MACHINE LEARNING METHODS FOR UNESCO CHINESE HERITAGE: COMPLEXITY AND COMPARISONS

K. Zhang[1], S. Teruggi[1], F. Fassi[1]

[1]3D Survey Group, ABC Dep, Politecnico di Milano, Via Ponzio 31, 20133 Milano, Italy.
kai2.zhang@mail.polimi.it, (simone.teruggi, francesco.fassi)@polimi.it

Commission II

KEY WORDS: Cultural Heritage, Point cloud, classification, machine learning, Chinese Architecture

ABSTRACT:

Recent years have seen the investigation and 3D documentation of architectural heritage becoming more accessible. The digitalization process could be more efficient when artificial intelligence is used in processing point cloud models. This article investigates the use of machine learning classification algorithms and a Multi-Level Multi-Resolution (MLMR) methodology to classify two point cloud projects in China, Nanchan Ssu, and Fokuang Ssu. Performances of multiple algorithms and solutions are compared, proving the applicability of MLMR on the point clouds. The practices pointed out the significance of corresponding features to classification rules and a sound logic in designing a systematic label tree with hierarchical semantic meanings.

1. INTRODUCTION

1.1 Point cloud and classification in cultural heritage

Nowadays, laser scanner and photogrammetric techniques are more than consolidated and are the two most used survey methodologies in the architectural and environmental worlds. The products of such survey activities are primarily 3D point clouds containing all the geometric information of the surveyed objects.

A Point cloud is a collection of elementary geometrical primitives representing the shape, size, position and orientation of objects in space. It is a set of points where each is represented by three coordinates in a cartesian coordinate system (x, y, z); it may combine colours, reflectance, intensity, normals and other information depending on the instrument used. Simple and powerful as they are, point clouds suit best the documentation works in the architectural field, especially when it comes to Cultural Heritage (CH). They are capable of collecting and representing large and complicate objects with high geometric accuracy.

To use the geometric information, it is usually necessary to perform a 3D modelling post-processing. This mostly consists in manual and time-consuming operations. Furthermore, these operations reduce the metric quality of the data as the simplification is not always dictated by the purpose as by the limitation of the software and it is subjective depending on the technical skills and from the personal interpretation of the operators. A solution to minimize time and costs of these modelling processes, is that of using the point clouds directly inside the production process. Recent research activities have been focusing on developing strategies that allow using these point models as real triangular meshes. The first step would be that of assigning semantic meaning to each point through their classification into predetermined classes.

Recent advancements in Artificial Intelligence (AI) (Machine Learning (ML) and Deep Learning (DL) methods) provide a solution to manage point cloud dataset in a more rational and semi-automatic way. Algorithms were developed to classify the datasets, distributing semantic meaning to each segment. Indeed, the labour on defining each item one by one can be largely saved. Actual applications of different techniques have varying performances that depends on the scenes analysed. The whole process requires improvements and greater generalization. Their potential has already been proofed in many other fields such as in natural language processing and image classification (Computer Vision), like AlexNet (Krizhevsky et al., 2012) and ResNet (He et al., 2016).

While the production of 3D point cloud has become increasingly convenient, when results would become precise and reliable, AI is expected to play an important role dealing with trillions of points (samples) collected from a complicated real scene.

1.2 Aim and content

The work performs and evaluates the automatic classification method based on the Multi-Level Multi-Resolution (MLMR) approach combined with ML algorithms.

The choice is to focalize the attention on Chinese Cultural Heritage, especially on wooden structures. Such structures are exposed with a high level of complexity, credited with value in different aesthetic pursuits. The geometrical attributes (diameter, surface roughness, etc) of most architectural elements vary from geographical surroundings and dynasties and are dissimilar from the counterpart of European ones. Two ancient wooden structures have been selected to test the validity of the approach: the great hall of Nanchan Ssu (China), the East Great Hall of Fokuang Ssu (China). They have different scale but close location and similar architectural styles. The comparison can be helpful on discussing relevant topics of classification.

2. STATE OF THE ART

Segmentation refers to the action of grouping points in subsets (commonly called segments) characterized by sharing one or more characteristics (geometric, radiometric, …). As summarized by Grilli et al. (2017), segmentation methods could be grouped into: edge-based, region growing, model fitting, hybrid method and AI (ML and DL) applications.

AI methods either rely on a set of provided training examples with manually annotated labels to learn to perform the classification tasks (supervised learning), or, seeks to build models that automatically understand how the data are organized (unsupervised learning). These methods are generally robust

against noise and occlusions, but require a large amount of training data and high computing power to run the algorithm.

Classification means to assign points to specific classes. These classes as prediction results are not linearly continuous and are indicated with semantic labels previously defined. The methods, using specific algorithms, can be sorted into supervised and unsupervised approach.

In a supervised ML approach, including support vector machines (SVM) (Oses et al., 2014), naive bayes (Mathias et al., 2011) and random forest (RF) (Ho, 1995, Breiman et al., 2001), semantic categories are learned from a dataset of manually annotated data. The trained model is then used to spread the semantic classification to the entire dataset. Normally, it is not mandatory to provide a large amount of annotated data to train the model.

More traditional methods, instead, typically relies on a range of hand-crafted shape descriptors, as feature vectors, from which to learn the classification pattern. As summarized by Griffiths and Boehm (2019) these descriptors include: local surface patches, spin images, intrinsic shape signatures, heat kernel signatures. Wang et al. (2015) discussed multi-scale and hierarchical feature extraction methods. While, Grilli et al. (2019) discussed which are the most relevant geometric covariance features to be used during the application of the algorithm.

Unsupervised approaches differ from supervised learning approaches, since features themselves are learned as part of the training process. Big data in recent years has made these methodologies, especially DL, accessible and popular. Among many others, convolutional neural networks (CNN), constitute the most representative approach in DL. Developed for 2D image analysis, it has proved really effective in different fields, e.g., object detection and model matching in street view scenario (Coenen et al., 2019). The progress made with 2D images acts as a foundation upon which to develop many 3D learning algorithms. Liu et al. (2019), in their survey on DL applications, summarized the main models for feature learning with raw point cloud as input. They were generally divided into point-based and tree-based approaches. The first directly takes the raw point cloud as the input for training the DL net. The second, firstly employs a k-dimensional tree (Kd-tree) structure to transform the point cloud in a regular representation and then feeds them into DL models.

2.1 Classification in CH

When confronting with a specific point cloud model in CH it is mandatory to meet the specific needs that particular heritage poses. Different classification techniques perform very differently towards singular unique objects with varying mass, material, coloristic appearance, surface variation etc.

A 2.5D approach utilizes features and labels from 2D images, projecting them onto 3D models to perform the classification. Grilli et al. (2018) in their work presented a classification approach that works with 2D data as input ("texture-based" approach). Optimized models, orthoimages ad UV maps were created for the test cases under study. Firstly, they performed the classification on orthoimages or the UV maps, and then they projected the 2D classification results onto the 3D objects.

Murtiyoso et al. (2020) proposed an algorithmic approach packaged into a toolbox created inside the MATLAB environment hosting all the codes and functions under one project: HERitAge by point CLoud procESsing for Matlab (M_HERACLES). The general workflow performs segmentation from the scale of a historical neighbourhood up to that of the architectural element. Firstly, the toolbox performs the segmentation from the scale level of a neighbourhood to that of individual buildings using GIS shapefile data to assist the process. Afterwards, it performs the segmentation from a building's scale

level to that of architectural elements like pillars and beams, utilizing several Euclidean geometry-based rules and a slicing approach to identify clusters.

AI approaches that directly work on the 3D CH point model started to appear only in recent years. Grilli et al. (2019) presented a ML approach that works directly on point clouds. Based on the use of the RF algorithm the classification process consists in providing a set of manually labelled samples together with computed geometric covariance features to train the model. Afterwards, the classification is spread to the whole dataset.

Starting from the approach presented by Grilli et al., Teruggi et al. (2020) proposed a MLMR approach tested on two large scale datasets: the Pomposa Abbey and Milan Cathedral, both in Italy. The full resolution dataset is subsampled and big macro-elements are classified training a specific RF classifier on a low resolution version of the point model. The result is then back interpolated on a higher resolution point cloud so that elements requiring high geometric accuracy can be subdivided. The process iterates up to the classification of single high detail architectural elements using the full resolution dataset (initial resolution). In this approach, each steps require to train a specific RF model, but, small amount of labelled data is required and the speed of training and classification proved effective. It hierarchically splits the data in sub-classes while the geometric details increase. Comparing to the Non-hierarchical classification, this method proved to be computationally economic, and allowed to achieve higher accuracy.

Pierdicca et al. (2020) presented a deep learning method, dynamic graph convolutional neural network (DGCNN), that for the first time supports features like normal and HSV colours coupled with the x, y, z coordinates of the points. The approach proved to be effective and has been applied to ArCH (Architectural Cultural Heritage) dataset. Limitations lie in the fact that the number of classes must be set a priori and must be uniform among all models in the used dataset for training. Up to today there doesn't exist an adequate size dataset of labelled example.

3. CASE STUDIES

The Nanchan Ssu (NCS) and the Fokuang Ssu (FKS), together with other temples and monasteries in Wutai Monti in the Middle of China, constitute a UNESCO site which have been inserted in the World Heritage List in 2009. They date back to the late VIII, mid-VIII century. Both of the cases were surveyed with TLS (Trimble TX8) and UAV (DJI Phantom 4 Pro, DJI FC6310) by the members of the Archaeology Centre for Architecture, Settlement and Landscape (ACASL), Tianjin University, China.

3.1 The great hall of Nanchan Ssu

NCS (Figure 1) is located in Lijiazhuang, south-east of Wutai in Shanxi Province. The great hall built in the Tang dynasty on a 1.2m high platform among the complex, is known as the earliest existing wooden structure in China. The earliest recording written under the west beam of the central bay proves that hall was reconstructed in 782 AD.

It is a one floor building, 9.2 m in high, with a rectangular (10.5x11.7 m) shaped floor plan divided in 9 bays marked by three rows. An altar of 0.7 m was built in the central bay, above which originally stood 17 painted statues (of which 14 statues remains today). The roof is supported by 10 wooden pillars, sealed with brick walls in which two wooden window and a door have been opened. During the intervention in the 80s, two additional poles were placed inside to support the beams.

The temple was surveyed in 2017. The dataset of the great hall consists of a point cloud project produced from 78 scans stations

(13 inside the building). The entire dataset contains 1026.4 million points, was subsampled to an average and uniform resolution of 5mm (Table 1).

Figure 1. The front façade of the great hall of NCS.

3.2 The east great hall of Fokuang Ssu

FKS (Figure 2) is located in the east-north of Wutai County, Shanxi province, 6 kilometers to the Doucun town, under the mountain of Fokuang (the west foot of south part of Wutai mount). From the inscription of the stone column in front of the great hall, the founding of the east great hall dates back to 857 AD.

Figure 2. The front façade of the east great hall of FKS.

The east great hall was built on a 1.2 m high platform. A single level construction 15.4 m high, with a rectangular (34x17.66 m) floor plan divided in 21 bays marked by three rows. The 0.9 m high altar was carved from the stone mountain body under the five central bays, above which stands 35 painted tall statues that reach from 1.95 to 5.3m in height. In the surrounding bays it is possible to find other 296 statues. The roof is supported by 32 wood pillars, sealed with brick wall with 4 wooden windows and 5 doors. In recent intervention, two additional poles were added to support the rear corners of the roof. The elements of the 'Dougong' (bracket sets) that hold up the roof have a cross section measuring 0.21 x 0.3 m, stretching the roof out from the main body of the hall of about 4 meters.

It is one of the earliest architectures above ground that remains, of great value to the Chinese nation and the East Asian cultural zone of architecture.

The temple was surveyed several times from 2015 to 2020. The dataset of the east great hall is obtained from a point cloud project of 2372.1 million points, resulted from 179 scans, among which 51 are interior scans. Due to the fact that the data acquisition, the terrestrial laser scan (TLS), is highly limited inside this project, especially within the roof structure, the used dataset is

subsampled into a mean 30 mm resolution one, containing 24.3 million points (Table 1).

Scene	Classes	Points (mil)	Mean Resol. (mm)	Subsampled (mm)
NCS	5	1026.4	15	15-60
FKS	16	2372.1	30	30-120

Table 1. Summary of 3d data acquisition and process.

4. METHODOLOGY

The projects under documentation features very large dimensions. They are rich in details and have a high variety of architectural elements. The complexity and uniqueness of the two buildings (NCS and FKS) make the unsupervised classification approach, especially the DL one, not applicable. It would be difficult and time consuming to prepare enough representative samples on which the model is trained. Additionally, the condition is the same in the non-hierarchical ML classification approach; the processing of such huge datasets requires large computational resources both to extract the necessary geometric feature for training and to predict on the whole dataset. Furthermore, a non-hierarchical classification with large number of semantic classes would easily results into low accuracy.

In this work a MLMR approach, as presented by Teruggi et al. (2020), is applied to deal with two Chinese CH architectures. The result was compared with that of non-hierarchical approach. In the meantime, machine learning algorithm decision tree (DT) and RF were tested.

In a MLMR approach, a set of multi-layered semantic classes organized into a tree-like cornice is designed firstly. Each class that defines macro-architectonic elements will be subdivided into classes that can better describe every portion of such objects. In this way the classification work is distributed to different levels, where corresponding and relatively lower resolution diminishes the burden on extraction of geometric features and training. The classification results will be back interpolated to the same portion of point cloud at higher resolution that depends on the amount of geometric detail of such level. The segments allow to classify and interpolate the whole dataset iteratively, until specific portions reach full-resolution. Comparing to other common classification solutions, MLMR approach resulted to be more accurate and computationally efficient for big scale huge datasets.

A DT, as a ML classifier algorithm, is a tree-like model of decisions and possible consequences. It iteratively divides and conquers multi-class classification tasks. It calculates the Entropy or Gini impurity, which represents the probability of wrong classification from feature, of all features to obtain information gain at each step. The model risks of overfitting by growing deep with more features, which proper pruning can avoid in a certain degree. RF is an ensemble learning method. It constructs a multitude of decision trees as base learners; the RF model combines the decision of all trees. RF model surpasses DT model on overfitting. For each joint of every individual tree, decision isn't made from the feature with best information gain, but from the best one of a randomly chosen subset of m features, m is usually much less than the amount of all the features. In addition, it takes a bootstrap sampling strategy to generate sampling trainsets and use them for training the base learners.

4.1 The great hall of Nanchan Ssu

Figure 3. Classes of the point cloud for NCS project.

Considering the monitoring needs and the original point cloud resolution and quality, the classification of the great hall of Nanchan Ssu is subdivided into three different levels (Figure 3). The first level refers to the basic framework of the hall, dividing the whole dataset into: roof, support, statue and ground. To recognize these categories, the classification is performed on a 60 mm resolution version of the entire dataset. The second level deepens these categories. The roof part is divided into tile set and truss. Support category comprises pillars, windows and wall subsets. The ground is divided into altar, platform, steps and earth. At this level, all elements are processed from a 30mm resolution subsampled point cloud. The third level classification is performed to classify the truss into longitudinal, transversal and diagonal elements. The ML RF classifier used needs covariance geometric features to distinguish among elements. Functions in software CloudCompare are used to compute such features. Those in use in this project contains roughness, anisotropy, planarity, linearity, surface variation, sphericity, verticality, Z coordinates, and normals in X and Y directions.

The visualization of geometric features reflects correlations with different elements. Vertical elements are distinguished from the horizontal ones (verticality), cylindrical from the planar ones (sphericity) and surfaces composed of different patterns can be distinguished by roughness.
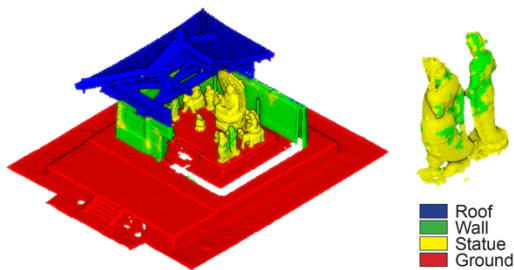

Figure 4. Classification result at level 1 of NCS.

	Roof	Wall	Statue	Ground	WGT. Average
PREC.	1.0	0.97	0.77	0.99	0.97
RECALL	1.0	0.78	0.97	0.99	0.96
F1	1.0	0.86	0.86	0.99	0.96

Table 2. Classification metrics at level 1 for NCS.

The first level of classification consists of 428,963 points, of which 78,655 have been manually annotated to be used as train set and 77,663 as evaluation set. The training of the model produced good results with weighted average of F1 score up to 0.96 (Table 2). Spreading the classification to the whole dataset, up to approximately 22% of points under the supports category were mis-classified as statues.

From Figure 4 it can be seen that around the corner of the wall, where pillars are partly exposed, similar anisotropy of the statue can be attributed lowering classification accuracy. Possibly because the algorithm generated a feature importance of anisotropy (at radius of 30 mm) up to 0.072, ranking only below that of the Z coordinates.
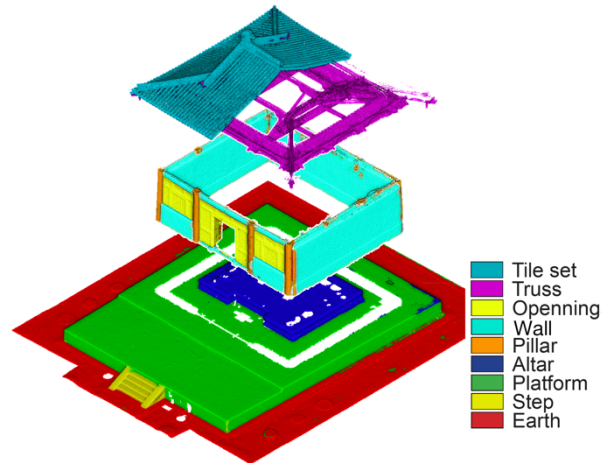

Figure 5. Classification result at level 2 in NCS.

Second level of classification on roof, wall and ground achieved satisfying results (Figure 5), with F1 score up to 0.99, 0.93, 0.99 respectively. When classifying the ground and the wall, mistakes can be seen mainly on labels attributed with fewer points, primarily when the trainset cannot provide enough points with representative geometric features.
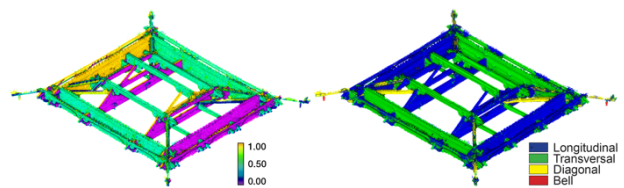

Figure 6. Results for level 3 classification on the truss of the great hall of NCS. Normals in x direction (left) and classification labels (right).

In classification level 3, the most critical point lies in the similarity on geometric features of points on the wooden elements of truss. This resulted in errors when attributing label to each point. Classification experiments showed that, connections between elements could not be correctly classified when the model for training is provided only the geometric covariance features at various radii. It is impossible to distinguish among points from similar and interweaving architectural elements.

In terms of component orientation, points may be assigned with different categories. It is necessary to enrich the geometric features array with normals computed in x and y directions with a radius of 60 mm (Figure 6). Scores upon the evaluation set achieved weighted average F1 score of 0.87 and generalized well on the whole dataset.
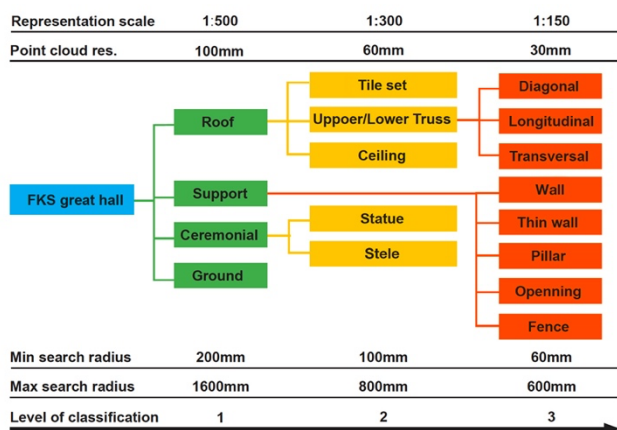
4.2 The east great hall of Fokuang Ssu

Figure 7. Classes of the point cloud for FKS project.

Considering the monitoring needs and the original point cloud resolution and quality, the classification of the great hall is subdivided into three different levels (Figure 7).

The first level refers to the basic framework of the hall, dividing the whole dataset into: roof, support, ceremonial object and ground. To recognize these categories, the classification is performed at a 100 mm resolution.

The second level is aimed to deepen the categories. Especially for the roof part, which is divided into tile set, truss, ceiling. Under the category of supports are comprised wall, thin wall, pillar, opening and fence. The ceremonial objects are divided into statue and stele. In this level, all elements are processed at a 60mm resolution.

The geometric features in use in this project contains roughness, anisotropy, planarity, linearity, surface variation, sphericity, verticality, Z coordinates. these covariance features show satisfying correlations with different elements.
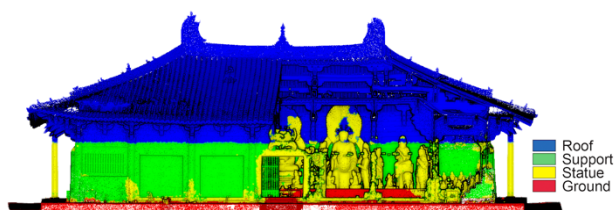

Figure 8. Classification result at level 1 for the east great hall in FKS.

	Roof	Support	Ceremonial	Ground	WGT. Average
PREC.	0.99	0.96	0.85	0.98	0.97
RECALL	1.0	0.84	0.91	0.96	0.97
F1	0.99	0.90	0.88	0.97	0.97

Table 3. Classification metrics at level 1 for FKS.

Classification level one consists of 624,049 points, with 51,028 points manually annotated as trainset. The training model tested on an evaluation set of 57,868 points has a satisfactory result. Nonetheless, since the ground truth roof category contains 400,946 points, taking up 2/3 of the dataset, the classification model turns to give Z coordinates or verticality a high weight, considering points under other categories 5.8 m above ground as

tolerable noise. It resulted into much mis-classification with the tall nimbus of the statue of Buddha (Figure 8). The problem is solved by adjusting the trainsets with more points that cover the nimbus. The result is then satisfying with F1 score up to 0.97 (Table 3).
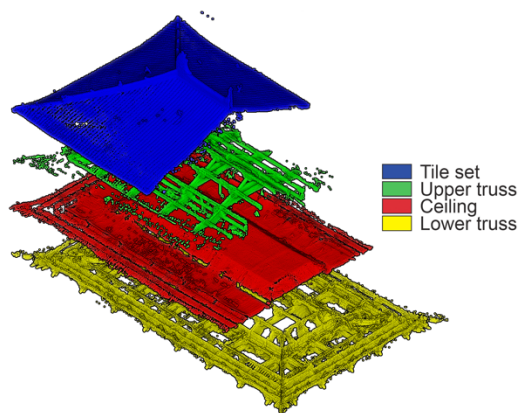

Figure 9. Classification result at level 2 on roof for FKS.

	Tile set	Upper Truss	Ceiling	Lower Truss	WGT. Average
PREC.	0.94	0.92	0.92	0.87	0.92
RECALL	0.92	0.82	0.96	0.93	0.92
F1	0.93	0.87	0.94	0.90	0.92

Table 4. Classification metrics at level 2 for FKS.

Classification level two (Figure 9) on the roof achieved 0.92 on F1 score (Table 4), the result shows clearly how the roof structure is composed: the whole edifice is covered with tile-sets, below which the wooden ceiling visually separated the truss.

Classification level three is aimed at classifying this last element. Some different sets of classes were tested to see if they improve the results (Figure 10), without adding normals in the X and Y direction as a feature to the dataset. The first trial classified elements from transversal, longitudinal, inclined ones and joints. It achieved 78% of precision on the evaluation set but, when spread to the whole dataset, the prediction didn't work well on defining elements across the bracket sets. The second, firstly separated upper truss from the lower one, classification result reaches 76% and 82% of precision, for the upper and lower truss. The upper part of the truss achieved better results at a global scene. Joints still get a high recall score.

It should be noted that classification rules should be coherent with the preparation of the feature and that the category should have a well thought structure. Therefore, when classifying only longitudinal, transversal, diagonal elements, with normals in X and Y direction computed, the result is visually clearer. It managed to classify the elements in X and Y direction, avoiding the confusion of prediction following a constructional way. Even so, the result still has a lot of confusion on classifying thin purlins around the border of the truss and points around the junctions of wooden elements; many of the longitudinal purlins were assigned to the category of transversal elements. With dataset of higher resolution and quality, better classification could be performed to define these elements.

The classification performed on the supports at 30 mm resolution datasets achieved F1 score of 0.86. Pillars (including those partly laid in the wall), walls, slab, ceiling, windows and doors are distinguished. Some horizontal reinforce rods behind the door slabs was mis-classified sharing the approximate surface of the wall.
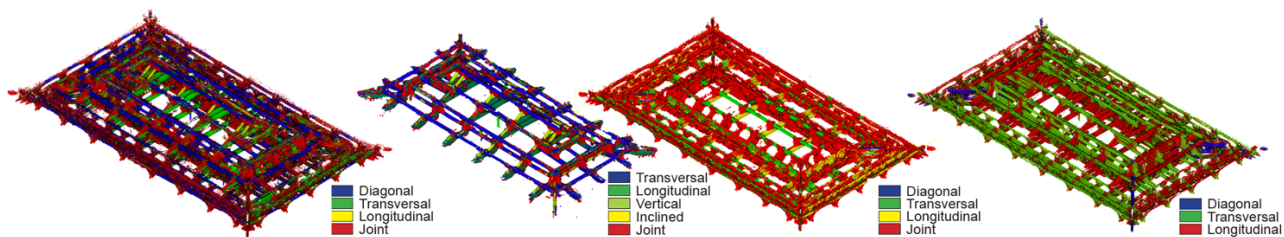
Figure 10. FKS truss classification tests in the 3rd MLMR classification level.

5. DISCUSSION

There are ways to mathematically evaluate the performance of the classification model. In most cases the accuracy gives the direct judgement of the model. Additionally, the confusion matrix, Precision, Recall, F1 score and Intersection over Union can be calculated. In every project, a manually annotated evaluation set collected from similar location to the train set is provided to evaluate the performance of the model. The visual presentation of the labels is most intuitive. In any case, it indicates mis-classified positions, where points usually share similar geometric feature values with the related mistaken labels. Observing the visual presentation helps in improving trainsets and choosing features.

5.1 Approaches

DT is simple and readable; the classification decision is represented clearly in the only tree generated by training. It can achieve results on big datasets but is often relatively inaccurate. When the model is expected to be complicated, the number of labels and features increase and the performance will no longer be satisfying. In the case study of NCS, the DT classifier gains 94.07% of accuracy when conducted on the level one classification task in MLMR approach, but following a non-hierarchical approach, the DT classifier obtained accuracy of 66.88% on the evaluation set (Figure 11). RF model has better performance in the point-based classification approach, even for unbalanced classification datasets (**Figure 12**).
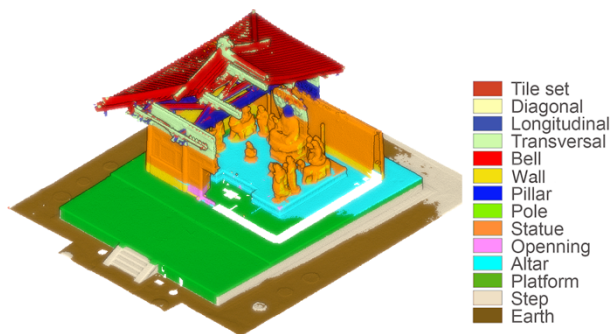

Figure 11. DT classification results for the great hall of NCS in Non-hierarchical approach.

A non-hierarchical classification approach is most direct and commonly used, it feeds the model once with the trainset of all labels. When applied to the great hall of NCS, it gets 0.9332 precision. (Figure 12). It took 21 sec. to train the model with 273,663 points, 7 sec. to predict the dataset of 1,472,351 points. The results seem pretty encouraging, with an average F1 score of 93%. However, it can't overcome the low score on pillars (Table 5), as seen in the level 2 MLMR classification.
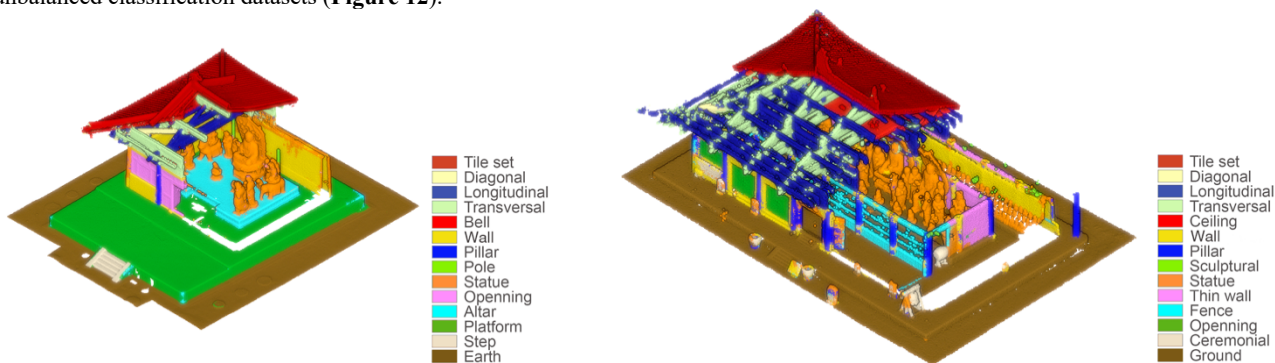

Figure 12. RF non-hierarchical classification results for the NCS (left) and FKS (right).

	Tile set	Diag. truss	Long. truss	Transv. truss	Bell	Wall	Pillar	Pole	Statue	Open.	Altar	Platf.	Step	Earth	WGT AVG
PREC.	0.97	0.88	0.94	0.80	1.00	0.97	0.84	0.99	0.85	0.94	0.99	0.97	0.99	1.00	0.94
REC.	1.00	0.59	0.74	0.93	0.82	0.82	0.41	0.39	0.98	0.98	0.97	1.00	0.72	1.00	0.93
F1	0.99	0.71	0.83	0.86	0.90	0.89	0.55	0.56	0.91	0.96	0.98	0.98	0.83	1.00	0.93

Table 5. RF non-hierarchical classification metrics for NCS.

	Fence	Diag. truss	Long. truss	Transv. truss	Ceiling	Tile set	Wall	Sculpt.	Open.	Thin wall	Cerem.	Pillar	Ground	Statue	WGT AVG
PREC.	0.93	0.48	0.62	0.29	0.87	0.92	0.87	0.93	0.79	0.73	0.52	0.76	0.98	0.77	0.79
REC.	0.90	0.04	0.45	0.52	0.90	0.92	0.88	0.52	0.89	0.52	0.31	0.53	0.98	0.88	0.77
F1	0.92	0.08	0.52	0.38	0.89	0.92	0.88	0.67	0.84	0.61	0.39	0.63	0.98	0.83	0.77

Table 6. RF non-hierarchical classification metrics for FKS.

When it is applied to predict detailed and large datasets like the east great hall of FKS, the performance is not optimal (Figure 12). Many points are not correctly predicted (Table 6). The necessity to compute many different features from small to large searching radii, resulted into higher complexity of the process. Under the same depth of trees, the performance was largely limited.

The MLMR approach performs the classification tasks at different resolutions, in each classification level, points gain a label that follows a designed tree-like structure. It avoids the heavy demand of representative datasets for deep learning training, and it also distributes the classification work to subsampled or pre-classified point clouds. As a result, points gain labels that indicate layers of semantic meanings.

In many CH projects, classification work is expected to detect objects of various scales. In most scenes, it demands high-resolution dataset to achieve full detail. The computation cost on extracting geometric features would be relatively high when performing on a full resolution dataset. In comparison, in a MLMR approach, the extraction will be conducted only on part of the whole dataset which required further classification; this avoid generating much redundant geometric information. For example, the computation on tile-sets and rubble won't be necessary in defining the roof and ground, since these parts of point cloud are definable at a relatively low resolution where it appears geometrically homogeneous. The geometric extraction should be conduct at a resolution corresponding to the detecting objects, otherwise it will waste computational resource.

The duration of this approach varies from the complexity of the project. The less hierarchy of the classes, the less difference on the duration on extracting geometric features; under this situation the manual annotation work is usually quick and clear. It will be sufficient to use the non-hierarchical approach rather than MLMR when dealing with a geometrically simple project of a small scale. On the contrary, when applying to a complex of layers of categories, the duration of extracting features will much increase. In addition, the preparation of training set will be troublesome, items under each category has to be manually pick out increasing labour cost.

5.2 Features

Features are to address the heterogeneity in point clouds. Algorithms process the data and set up correlations from labels to features with, for instance, information gain. High-quality features allow better interpreting the models and enhancing algorithm performance concerning both speed and accuracy.

Features too much closely related to labels might be given a high weight that risk overriding other features, causing mistakes. In RF model, voting process might consider trees that use features with poor relation to certain labels therefore, a casual choice of features might decrease the score of the model. It results that finding robust and relevant features should be prioritized.

Coordinates are the most basic identification of a point. In most cases, the Z coordinate is of high importance above all others. Since all architectures follow the rule by gravity, points located on different height build up different elements.

Geometric covariance features are the most used features, they represent inter-point information in a certain radius. The radius that represents the local neighbourhood should be carefully chosen to avoid excluding detailed geometric information nor having a non-numeric result (too small radius). When computing geometric features, different sets of features were generated under different radii. The combination of many radii helps the algorithm to make better decisions.

Normals can help a lot when the under-defining objects are distinguishable with its orientation from the others. When it comes to the classification of roof truss in Chinese wooden architectures, where geometric features are confusing, it contributes to achieve improved results.

6. CONCLUSION AND FUTURE WORKS

The work presented the hierarchical classification procedure, MLMR, on two Chinese CH complexes, the great hall of the NanChan Ssu and the east great hall of FoKuang Ssu. MLMR classification results have been compared with that of non-hierarchical classification approaches; DT and RF algorithms are tested. Two tested datasets have different scale and complexity and from the whole process and results, many conclusions can be drawn.

In the data processing procedure, it's tested that providing well defined and selected feature for the dataset and preparing trainsets with representative points largely affects the classification result. Regarding the algorithm, the tested DT classifier proved to be inadequate in classifying complicated datasets with numerous labels. RF classifier overcomes defect of DT on overfitting, proved to be robust in various scenes.

The non-hierarchical approach can achieve acceptable outcome when the project is relatively small and simple and has a low-number of categories subdivisions. But when applied to a more complicate project like in FKS, the results will not be optimal and require more computational resources to process the whole dataset.

The MLMR classification approach proved its capability on different types of architecture, with different materiality and scale of the space. Yet, it still meets the problems of classification rule and features definition. The labour resource is mainly used on dataset and trainset preparation, while the computational resource is primarily used on extracting geometric features. In the case study of NCS, the classification process didn't take a relatively long time (6.9 minutes for MLMR prediction while 86 minutes for manual annotation) because of the small scale of the dataset. The comparison of MLMR and non-hierarchical approach shows that the former one suits complexity and large scale, it saves large computational resource and time on feature computation, and allow manual adjustment on each level, providing a hierarchical semantic label tree for point cloud management.

Improving the quality of the point cloud in terms of attributes and density allows for more advanced classification. In the classification process, especially in the FKS project, the colour information and higher density of the roof truss area will make the architectural elements more distinguishable.

Some aspects may deserve further attention and development:

- Test and combine different segmentation algorithms to generate better results on all datasets.
- Applying other algorithms and approaches, such as model-fitting and deep-learning, to generalize various Chinese wooden architecture.
- Compiling general categories of architectural component for Chinese wooden structures.

ACKNOWLEDGEMENTS

REFERENCES

Breiman, L, 2001. Random Forests. *Machine Learning* 2001, 45, 5–32. doi:10.1023/A:1010933404324.

Coenen, M., Rottensteiner, F., Heipke, C, 2019. Precise Vehicle Reconstruction for Autonomous Driving Applications. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. IV-2-W5, pp. 21–28. doi.org/10.5194/isprs-annals-IV-2-W5-21-2019.

Griffiths, D., Boehm, J, 2019. A Review on Deep Learning Techniques for 3D Sensed Data Classification. *Remote Sensing* 2019, 11, 1499. doi:10.3390/rs11121499.

Grilli, E., Dininno, D., Petrucci, G., Remondino, F, 2018. From 2D to 3D Supervised Segmentation and Classification for Cultural Heritage Applications. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII–2. pp. 399–406. doi.org/10.5194/isprs-archives-XLII-2-399-2018.

Grilli, E., Farella, E.M., Torresani, A., Remondino, F., 2019. Geometric Features Analysis for The Classification of Cultural Heritage Point Clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-2-W15, pp. 541–548. doi.org/10.5194/isprs-archives-XLII-2-W15-541-2019.

Grilli, E., Menna, F., Remondino, F, 2017. A Review of Point Clouds Segmentation and Classification Algorithms. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W3, 339–344. doi:10.5194/isprs-archives-XLII-2-W3-339-2017.

Grilli, E., Özdemir, E., Remondino, F, 2019. Application of Machine and Deep Learning Strategies for The Classification of Heritage Point Clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-4-W18, pp. 447–454. XLII-4/W18, 447–454. doi.org/10.5194/isprs-archives-XLII-4-W18-447-2019.

He, K., Zhang, X., Ren, S., Sun, J, 2016. Identity Mappings in Deep Residual Networks. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. *Lecture Notes in Computer Science*, vol 9908. Springer, Cham. doi.org/10.1007/978-3-319-46493-0_38.

Ho, T.K., 1995. Random Decision Forests. *Proceedings of the Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995, Vol. 1, pp. 278–282 vol.1.

Liu, W., Sun, J., Li, W., Hu, T., Wang, P, 2019. Deep Learning on Point Clouds and Its Application: A Survey. *Sensors* 2019, 19, 4188. doi.org/10.3390/s19194188.

Krizhevsky, A., Sutskever, I., Hinton, G.E, 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2012, Vol. 25. doi.org/10.1145/3065386.

Mathias, M., Martinovic, A., Weissenberg, J., Haegler, S., Van Gool, L, 2012. Automatic Architectural Style Recognition. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XXXVIII-5-W16, pp. 171–176. doi.org/10.5194/isprsarchives-XXXVIII-5-W16-171-2011.

Murtiyoso, A., Grussenmeyer, P, 2020. Virtual Disassembling of Historical Edifices: Experiments and Assessments of an Automatic Approach for Classifying Multi-Scalar Point Clouds into Architectural Elements. *Sensors* (Basel) 2020, 20. doi.org/10.3390/s20082161.

Oses, N., Dornaika, F., Moujahid, A, 2014. Image-Based Delineation and Classification of Built Heritage Masonry. *Remote Sensing* 2014, 6, 1863–1889. doi:10.3390/rs6031863.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E., Lingua, A.M, 2020. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sensing* 2020, 12, 1005. doi:10.3390/rs12061005.

Teruggi, S., Grilli, E., Russo, M., Fassi, F., Remondino, F, 2020. A Hierarchical Machine Learning Approach for Multi-Level and Multi-Resolution 3D Point Cloud Classification. *Remote Sensing* 2020, 12, 2598. doi:10.3390/rs12162598.

Wang, Z., Zhang, L., Fang, T., Mathiopoulos, P.T., Tong, X., Qu, H., Xiao, Z., Li, F., Chen, D, 2015. A Multiscale and Hierarchical Feature Extraction Method for Terrestrial Laser Scanning Point Cloud Classification. *IEEE Transactions on Geoscience and Remote Sensing* 2015, 53, 2409–2425. doi:10.1109/TGRS.2014.2359951.