fields and charcoal kilns from LiDAR-derived 2D images.

This paper presents a novel methodology to address the automatic extraction and identification of object instances within cultural heritage sites represented as 3D point clouds. The contributions of this methodology are: (i) how to operate directly on point clouds in a way suitable for real world application to cultural heritage and archaeology contexts, (ii) how to use the discriminative power of CNNs while mitigating their limitations and (iii) how to address the inherent challenges associated with point cloud data while benefiting from their 3D nature. Additionally, we propose a conditional multi-scale partitioning scheme within the methodology to ensure ground level objects are detected. In contrast to previous methodologies applied in a cultural heritage and archaeology context, the methodology presented in this paper involves methods applied directly to the 3D point cloud data rather than first transforming them to another structure.

The remainder of this paper is organised as follows. The methods and structure of the proposed methodology are detailed in Section 2; Section 2.1 details the segmentation process and extraction of geometric features and Section 2.2 describes the approach used for classification. Section 3 details the experimental results and Section 4 discusses the findings along with suggestions for future research. Finally, Section 5 concludes the paper.

2 METHODOLOGY

In the proposed methodology we consider the input point cloud P as a set of 3D points $\{p_i | i = 1, \dots, n\}$ such that $p_i \in P$ where n is the total number of points in P . The points P_i are a vector of coordinates (x, y, z) . The input P represents scenes of cultural heritage sites and is assumed to have been registered and pre-processed to remove outlier and duplicate points. The goal of the

methodology is to partition the scene into segments $S = \{S_1, \dots, S_n\}$ and provide a label L for each from a set of semantic classes C . To do so, it is comprised of two steps: segmentation and classification.

1. Segmentation is an unsupervised process that seeks to partition the point cloud scene into regions based on the continuity and homogeneity of the properties. We define the regions as local neighbourhoods and compute features that describe their similarities. We further, embed this information into an attributed graph structure and approximate the segments with smooth pre-defined shapes by a generalised minimal partition model, a type of loss function (Landrieu and Simonovsky 2017). This serves two purposes. First, the resulting segments effectively represent the objects contained within the scene, either in parts or as a whole. Second, by considering segments rather than individual points, the classification task is made easier as there is guaranteed to be fewer segments than points. See [Figure 1](#) for an illustration of the segmentation method.
2. For classification, we explore the idea of transfer learning and pre-train the ConvPoint network (Boulch 2019) on generic object models from the ModelNet40 (Wu et al. 2015) dataset. Transfer learning allows us to take advantage of the discriminative power of the CNN and couple it with the flexibility from more classic models for per-class training. We apply each partitioned segment to the pre-trained network to generate a set of high-level abstract features. These features represent a global descriptor for the segments and are input into a multilayer perception (MLP) network classifier to predict the class labels. See [Figure 4](#) for illustration of the classification method.

2.1 POINT CLOUD SEGMENTATION

In general, point cloud data are unstructured. That is, there is no defined neighbourhood to connect each point in space. This is in contrast to 2D images, where each pixel sits on a grid and has explicit neighbouring pixels. To extract meaningful features then, some form of structure must be imposed and designed specific to point cloud data. The common approach is to avoid processing the 3D data directly, instead rasterising the point cloud into multiple 2D representations (Su et al. 2015). An alternative is for the points to be placed within volumetric containers such as voxels (Qi et al. 2016b).

Methods such as Spin Images (Johnson and Hebert 1999), kernel signatures (Aubry, Schlickewei and Cremers 2011; Bronstein and Kokkinos 2010), and inner-distance descriptors (Ling and Jacobs 2007) use a local estimate of the underlying surface around the point. Recent kernel methods build on this.

Figure 1 Concept of the proposed segmentation methodology. Solid arrows represent the flow of processes; dashed arrows represent conditional processes. A complex 3D point cloud scene is taken as input and divided into multiple segments based on the set of features extracted in relation to the points' local neighbourhood. These segments are subsets of the original scene, themselves being point clouds, and are assumed to represent the objects contained in the scene.

Sparse kernels (Graham 2015, 2015), deformable kernels (Dai et al. 2017; Su et al. 2018), and continuous kernels (Boulch 2019) have been used to achieve leading accuracy scores for object recognition and semantic segmentation against benchmarks.

Recently, many researchers have begun to explore structure applied more directly to 3D point clouds, such as tree-based (Klokov and Lempitsky 2017; Riegler, Ulusoy and Geiger 2017), graph-based (Simonovsky and Komodakis 2017) and set-based approaches (Qi et al. 2016a, 2017). A more traditional, but equally direct solution, is to perform a point-wise search and define a structure based on the points' local neighbourhoods.

The term 'segmentation' in the context of point cloud data means the partitioning of spatial regions within the scene, based on some criteria. We can distinguish between two classes of point cloud segmentation problems commonly found in the literature. The first class of problems is to segment the scene based on some geometric similarity or characteristic and can be seen as the inference of object detection or localisation. The second class of problem is semantic segmentation, a fine-grained instance of classification. This segmentation performs point-wise classification, where individual points are provided with a label. Thereby, the scene is partitioned based on semantic similarity.

A simple form of the first class of segmentation problem is to partition the foreground and background of a scene (Dohan, Matejek and Funkhouser 2015; Golovinskiy, Kim and Funkhouser 2009). Because this type of segmentation represents regions of similarity, it is used regularly as precursor to object classification (Golovinskiy, Kim and Funkhouser 2009; Shapovalov, Velizhev and Barinova 2010). Spina et al. 2011 demonstrated this type of point cloud segmentation in a cultural heritage context. Similar to Hackel, Wegner

and Schindler 2016 and Guinard and Landrieu 2017 our method concerns the use of local point neighbourhoods by which we extract features to represent local regions of the scene. This method was chosen in contrast to semantic segmentation, which would require a network to be trained for specific terrain types as well as objects.

We consider a point-wise search to define local neighbourhoods. One such strategy is to search using a fixed-radius r , whereby a spherical (Lee and Schenk 2002) or cylindrical (Filin and Pfeifer 2005) representation is used to define the neighbourhood. Another is to consider the k -nearest neighbours around each point, based on some form of distance metric. This typically involves 2D (Niemeyer, Rottensteiner and Soergel 2014) or 3D (Jonathan et al. 2001) distances. As noted by Weinmann et al. 2015, for this solution to remain practical across varying scene types, search-based solutions require some form of optimization. This is either in terms of r or k , respectively. We define the points' local neighbourhoods through a k -nearest neighbour search in Euclidean space and optimise k based on eigenentropy, as advocated by Guinard and Landrieu 2017. This approach is suited to different point densities and gives more precise control over neighbourhood size (Weinmann et al. 2015).

2.1.1 Feature extraction

In this section, we present the features and algorithms used for the segmentation process.

The first stage of the proposed methodology is the segmentation process. Here, features that characterise the local dimensionality of the scene are extracted. For each point p_i , the k -nearest neighbouring points in the point cloud P are selected and the covariance matrix of their positions is calculated. From this we obtain the set of eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and corresponding eigenvectors u_1, u_2, u_3 . To determine the optimal size for k , a specific

energy function, the same as in (Weinmann et al. 2015), is used to minimise the eigenentropy E of the vector $(\lambda_1/\Lambda, \lambda_2/\Lambda, \lambda_3/\Lambda)$:

$$E = - \sum_{i=1}^3 \lambda_i / \Lambda \ln(\lambda_i / \Lambda),$$

with $\Lambda = \sum_{i=1}^3 \lambda_i$. This results in neighbourhoods which have maximum homogeneity or minimum disorder of points within the neighbourhood. The size of k is varied between $k_{min} = 10$ and $k_{max} = 100$ in increments of 1 (i.e., $\Delta k = 1$).

Using the eigenvalues, we construct a set of features $f_i \in \mathbb{R}^4$, which characterise the neighbourhood's local dimensionality and geometry. We use linearity, planarity, scattering (Demantké et al. 2012) and verticality (Guinard and Landrieu 2017):

$$\text{Linearity} = \lambda_1 - \lambda_2 / \lambda_1,$$

$$\text{Planarity} = \lambda_2 - \lambda_3 / \lambda_1,$$

$$\text{Sphericity} = \lambda_3 / \lambda_1,$$

$$\text{Verticality} = \sum_{j=1}^3 \lambda_j | \langle [0,0,1], u_j \rangle |.$$

The first three features are often referred to as dimensionality. Linearity describes how well the neighbourhood represents a 1-dimensional straight line, while planarity describes how well it fits to a 2-dimensional plane. Similarly, sphericity (also referred to as scattering in the literature) measures how well the neighbourhood resembles a

sphere. Verticality indicates the geometric orientation of the neighbourhood; for example, $\text{Verticality}_{min} = 0$ represents a horizontal orientation whereas $\text{Verticality}_{max} = 1$ represents a vertical orientation (Guinard and Landrieu 2017). Examples of these features can be seen in [Figure 2](#).

2.1.2 Adjacency graph structure

A graph structure can be used to capture how different entities are related to one another. The graph nodes (or vertices as they are sometimes called) represent a singular entity, while edges connecting nodes represent the relationship between entities. The edges may be either directed, such that they can be traversed only in a single direction, or undirected, such that they can be traversed in either direction. Graphs are commonly used in machine learning to represent probabilistic models. For example, Bayesian networks, Markov random fields (MRF), and conditional random fields (CRF), are all graphical models. Additionally, graphical models may also be used as the basis of a graph CNN, a generalisation of convolution operations to arbitrarily structured graphs (Landrieu and Simonovsky 2017).

Applied to point clouds, graphical models can be used as both a structure and for data analysis (Bronstein et al. 2017). Niemeyer et al. 2011 proposed graphical models to encode the spatial relationship between points into a graph structure called an adjacency graph. Furthermore, they showed how point cloud density and number of adjacent points affect this construction. Regarding this, they concluded that a larger neighbourhood has

Figure 2 Geometric features shown for an example image containing grave markers: (a) linearity, (b) planarity, (c) scattering and (d) verticality. Point cloud data provided by Atlantic Geomatics (UK) Limited.

the potential to better represent adjacency, albeit at a significant computational trade-off. Refining this conclusion, Guinard and Landrieu 2017 advocated a graph that represents the adjacency of the 10 nearest points. Note that the neighbourhood of points represented in the adjacency graph is different to the neighbourhood used for feature extraction.

To encode the spatial relationship between points, the point cloud is represented using an undirected adjacency graph $G_{nn} = (V, E_{nn})$. The set of nodes $V = \{V_1, \dots, V_n\}$ is constructed from each point in the point cloud, whereby each point p_i is represented by its associated features vector f_p and the edges E_{nn} encode the adjacency relationship of the 10 nearest neighbour points (Niemeyer et al. 2011). Segmentation is then a process in which the graph is split optimally into non-overlapping connected components. These splits are computed using the l_0 -cut pursuit algorithm (Landrieu and Obozinski 2017) and defined as the vector $g^* \in \mathbb{R}^{4 \times n}$ which minimises the following generalised minimum partition model:

$$g^* = \arg \min \sum_{i \in V} \|g_i - f_i\|^2 + \rho \sum_{(i,j) \in E_{nn}} w_{i,j} [g_i - g_j \neq 0],$$

with g as the variable value used to determine the optimal minimisation. The Iverson bracket $[\cdot]$ yields 0 if the internal expression is true, and 1 everywhere else. The edge weight $w \in \mathbb{R}_+^{[E]}$ is chosen to be linearly decreasing with respect to the edge length and factor ρ is the regularisation strength, which determines the coarseness of the resulting partition (Landrieu and Simonovsky 2017). This formulation ensures that the

resulting point cloud segments correspond to similar values of f without the need to define a maximum size for the segments. The point cloud segments are represented as the set $S = \{S_1, \dots, S_h\}$, where h is the number of segments returned by the cut adjacency graph. For clarity, the segments are the non-overlapping connected components. The segments are subsets of the original point cloud and the number of points vary per segment, see [Figure 3](#).

To increase the chances of finding smaller objects that may have been missed in the initial segmentation, a conditional multi-scale partitioning scheme is proposed. This secondary conditional partition considers only the largest 10% of planar segments. These are passed through the segmentation process again, with the neighbourhood for feature extraction adjusted to within a radius defined by the point density. If new components are found, then they are added to the set of segments. Otherwise, the segment is assumed to have continuous local shape and considered to be a single segment. This process ensures ground level objects are detected in large segments of ground level points.

2.2 CLASSIFICATION

Using the 3D point clouds directly (instead of converted 2D representations) in the classification method is essential to the real-world applicability, efficiency and generalisability of the overall methodology. Conversions to other representations would not only result in information loss but would require an in-between representation, such as mesh models (Su et al. 2015). This conversation alone remains a difficult task when applied to fine resolution real-world 3D point cloud

Figure 3 An example of a partitioned scene. The segments are assigned a colour randomly for demonstration purposes. Point cloud data provided by Atlantic Geomatics (UK) Limited.

data. Additionally, 2D multi-view methods are sensitive to viewpoint selection and occlusions within query instances. The real-world extracted objects are likely to contain noise from background objects (i.e., vegetation) and registration artefacts. While more recent multi-view 2D methods achieve leading accuracy scores on benchmarks, they rely on observational colour (*RGB*) information (Yu, Meng and Yuan 2018). Many fine-resolution point cloud datasets (especially those from LiDAR sensors) do not include this information as it requires specialised equipment and processing to collect and register the colour dimensions to the spatial points. Even if they are included, several factors present during the scanning process (e.g., glare, moisture, motion blur, camera focus, etc.) can contribute to inconsistent *RGB*

values. This is not to say that *RGB* or other multispectral data should not be used when available, but that to remain generally applicable, the classification method should ingest directly the point cloud data and not be dependent on any additional observed features.

For the classification sequence the transfer learning paradigm is followed. The ConvPoint network is pre-trained using generic object models from the ModelNet40 benchmark. An example of the adapted ConvPoint network is provided in Figure 5. The ConvPoint CNN was chosen because of its flexibility. It does not require a set input size and is robust to the permutation, scale and translation of the input 3D point cloud (Boulch 2019). As the name suggests, the Modelnet40 dataset is a collection of 3D models from across 40 different object

Figure 4 Illustration of the proposed classification methodology. The segments produced by the segmentation methodology are used as input; each is a 3D point cloud. Solid arrows represent the flow of processes. The dotted arrow indicates that the descriptors in the training set are created using the same pre-trained CNN model. The output to this process is a set of labels that relate the input point cloud segments to their predicted class.

Figure 5 The ConvPoint CNN adapted as classification feature extractor. The CNN is composed of five convolutional layers. Each consists of a convolution operation, one-dimensional batch normalisation and uses the rectified linear unit for the activation function. At each layer, the number of points per object is reduced, which inversely increases the number of descriptive features. The features from the last layer are used for classification, rather than continue to the fully connected output layer of the original CNN.

classes. It is important to note that the classes available in ModelNet40 do not include directly any of the target labels for classification (e.g., cultural heritage objects). ConvPoint directly ingests the spatial coordinate points through an adaptation of discrete kernel convolutions to be continuous. A simple MLP learns a dense geometrical weighting function that independently distributes the input points onto a kernel. At each layer, the convolution operation effectively mixes the estimation in the feature space and geometrical space. The derived kernel is then an explicit set of points associated with weights. Normalisation is added according to the input set size (Boulch 2019). The final fully connected layer of the pre-trained network is used to leverage the weighted layers as a fixed feature extractor.

Using the last convolutional layer, we compute a 1 by 512 feature vector x_h , where $\{x_h \mid 1, \dots, h\}$ is a global descriptor for each segment S_h . This vector is used to define an abstract feature space which is optimised for the separation of the training objects. Transfer learning, as a concept, assumes that this feature space can also be used to separate the new test objects.

A simple MLP was trained to learn the difference in the feature space, therefore, leverage the knowledge learned to classify new data and apply semantic labels L . The MLP itself is formulated as one hidden layer with 100 units, one output layer and uses the logistic sigmoid activation function,

$$f(x) = 1/(1 + \exp(-x)).$$

The features in x are assumed to be normally distributed and as such each is standardised by setting the features' mean at 0 and scaling to unit variance of 1; e.g., compute the standard score $z = (x - \text{mean}(x))/\text{std}(x)$ per feature. In doing so, we found this to increase classification accuracy results by at least 5%. See [Figure 5](#) for an illustration of the transfer learning procedure. We test and compare a variety of supervised classifiers for the classification task, which can be found in Section 3.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 DATASETS AND EVALUATION

3.1.1 ModelNet10

The Princeton ModelNet project provides a collection of 3D CAD object models split into two benchmarks: a 40-class subset and 10-class subset known as ModelNet40 and ModelNet10, respectively. The ModelNet10 data set was used to analyse the performance of the proposed transferred ConvPoint global descriptor in the classification process. The dataset was divided into training and validation sets. The CAD models were converted into 3D point clouds by randomly sampling points along the model surfaces. [Table 1](#) shows summary

CLASS NUMBER	CLASS NAME	TRAINING	TEST	TOTAL
1	bathtub	106	50	156
2	bed	515	100	615
3	chair	889	100	989
4	desk	200	86	286
5	dresser	200	86	286
6	monitor	465	100	565
7	nightstand	200	86	286
8	sofa	680	100	780
9	table	392	100	492
10	toilet	344	100	444

Table 1 Classification index for the ModelNet10 dataset; including number of training and test samples.

statistics for all 10 classes and their corresponding training and test samples.

3.1.2 Cultural Heritage Scenes

Two separate cultural heritage sites represented as 3D point cloud scenes were chosen for the evaluation of the proposed methodology applied to real-world data. The digitised cultural heritage sites were provided by the burial ground management system team at Atlantic Geomatics (UK) Limited. Scene 1 is a burial ground from Adlington civil parish in North West England. The scene is a subset of a much larger scene; the same large scene from which the classification training data were acquired. Scene 2 is a separate dataset: a burial ground located in Staines-upon-Thames in South East England. It is not taken from a larger scene. The scenes were collected by a terrestrial LiDAR sensor platform with a relative accuracy of 2 to 3 cm. In the analysis, four separate semantic classes of objects were targeted; memorial objects (grave markers such as headstones, stone crosses, sarcophagus, etc.), infrastructure (buildings, walls, gates, street poles, etc.), vegetation (tall grasses, shrubs, trees, canopy leaves, etc.) and ground (grass terrain, roads, paths, etc.).

3.1.3 Evaluation metrics

Following the standard convention from the field of machine learning, Precision, Recall and F_1 -score, along with their the macro- and weighted-average variations, were used as evaluation metrics. In a classification context, these metrics are ratios defined with respect to the number of true positives TP , false positives FP and false negatives FN returned per class. The recall, defined as $TP/(TP + FN)$, indicates the classifier's ability to find all positive samples. Likewise, precision is the fraction $TP/(TP + FP)$ that reflects the ability to return more relevant results than irrelevant ones. The F_1 is a measure of the classifier's accuracy. Formulated as $F_1 = 2(\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$, it considers both precision

and recall. All three metrics produce a score in the range [0, 1], reaching their worst value at 0 and best value at 1. The macro-average variation is then the mean of all scores divided by the number of classes. Similarly, the weighted-average is the score of each class weighted by the number of samples from that class.

3.1.4 Processing Platform

Experiments were run on a Unix machine with 2.7 GHz Intel Core i5, 16GB RAM and SSD. The combined process of segmentation and classification had an average run-time of 25 minutes for a point cloud of roughly 7 million points.

Several factors present during the scanning process can contribute to inconsistent observational colour (R , G , B) values. Furthermore, while some processes of point cloud acquisition, such as photogrammetry, inherently provide data as R , G , B , specialised equipment and processing is needed to register the colour dimensions to the point clouds generated by LiDAR sensors. As a result, many fine spatial resolution point cloud data sets do not include this information. In the proposed methodology we restrict the points to contain only the x , y , z coordinate information.

3.2 ANALYSIS OF TRANSFERRED DESCRIPTOR

3.2.1 Comparison of Classification Algorithms within the Proposed Methodology

Baselines are used generally to determine how well an algorithm performs. Thus, it can be a problem when a baseline for the specific domain does not exist. Applying the intuition behind transfer learning we conducted an initial experiment to gauge the effectiveness of our approach. We chose to assess a variety of supervised classification algorithms from the Scikit-learn Python package (Pedregosa et al. 2011) and test each against the ModelNet10 benchmark. A ‘best case’ for the baseline can be provided with the test data matching the data used to train the CNN feature extractor. To better reflect the objects recovered from the segmentation process, we chose to vary the number of points sampled, per model, to between 32 and 2048 points. The ModelNet10 point clouds were then given a global descriptor set using the adapted ConvPoint network. This allowed us to explore how

different classifiers interact with the data and determine the most appropriate approach for classification.

We interpret from the results in [Table 2](#) that the multi-layer perceptron (MLP) network implementations are the most promising among the tested methods, although the linear support vector machine (SVM) achieves similar scores, placing it behind the MLP by as little as 1% in the majority of metric categories. Within this experiment we investigated and compared the behaviours of different MLP activation functions. In particular, the weighted-average F_1 -score for the MLP with sigmoid activation performed particularly well, with an increase of at least 1–2% over the other MLP formulations. Consequently, this translates to a 7% and 5% F_1 -score increase over the next most accurate methods behind the SVM; random forest and k -nearest neighbours classifiers, respectively. The Gaussian Naive Bayes and decision tree classifiers performed the least accurate, where the MLP(sigmoid) had an increase of 21% and 19%, respectively.

3.2.2 Handcrafted global descriptors versus the transfer learning approach

The comparison of the proposed transfer learning global descriptor and commonly applied global features available in the open source Point Cloud Library (Rusu and Cousins 2011) are shown in [Table 3](#). These global

CLASSIFIER	PRECISION	RECALL	F1-SCORE
k-Nearest Neighbours	0.82/0.82	0.80/0.82	0.80/0.81
Gaussian Naive Bayes	0.71/0.74	0.68/0.70	0.68/0.70
Linear SVM	0.87/0.86	0.85/0.86	0.86/0.86
Random Forest	0.85/0.84	0.81/0.83	0.82/0.83
Decision Tree	0.67/0.68	0.66/0.68	0.66/0.67
MLP (sigmoid)	0.87/0.87	0.86/0.87	0.87/0.87
MLP (tanh)	0.87/0.87	0.86/0.87	0.86/0.87
MLP (relu)	0.86/0.86	0.85/0.86	0.85/0.86

Table 2 Results of the experiment to determine a baseline for classification. The values before the slash represent the macro-average score and after the slash the weighted-average score. The largest value for each metric is shown in bold.

DESCRIPTOR	PRECISION	RECALL	F1-SCORE
VFH(Rusu et al. 2010)	0.71/0.71	0.69/0.70	0.70/0.70
CVFH(Aldoma et al. 2011)	0.67/0.67	0.65/0.65	0.64/0.64
ESF(Wohlkinger and Vincze 2011)	0.05/0.05	0.17/0.19	0.07/0.08
GASD(Silva do Monte Lima and Teichrieb 2016)	0.81/0.80	0.79/0.80	0.79/0.80
Proposed transferred descriptor	0.87/0.87	0.86/0.87	0.87/0.87

Table 3 Classification results comparing global descriptors from the Point Cloud Library to the proposed transferred global descriptor on the ModelNet10 dataset. The values before the slash represent the macro-average score and after the slash the weighted-average score. The largest value for each metric is shown in bold.

descriptors are so called “handcrafted” and as such, have been designed specifically to encode certain aspects of a point cloud. This is in contrast to the transferred ConvPoint feature descriptor, which instead learned what features to encode from training data. The handcrafted descriptors were applied to classification of the ModelNet10 dataset. This experiment used the same methodology as described in Section 2.2, with their substitution as the global descriptor. The MLP(sigmoid) classifier was used as the classification model. It can be seen that the transferred descriptor from the ConvPoint network outperformed all handcrafted global descriptors across all three evaluation metrics, achieving a 7% increase in the weighted-average F1-score over the next best global descriptor implementation.

3.3 EVALUATION ON CULTURAL HERITAGE SITES

3.3.1 Comparison of Classifier Models with Real World Data

Real world data are often of varying quality and measured point clouds of real-world surfaces are no exception. They are commonly affected by variations in point density and objects of interest can occasionally become occluded during the scanning process. Therefore, we compare and investigate the behaviours of different classification models when applied to real world data of cultural heritage sites (Figure 6), although this should apply generally to any real-world dataset.

The classification results obtained for real world data from a pre-segmented cultural heritage scene are presented in Table 4. Analysis of these scores indicates that the MLP classifier outperformed the other tested methods; thus, the MLP demonstrated an ability to handle the real-world data. This is in support of the earlier assessments of the MLP classifiers. It is interesting to note that, given real world data, the random forest

CLASSIFIER	PRECISION	RECALL	F1-SCORE
k-Nearest Neighbours	0.89/0.89	0.76/0.89	0.79/0.88
Gaussian Naive Bayes	0.64/0.78	0.65/0.75	0.62/0.75
Linear SVM	0.86/0.90	0.82/0.90	0.84/0.89
Random Forest	0.91 /0.89	0.76/0.89	0.80/0.88
Decision Tree	0.72/0.82	0.72/0.81	0.71/0.81
MLP (sigmoid)	0.88/ 0.91	0.83/0.91	0.85/0.91
MLP (tanh)	0.87/ 0.91	0.83/0.91	0.84/ 0.91
MLP (relu)	0.87/ 0.91	0.83 /0.90	0.85 /0.90

Table 4 Comparison of different classification methods applied to memorial objects from a pre-segmented scene. Performance is evaluated using precision, recall and F_1 -score; the values before the slash represents the macro-average score and after the slash the weighted-average score. The largest score for each metric is shown in bold.

model performed closely with the linear SVM, and in fact achieved the largest macro-average precision score; this is in contrast to the earlier experiment. Based on these results, we concluded that an MLP with sigmoid activation function is the most suitable, of the tested classifiers, for use within the proposed methodology.

3.3.2 Evaluation of Methodology on Cultural Heritage Scenes

We applied the methodology to two separate cultural heritage scenes, the results of which are shown in Tables 5 and 6. The same training data were used in both scenes to train the MLP classifier. For both scenes, the proposed approach achieved a weighted average of at least 91% across all metrics. In general, classification of memorial objects was highly accurate, with an F_1 -score of 92% and 95% for scenes 1 and 2, respectively. The results from scene 2 illustrates how the proposed methodology generalises to a different dataset without retraining, even across two different and spatially distant regions. Classification accuracies for memorial, vegetation and infrastructure objects were similar to the scores in scene 1. However, there is a decrease when correctly determining segments that contain ground points. This is likely explained by the difference in terrains (e.g., slopes, flats, hills, etc.) between scenes; without the addition of these landscape characteristics the classification model can struggle to accommodate these changes.

	PRECISION	RECALL	F1-SCORE
Memorial	0.95	0.88	0.92
Infrastructure	0.56	0.83	0.67
Vegetation	0.94	0.94	0.94
Ground	0.85	0.94	0.89
macro avg.	0.83	0.90	0.90
weighted avg.	0.92	0.91	0.91

Table 5 Precision, recall and F1-score of MLP classification applied to Scene 1. Scores are an average result after 100 runs.

	PRECISION	RECALL	F1-SCORE
Memorial	0.94	0.95	0.95
Infrastructure	0.73	0.63	0.68
Vegetation	0.91	0.91	0.91
Ground	0.82	0.83	0.83
macro avg.	0.85	0.83	0.84
weighted avg.	0.91	0.91	0.91

Table 6 Precision, recall and F1-score of MLP classification applied to Scene 2. Scores are an average result after 100 runs.

Figure 6 Examples of a classified region from different classifier methods. Ground points are represented in yellow; vegetation is in blue; infrastructure is in orange and memorial objects are marked in green and red showing two sub-class identifications (headstone and cross). Point cloud data was provided by Atlantic Geomatics (UK) Limited.

4 DISCUSSION

Misclassification of the cultural heritage data lies in the infrastructure objects class. This is, in part, to be expected as memorial objects are often subjective. Cultural heritage sites normally contain various items of street furniture, and those in and of themselves might be a type of monument, e.g., a bench object may be classified into either the memorial or infrastructure class depending on semantics alone, with no visually distinct reasoning. The same can be true for trees and shrubs. The experiment classification index adheres strictly to memorial and non-memorial objects based on the manually labelled scene, which does not take this into account. This raises questions of how to impose semantic meaning to objects with little or no visually discerning attributes.

Pre-segmentation can also influence the classification results. It is possible for buildings and walls to be partitioned into smaller parts which share characteristics with headstone monuments or even vegetation, thus, potentially resulting in misclassification. In this sense, classification results are contingent on the quality of the segmentation process. An alternative to an object-wise segmentation would be to use a region-growing or point-wise algorithm. However, point density and noise have a direct impact on the time complexity of such methods. As a result, they have a limited ability to segment large-scale point clouds (Landrieu and Simonovsky 2017). Experiments showed that the proposed methodology

is capable of running on a personal computer. However, we note that RAM capacity was a limiting factor. Considerably large point clouds may need to be divided manually into smaller regions beforehand or else down sampled, provided that there is no great loss in visual representation. The objects in question should be easily identified by an operator.

In the future, we are interested in exploring how a more fine-grained classification could be achieved *within* the object classes. For example, many different burial ground monument types are found in a single cemetery. Additionally, grave markers from different geographical areas, different time periods and coming from different cultures, likely appear distinct and different from one another. With the general public becoming more interested in family ancestry and genealogy, there is a real need for this information to be available and to be provided at scale. Similarly, we are interested in ways to incorporate new object variations, unknown objects and additional classes to the methodology. The grounds, building infrastructure and serviceable equipment, etc., are all objects of importance for the maintenance of cultural heritage sites. It is, therefore, of value if the classification model does not have to be completely retrained each time a new variation is seen. Furthermore, based on the results of the transfer learning approach within this research, we are motivated to explore the use of various point cloud specific neural networks as feature extractors and evaluate their relative performances.

5 CONCLUSIONS

We presented a new methodology for the automatic identification and extraction of objects from 3D point cloud representations of cultural heritage sites. This methodology addressed how point cloud data can be used directly to map and extract objects from archaeological and cultural heritage contexts, without the need to rasterise or transform the data into another representation beforehand. Benchmarking exercises established that, compared to several classification methods, the proposed methodology achieves a statistically higher accuracy for both artificial and real-world datasets. We applied the methodology to the task of locating, extracting and labelling grave marker objects from two cultural heritage sites. The results demonstrated that the proposed approach can leverage transfer learning to separate objects from the scene and distinguish between multiple classes. We believe that this is the first time that such a methodology has been developed for the automatic and direct extraction and labelling of memorial objects from cultural heritage sites using 3D point cloud data.

ACKNOWLEDGEMENTS

The authors thank members at Atlantic Geomatics (UK) Limited for providing access to the raw LiDAR data and reference datasets.

FUNDING INFORMATION

This research was supported by Cumbria Innovations Platform (CUSP) at Lancaster University and the European Regional Development Fund (ERDF).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Nicholas Arnold orcid.org/0000-0003-3968-6233
Lancaster University, GB

Plamen Angelov orcid.org/0000-0002-5770-934X
Lancaster University, GB

Tim Viney
Atlantic Geomatics, GB

Peter Atkinson orcid.org/0000-0002-5489-6880
Lancaster University, GB

REFERENCES

- Aldoma, A, Vincze, M, Blodow, N, Gossow, D, Gedikli, S, Rusu, RB and Bradski, G.** 2011. CAD-model recognition and 6DOF pose estimation using 3D cues. In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. November 2011. pp. 585–592. DOI: <https://doi.org/10.1109/ICCVW.2011.6130296>
- Aubry, M, Schlickewei, U and Cremers, D.** 2011. *The wave kernel signature: A quantum mechanical approach to shape analysis*. 1 November 2011. pp. 1626–1633. DOI: <https://doi.org/10.1109/ICCVW.2011.6130444>
- Bello, SA, Yu, S and Wang, C.** 2020. Review: deep learning on 3D point clouds. *arXiv:2001.06280 [cs]*. DOI: <https://doi.org/10.3390/rs12111729>
- Boulch, A.** 2019. ConvPoint: continuous convolutions for cloud processing. *arXiv:1904.02375 [cs]*. DOI: <https://doi.org/10.1016/j.cag.2020.02.005>
- Bronstein, MM, Bruna, J, LeCun, Y, Szlam, A and Vandergheynst, P.** 2017. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42. DOI: <https://doi.org/10.1109/MSP.2017.2693418>
- Bronstein, MM and Kokkinos, I.** 2010. Scale-invariant heat kernel signatures for non-rigid shape recognition. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 2010. pp. 1704–1711. DOI: <https://doi.org/10.1109/CVPR.2010.5539838>
- Cannell, RJS, Gustavsen, L, Kristiansen, M and Nau, E.** 2018. Delineating an Unmarked Graveyard by High-Resolution GPR and pXRF Prospection: The Medieval Church Site of Furulund in Norway. *Journal of Computer Applications in Archaeology*, 1(1): 1–18. DOI: <https://doi.org/10.5334/jcaa.9>
- Chase, ASZ, Chase, DZ and Chase, AF.** 2017. LiDAR for Archaeological Research and the Study of Historical Landscapes. In: Masini, N and Soldovieri, F (eds.), *Sensing the Past. Geotechnologies and the Environment*. Cham: Springer International Publishing. pp. 89–100. DOI: https://doi.org/10.1007/978-3-319-50518-3_4
- Dai, J, Qi, H, Xiong, Y, Li, Y, Zhang, G, Hu, H and Wei, Y.** 2017. Deformable Convolutional Networks. *arXiv:1703.06211 [cs]*. DOI: <https://doi.org/10.1109/ICCV.2017.89>
- Demantké, J, Mallet, C, David, N and Vallet, B.** 2012. Dimensionality Based Scale Selection in 3d Lidar Point Clouds. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-5/W12: 97–102. DOI: <https://doi.org/10.5194/isprsarchives-XXXVIII-5-W12-97-2011>
- Dohan, D, Matejek, B and Funkhouser, T.** 2015. Learning Hierarchical Semantic Segmentations of LIDAR Data. In: *2015 International Conference on 3D Vision*. October 2015 pp. 273–281. DOI: <https://doi.org/10.1109/3DV.2015.38>
- Favorskaya, MN and Jain, LC.** 2017. Overview of LiDAR Technologies and Equipment for Land Cover Scanning. In:

- Favorskaya, MN and Jain, LC (eds.), *Handbook on Advances in Remote Sensing and Geographic Information Systems: Paradigms and Applications in Forest Landscape Modeling*. Intelligent Systems Reference Library. Cham: Springer International Publishing. pp. 19–68. DOI: https://doi.org/10.1007/978-3-319-52308-8_2
- Filin, S** and **Pfeifer, N.** 2005. Neighborhood Systems for Airborne Laser Data. *Photogrammetric Engineering & Remote Sensing*, 71: 743–755. DOI: <https://doi.org/10.14358/PERS.71.6.743>
- Fryskowska, A, Kedzierski, M, Walczykowski, P, Wierzbicki, D, Deliś, P** and **Lada, A.** 2017. Effective Detection of Sub-Surface Archeological Features from Laser Scanning Point Clouds and Imagery Data. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5: 245–251. DOI: <https://doi.org/10.5194/isprs-archives-XLII-2-W5-245-2017>
- Golovinskiy, A, Kim, VG** and **Funkhouser, T.** 2009. Shape-based recognition of 3D point clouds in urban environments. In: September 2009 IEEE. pp. 2154–2161. DOI: <https://doi.org/10.1109/ICCV.2009.5459471>
- Graham, B.** 2015. Sparse 3D convolutional neural networks. *arXiv:1505.02890 [cs]*. DOI: <https://doi.org/10.5244/C.29.150>
- Guinard, S** and **Landrieu, L.** 2017. Weakly Supervised Segmentation-Aided Classification of Urban Scenes from 3d Lidar Point Clouds. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1/W1: 151–157. DOI: <https://doi.org/10.5194/isprs-archives-XLII-1-W1-151-2017>
- Guyot, A, Hubert-Moy, L** and **Lorho, T.** 2018. Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques. *Remote Sensing*, 10(2): 225. DOI: <https://doi.org/10.3390/rs10020225>
- Hackel, T, Wegner, JD** and **Schindler, K.** 2016. *Contour Detection in Unstructured 3D Point Clouds*. June 2016 IEEE. pp. 1610–1618. DOI: <https://doi.org/10.1109/CVPR.2016.178>
- Jaklič, A, Erič, M, Mihajlović, I, Stopinšek, Ž** and **Solina, F.** 2015. Volumetric Models from 3d Point Clouds: The Case Study of Sarcophagi Cargo from a 2nd/3rd Century Ad Roman Shipwreck Near Sutivan on Island Bra, Croatia. *Journal of Archaeological Science*, 62: 143–152. DOI: <https://doi.org/10.1016/j.jas.2015.08.007>
- Johnson, AE** and **Hebert, M.** 1999. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5): 433–449. DOI: <https://doi.org/10.1109/34.765655>
- Jonathan, E, Roberts, C, Presentations, S, Linsen, L** and **Prautzsch, H.** 2001. *Local Versus Global Triangulations*.
- Kazimi, B, Thiemann, F, Malek, K, Sester, M** and **Khoshelham, K.** 2018. *Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data*. 1 September 2018 p.
- Klokov, R** and **Lempitsky, V.** 2017. Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models. In: 2017 IEEE International Conference on Computer Vision (ICCV). October 2017 Venice: IEEE. pp. 863–872. DOI: <https://doi.org/10.1109/ICCV.2017.99>
- Kramer, I.** 2015. *An archaeological reaction to the remote sensing data explosion. Reviewing the research on semi-automated pattern recognition and assessing the potential to integrate artificial intelligence*. PhD Thesis.
- Landrieu, L** and **Obozinski, G.** 2017. Cut Pursuit: fast algorithms to learn piecewise constant functions on general weighted graphs. *SIAM Journal on Imaging Sciences*, 10(4): 1724–1766. DOI: <https://doi.org/10.1137/17M1113436>
- Landrieu, L** and **Simonovsky, M.** 2017. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. *CoRR*, abs/1711.09869. DOI: <https://doi.org/10.1109/CVPR.2018.00479>
- Lee, I** and **Schenk, A.** 2002. Perceptual organization of 3D surface points. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34.
- Ling, H** and **Jacobs, DW.** 2007. Shape Classification Using the Inner-Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2): 286–299. DOI: <https://doi.org/10.1109/TPAMI.2007.41>
- Niemeyer, J, Rottensteiner, F** and **Soergel, U.** 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 152–165. DOI: <https://doi.org/10.1016/j.isprsjprs.2013.11.001>
- Niemeyer, J, Wegner, JD, Mallet, C, Rottensteiner, F** and **Soergel, U.** 2011. Conditional Random Fields for Urban Scene Classification with Full Waveform LiDAR Data. In: Stilla, U, Rottensteiner, F, Mayer, H, Jutzi, B and Butenuth, M (eds.), *Photogrammetric Image Analysis. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer. pp. 233–244. DOI: https://doi.org/10.1007/978-3-642-24393-6_20
- Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R, Dubourg, V, Vanderplas, J, Passos, A, Cournapeau, D, Brucher, M, Perrot, M** and **Duchesnay, E.** 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830.
- Qi, CR, Su, H, Mo, K** and **Guibas, LJ.** 2016a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv:1612.00593 [cs]*.
- Qi, CR, Su, H, Niessner, M, Dai, A, Yan, M** and **Guibas, LJ.** 2016b. Volumetric and Multi-View CNNs for Object Classification on 3D Data. *arXiv:1604.03265 [cs]*. DOI: <https://doi.org/10.1109/CVPR.2016.609>
- Qi, CR, Yi, L, Su, H** and **Guibas, LJ.** 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *arXiv:1706.02413 [cs]*. DOI: <https://doi.org/10.1109/CVPR.2017.701>
- Riegler, G, Ulusoy, AO** and **Geiger, A.** 2017. OctNet: Learning Deep 3D Representations at High Resolutions. *arXiv:1611.05009 [cs]*. DOI: <https://doi.org/10.1109/CVPR.2017.701>
- Royo, S** and **Ballesta-Garcia, M.** 2019. An Overview of Lidar Imaging Systems for Autonomous Vehicles. *Applied Sciences*, 9(19): 4093. DOI: <https://doi.org/10.3390/app9194093>

- Rusu, RB, Bradski, G, Thibaux, R and Hsu, J.** 2010. Fast 3D recognition and pose using the Viewpoint Feature Histogram. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October 2010. Taipei: IEEE. pp. 2155–2162. DOI: <https://doi.org/10.1109/IROS.2010.5651280>
- Rusu, RB and Cousins, S.** 2011. 3D is here: Point Cloud Library (PCL). In: *2011 IEEE International Conference on Robotics and Automation*. May 2011. pp. 1–4. DOI: <https://doi.org/10.1109/ICRA.2011.5980567>
- Sevara, C, Pregesbauer, M, Doneus, M, Verhoeven, G and Trinks, I.** 2016. Pixel versus object—A comparison of strategies for the semi-automated mapping of archaeological features using airborne laser scanning data. *Journal of Archaeological Science: Reports*, 5: 485–498. DOI: <https://doi.org/10.1016/j.jasrep.2015.12.023>
- Shapovalov, R, Velizhev, E and Barinova, O.** 2010. Nonassociative markov networks for 3d point cloud classification. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII, Part 3A*. 2010. pp. 103–108.
- Silva do Monte Lima, JP and Teichrieb, V.** 2016. An Efficient Global Point Cloud Descriptor for Object Recognition and Pose Estimation. In: *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. October 2016. pp. 56–63. DOI: <https://doi.org/10.1109/SIBGRAPI.2016.017>
- Simonovsky, M and Komodakis, N.** 2017. Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs. *arXiv:1704.02901 [cs]*. DOI: <https://doi.org/10.1109/CVPR.2017.11>
- Spina, S, Debattista, K, Bugeja, K and Chalmers, A.** 2011. *Point Cloud Segmentation for Cultural Heritage Sites*. 1 January 2011. pp. 41–48. DOI: <https://doi.org/10.2312/VAST/VAST11/041-048>.
- Su, H, Jampani, V, Sun, D, Maji, S, Kalogerakis, E, Yang, M-H and Kautz, J.** 2018. SPLATNet: Sparse Lattice Networks for Point Cloud Processing. *arXiv:1802.08275 [cs]*. DOI: <https://doi.org/10.1109/CVPR.2018.00268>
- Su, H, Maji, S, Kalogerakis, E and Learned-Miller, E.** 2015. Multi-view Convolutional Neural Networks for 3D Shape Recognition. *arXiv:1505.00880 [cs]*. DOI: <https://doi.org/10.1109/ICCV.2015.114>
- Trier, ØD, Cowley, DC and Waldeland, AU.** 2019. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automated mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, 26(2): 165–175. DOI: <https://doi.org/10.1002/arp.1731>
- Trier, Ø, Salberg, A-B, Pilø, L, Tønning, C, Johansen, H and Aarsten, D.** 2016. *Semi-automatic mapping of cultural heritage from airborne laser scanning using deep learning*. 17 April 2016.
- Verschoof-van der Vaart, WB and Lambers, K.** 2019. Learning to Look at Lidar: The Use of R-CNN in the Automated Detection of Archaeological Objects in Lidar Data from the Netherlands. *Journal of Computer Applications in Archaeology*, 2(1): 31–40. DOI: <https://doi.org/10.5334/jcaa.32>
- Weinmann, M, Schmidt, A, Mallet, C, Hinz, S, Rottensteiner, F and Jutzi, B.** 2015. Contextual Classification of Point Cloud Data by Exploiting Individual 3d Neighbourhoods. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. March 2015. Copernicus GmbH. pp. 271–278. DOI: <https://doi.org/10.5194/isprsannals-II-3-W4-271-2015>
- Weitman, S.** 2012. *Using Archaeological Methods in Cemetery Surveys with Emphasis on the Application of Lidar*. Georgia Southern University.
- Wohlkinger, W and Vincze, M.** 2011. Ensemble of shape functions for 3D object classification. In: *2011 IEEE International Conference on Robotics and Biomimetics*. December 2011. pp. 2987–2992. DOI: <https://doi.org/10.1109/ROBIO.2011.6181760>
- Wu, Z, Song, S, Khosla, A, Yu, F, Zhang, L, Tang, X and Xiao, J.** 2015. 3D ShapeNets: A deep representation for volumetric shapes. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015. pp. 1912–1920. DOI: <https://doi.org/10.1109/CVPR.2015.7298801>
- Yu, T, Meng, J and Yuan, J.** 2018. Multi-view Harmonized Bilinear Network for 3D Object Recognition. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 2018. Salt Lake City, UT: IEEE. pp. 186–194. DOI: <https://doi.org/10.1109/CVPR.2018.00027>
- Zacharek, M, Delis, P, Kedzierski, M and Fryskowska, A.** 2017. Generating Accurate 3d Models of Architectural Heritage Structures Using Low-Cost Camera and Open Source Algorithms. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W1: 99–104. DOI: <https://doi.org/10.5194/isprs-archives-XLII-5-W1-99-2017>
- Zingman, I, Saupe, D, Penatti, OAB and Lambers, K.** 2016. Detection of Fragmented Rectangular Enclosures in Very High Resolution Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8): 4580–4593. DOI: <https://doi.org/10.1109/TGRS.2016.2545919>

TO CITE THIS ARTICLE:

Arnold, N, Angelov, P, Viney, T and Atkinson, P. 2021. Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds. *Journal of Computer Applications in Archaeology*, 4(1), 79–93. DOI: <https://doi.org/10.5334/jcaa.66>

Submitted: 30 September 2020   Accepted: 23 March 2021   Published: 23 April 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Computer Applications in Archaeology is a peer-reviewed open access journal published by Ubiquity Press.