SHAPE BASED CLASSIFICATION OF SEISMIC BUILDING STRUCTURAL TYPES

R. Sulzer[1,*], P. Nourian[1], M. Palmieri[2], J.C. van Gemert[1]

[1] Delft University of Technology, Netherlands - (P.Nourian, J.C.vanGemert)@tudelft.nl
[2] Arup Amsterdam, Netherlands - Michele.Palmieri@arup.com

KEY WORDS: Seismic Building Structural Type, Classification, Machine Learning, Point Cloud, Shape Descriptor, Shape-DNA

ABSTRACT:

This paper investigates automatic prediction of seismic building structural types described by the Global Earthquake Model (GEM) taxonomy, by combining remote sensing, cadastral and inspection data in a supervised machine learning approach. Our focus lies on the extraction of detailed geometric information from a point cloud gained by aerial laser scanning. To describe the geometric shape of a building we apply Shape-DNA, a spectral shape descriptor based on the eigenvalues of the Laplace-Beltrami operator. In a first experiment on synthetically generated building stock we succeed in predicting the roof type of different buildings with accuracies above 80%, only relying on the Shape-DNA. The roof type of a building thereby serves as an example of a relevant feature for predicting GEM attributes, which cannot easily be identified and described by using traditional methods for shape analysis of buildings. Further research is necessary in order to explore the usability of Shape-DNA on real building data. In a second experiment we use real-world data of buildings located in the Groningen region in the Netherlands. Here we can automatically predict six GEM attributes, such as the type of lateral load resisting system, with accuracies above 75% only by taking a buildings footprint area and year of construction into account.

1. INTRODUCTION

Knowledge about the seismic vulnerability of existing building stock is of vital importance in seismic risk management, e.g. for the design and development of seismic retrofit strategies. The seismic building structural type (SBST) reflects the main load-bearing structure of a building and therefore its behaviour under seismic load (Geiß et al., 2015). However, for numerous areas in earthquake prone regions this information is often outdated, un-available, or simply not existent (Geiß, 2015). Traditional methods to gather this information, such as building-by-building inspections, are costly and highly time-consuming, making them unfeasible for assessing large building inventory. For this reason the use of remote sensing data and ancillary geo-information has been proposed to allow a fast acquisition of SBST information on urban and regional scale. Subsequently, machine learning algorithms may be used to analyse the gathered information, e.g. to classify a building stock into groups with similar SBSTs (Borzi et al., 2011, Christodoulou et al., 2017, Geiß et al., 2015, Lugari, 2014, Pittore and Wieland, 2013, Sarabandi, 2007). However, existing approaches for such a workflow often deliver highly aggregated results in terms of their spatial or typological granularity, and therefore prevent a precise seismic assessment.

The GEM foundation[1] provides an internationally standardised scheme for seismic risk assessment. Within this scheme, the GEM building taxonomy was developed, to allow the uniform classification of buildings with regard to their expected seismic behaviour. To this end, the GEM taxonomy describes a building with 13 attributes that uniquely determine its SBST, such as the lateral load resisting system, the floor or the exterior wall type.

In this research we aim to predict GEM attributes for buildings located in the Groningen region in the Netherlands. In this region seismic risk is induced by the extraction of gas from the large Groningen gas field (Muntendam-Bos et al., 2015). This leads to a unique situation in which the traditional, mainly non-resilient building stock of the Groningen region is exposed to recurring seismic events with minor intensities. Arup Amsterdam is commissioned with a large scale seismic assessment of the region. Based on existing building information, such as building plans and a small number of in situ building inspections, Arup has developed a method to determine GEM attributes for around 10% of the 250.000 affected buildings. We use this data as a training set in a supervised machine learning approach, by combining it with openly available geo-data. For now, we investigate into predictive accuracies on the training set. In the future, we want to predict GEM attributes for the entire building stock of the Groningen region.

Previous research in this field often concludes that shape features are important properties when trying to infer seismic building structural types (SBSTs). An example are pitched roofs with large spans that can be an indication for frame-based structures (Nederlandse Aardolie Maatschappij BV, 2015). However, extracting such properties from unstructured remote sensing data is often not trivial. To this end we explore a new method to describe the shape of buildings by using Shape-DNA. Shape-DNA represents the global shape of an object with the normalized beginning sequence of the eigenvalues of the Laplace-Beltrami operator (LBO). By using Shape-DNA to describe the shape of a building we expect to improve the prediction of SBSTs.

We make the following contributions. We start in a controlled environmet by generating synthetic building models in the form of polygonal meshes with different mesh densities, three different roof types and different extensions and classify them only by using Shape-DNA. Then, for the real-world Groningen SBST prediction, we first extract building features, such as the footprint area and year of construction, from a cadastral dataset. We generate 3D building models in the form of polygonal meshes by using footprint polygons and an aerial laser scanning (ALS) point cloud of the Groningen region. From the building meshes

we extract Shape-DNA and other geometric features. We then represent each building by a feature vector. Supervised machine learning techniques, such as Random Forests (RF) and Support Vector Machines (SVM) are used to learn patterns in the training set based on the feature representations and estimate the GEM attributes of buildings. We also assess the influence of each of our features for the SBST prediction, based on predictive accuracies and feature importance scores.

2. RELATED WORK

Automatic prediction of SBSTs for detailed pre-event seismic vulnerability assessment has only been attempted in a few, largely heterogeneous approaches (Borzi et al., 2011, Christodoulou et al., 2017, Geiß, 2015, Lugari, 2014, Pittore and Wieland, 2013, Sarabandi, 2007). The works differ in their input data, from terrestrial images (Pittore and Wieland, 2013) to aerial (Lugari, 2014) and satellite imagery (Borzi et al., 2011, Geiß, 2015, Lugari, 2014, Pittore and Wieland, 2013, Sarabandi, 2007), aerial laser scanning point clouds (Christodoulou et al., 2017, Geiß, 2015, Lugari, 2014) or a combination of the aforementioned. They extract topologic (Geiß, 2015, Sarabandi, 2007), geometric (Borzi et al., 2011, Christodoulou et al., 2017, Geiß, 2015, Lugari, 2014, Pittore and Wieland, 2013, Sarabandi, 2007), visual (or spectral) (Borzi et al., 2011, Geiß, 2015, Lugari, 2014, Pittore and Wieland, 2013), geographic and semantic features on building and building block level (Geiß et al., 2015, Wieland et al., 2012). Furthermore, different definitions of SBSTs, e.g. stemming from the European Macroseismic Scale (EMS) are predicted. In this paper, we build on top of the ideas of these works. However, we predict SBSTs defined by the GEM building taxonomy, which describes a building in significantly more detail than previously used typologies. Additionally, we investigate into improving the commonly used geometric features by applying a spectral shape descriptor.

Geometric shape features Geometric features are often regarded as the most important proxies to determine SBSTs. Simple features, such as the footprint area (Borfecchia et al., 2010, Lugari, 2014) or the building height (Pittore and Wieland, 2013) can already be useful to distinguish smaller from larger structures. However, to predict a more detailed description of SBSTs, a more detailed description of the building geometry may be necessary. Examples of features that describe the building geometry are the shape of the footprint or the roof type of a building. The shape of the footprint can e.g. be described by its slenderness, convexity and irregularity (Sarabandi, 2007) or compactness, elongation, density and many other parameters (Geiß et al., 2015). Most of these features are calculated by using some sort of quotient of the area, perimeter, longest axis or minimum bounding rectangle of the footprint. A way to include information about the roof of a building is to first classify the roof shape into different types such as flat simple, flat multi-level, pitched and complex (Borfecchia et al., 2010, Lugari, 2014) or flat, low slope and steep slope (Sarabandi, 2007) and then use this information as a categorical feature in the SBST classification. In our approach we do not want to first classify the roof structures into different typologies. Instead we extract shape features from a complete 3D representation of the building and directly included them in the SBST classification. We do this for a variety of reasons. Extracting local shape features, such as the roof angle, from a semantically unstructured geometric representation of a building is not trivial. For complex shapes, extensions on the building of secondary importance can

significantly influence the values of such features (see e.g. Figure 4). Moreover, it is not obvious which local properties are generally relevant for the classification of a variety of different SBSTs. Even though features, such as the roof steepness, may be suited to distinguish lateral load resisting systems, such as wall based systems from frame based systems, they lack discriminative power to distinguish them from other types. Other relevant features may be subject to common building practice or rules and regulations in a particular area or construction period. Thus, finding relevant shape features for an entire building stock can be a cumbersome task, even for an expert in the field. On the other hand, global building properties that are easier to obtain, such as the volume or the surface area of a building might often not be sufficient for describing complex building geometries. It is therefore desirable to describe the geometric shape of a building in a way that all relevant building features can be captured and extracted. To achieve this we make use of a global shape descriptor called Shape-DNA.

Visual features Several of the aforementioned studies (Borzi et al., 2011, Geiß, 2015, Lugari, 2014, Pittore and Wieland, 2013) use visual information extracted from satellite, aerial or terrestrial images. However, they often come to the conclusion that visual information is only of secondary importance compared to shape features. Furthermore, visual information can sometimes be misleading, e.g. the facade of a building does often not allow conclusions about its structure (Christodoulou et al., 2017). In this paper we do not include visual information in the classification, however, we do believe that it is worth to investigate into its use for SBST classification in the future, especially considering the vast amount of sources for this type of data.

Semantic features Semantic features may be the building age, which is often estimated using multitemporal images, or the buildings' occupancy type (residential, commercial, industrial), that can e.g. be gained from tax assessors data (Sarabandi, 2007). In this work we make use of a cadastral dataset that includes both, the year of construction and occupancy type. However, we only use the year of construction as a feature in our classification and not the occupancy type (see section 3.4). The year of construction can help to identify common building practice or rules and regulations applied at different construction eras.

Geographic features In our approach we only use building-level information for classification. One exception is a binary feature describing the presence of a directly adjacent building (see section 3.4). Generally, the building geometry allows characterising individual buildings, whereas information on building block level characterises the geographic setting in which the respective buildings are embedded in (Geiß et al., 2015). This can be valuable information in many stages of the process. In the classification process it can help to prevent misclassifications due to unlikely geometries of single structures. It can provide information on different spatial scales, also with regard to data collection. This should be addressed in future implementations.

Classifiers To predict SBSTs often a supervised learning approach is used, that takes in a training set of buildings represented by a vector containing (some of) the aforementioned features and labeled with their corresponding SBSTs. Based on the feature representation, classifiers such as a Random Forest (RF), a Super Vector Machine (SVM) (Geiß, 2015) or an Artificial Neural Network (Lugari, 2014) can find patterns in the labeled training set and predict the class of unknown buildings. One downside of the

latter classifier can be that it requires a larger amount of training samples, thus we make use of RFs and SVMs in our approach.

Seismic structural building type Next to different input data, extracted features and classifiers, a large source of variation in previous works is reflected in the predicted typologies. While some works use their own definitions of building typologies (Borzi et al., 2011, Geiß et al., 2015, Lugari, 2014), others predict typologies defined in HAZUS-MH[2] (Sarabandi, 2007), the World Housing Encyclopedia[3] or the European Macroseismic Scale[4] (Wieland et al., 2012). These typologies are often too broad (EMS), are not primarily designed for SBST classification (WHE) or focus on specific regions (HAZUS) (Brzev et al., 2013). In our approach we predict GEM typologies as they are internationally standardised and were developed to allow the uniform classification of buildings with regard to their expected seismic behaviour. The GEM taxonomy describes a building with 13 attributes that uniquely determine its SBST. For a detailed overview of all attributes we refer the reader to the GEM Building Taxonomy (Brzev et al., 2013). It is, however, important to note that the possibility to predicted a certain typology is constrained by the available training data. In a supervised learning approach it is only possible to predict typologies that are present in the learning sample. The generation of our training sample required in-depth seismic engineering knowledge and the availability of a variety of datasets (see section 3.4). Acquiring such a training sample can be challenging in many cases.

3. METHOD

3.1 Generation of building models

3.1.1 Synthetic building models In this section we report on the generation of synthetic building models from piecewise planar segments. We generate building models with flat, pitched and gabled roofs and apply different mesh densities. To this end, we randomly draw the footprint length l as $4m < l < 8m$ and width w as $5m < w < 9m$, the building height h as $3m < h < 12m$, as well as the roof angle α as $15° < \alpha < 60°$ for buildings with pitched and gabled roofs. Additionally we create models with a cubic extension with $l, w, h \in \{1m, 2m, 3m\}$. This results in a dataset with in total 1200 building models: 100 buildings with different dimensions $\times$ 3 roof types $\times$ 4 extension types.

Next we sample points of each planar building segment in a $20cm \times 20cm$ grid. We apply a Delaunay triangulation to each point segment and glue the resulting meshes together to get one polygonal mesh for each building model. For an additional dataset, we generate one mesh for each roof type where points are sampled in a $50cm \times 50cm$ grid. To resemble the meshing we achieve for the building models generated from remote sensing data we also generated one mesh without vertices on the vertical walls (cf. Figure 2e). Additionally, we remesh these models and apply Laplacian smoothing (see also section 3.1.2) to achieve a better mesh quality for the models without wall points. This last dataset will be used to assess the influence of the meshing on the Shape-DNA (section 4.1).

3.1.2 Building models from remote sensing and cadastral data While this step is not strictly necessary for SBST classification it can facilitate the extraction of 3D building information. As an example, a 3D model of the building constructed with

piecewise (planar) segments can help to extract local shape features, such as the number of roof segments or the facade area. However, even though 3D building reconstruction from remote sensing data has been an active area of research for the past 20 years, it is still a cumbersome task to reconstruct urban scenes with buildings of arbitrary shape (Haala and Kada, 2010). We do make use of one tool that is frequently applied in 3D building reconstruction: we apply a region growing segmentation to the ALS point cloud of the building roof to identify different roof segments (Grilli et al., 2017). We then use the number of roof segments and the average of all segment angles as features (see also section 3.3.2). Since further progress in reconstruction would require considerable effort, we want to, at least for now, avoid this topic. Instead, we apply a Delaunay triangulation to the ALS point cloud and the footprint polygon and merge them to gain a 3D polygonal mesh that allows the extraction of Shape-DNA. The full process is depicted in the flowchart in figure 1.
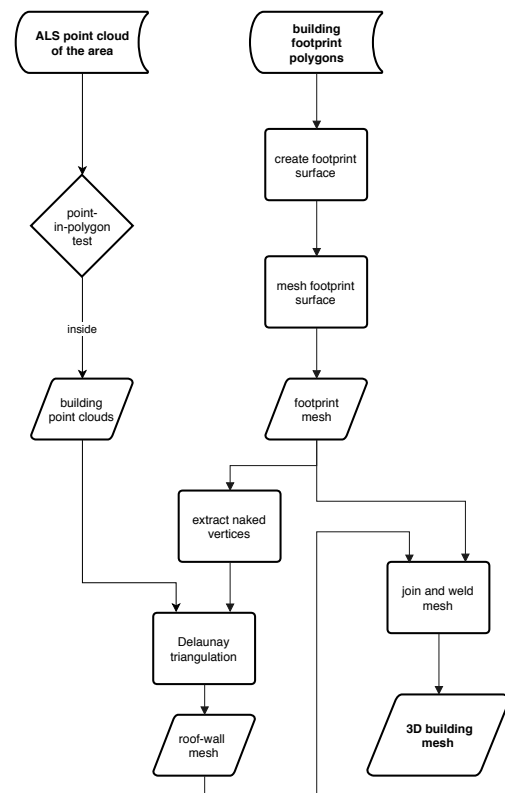

Figure 1. Flowchart of 3D building modeling from ALS point cloud and polygon footprints

In a first step, we identify the points of the area point cloud that lie inside the footprint polygons and subsequently associate these points with the corresponding footprint. This results in a "small" point cloud of every building, which we will call the building point cloud in the following. We now describe the inside of the footprint polygon with a surface (Figure 2a). We mesh the surface with the same density as the building point cloud (Figure 2b). Now we extract only the naked (outer) vertices of the footprint mesh.

Furthermore we project the building point cloud on the global XY plane in which also the footprint mesh lies (figure 2c). In this 2D space, the Delaunay triangulation provides a fast triangulation of a number of points N. We apply a Delaunay triangulation to the points, but only use edges that lie inside the footprint polygon (Figure 2d). Now we glue the upper mesh, consisting of roof

---

[2]HAZUS - https://www.fema.gov/hazus
[3]World Housing Encyclopedia - http://www.world-housing.net
[4]European Macroseismic Scale - https://www.gfz-potsdam.de/ems98
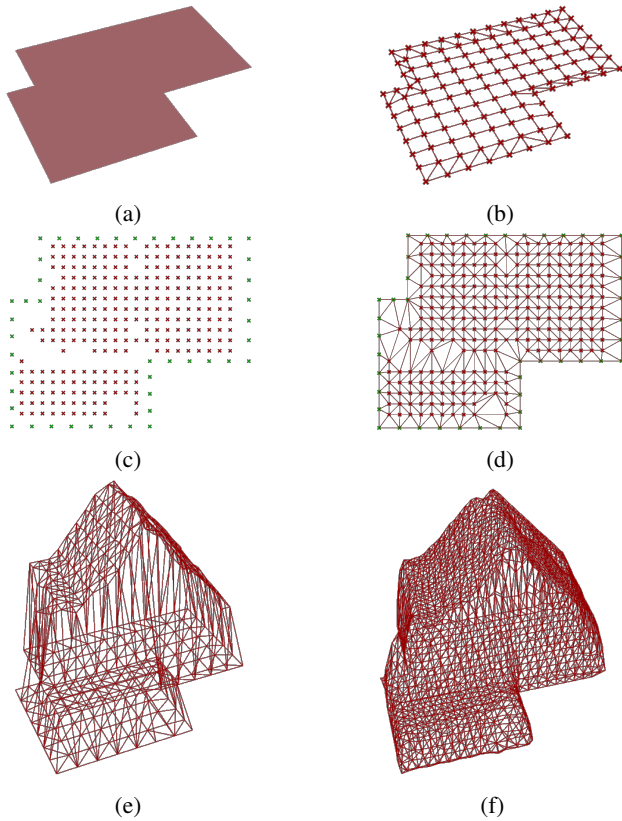
(a)

(b)

(c)

(d)

(e)

(f)

Figure 2. Generation of building mesh: (a) XY plane trimmed with footprint curve; (b) footprint approximated with mesh; (c) naked vertices of footprint mesh (green) and building point cloud projected to XY plane (red); (d) Delaunay mesh of building (top view); (e) Delaunay triangulation in 3D; (f) refined and smoothened mesh.

and walls, and the footprint mesh together. Figure 2e shows the resulting mesh in 3D.

On the building walls we don't have any points and, thus, end up with large triangles spanning from the edge of the roof to the building footprint. This is especially undesirable as the roof-wall and wall-footprint areas include important shape characteristics of the building. To combat this, we use a global mesh refinement that divides every edge of the mesh into k equidistant segments to generate k^2 similar smaller triangles (Reuter et al., 2006). Last, we iteratively apply a form of Laplacian smoothing to the mesh (Vollmer et al., 1999) to further even out the mesh density on the walls (Figure 2f). The meshes will now be used to extract the Shape-DNA of the buildings.

3.2 Shape-DNA

Shape-DNA is a global shape descriptor based on the spectrum (i.e. the eigenvalues and eigenfunctions) of the Laplace-Beltrami operator (LBO) of 2D and 3D manifolds. In the continuous case, the LBO is defined as

$$\Delta f := div(grad\ f), \quad (1)$$

where $grad$ and div are the gradient and divergence on the manifold. The Laplacian eigenvalue problem is given as:

$$\Delta f = -\lambda f. \quad (2)$$

The LBO is intrinsic and as a result, it is invariant to isometric (metric-preserving) deformations of the manifold (Masci et al., 2015). For this reason it has been used for the analysis and retrieval of non-rigid shapes (such as organic objects), where deformations are often near isometric (Lian et al., 2011). However, non-isometric transformations change the spectrum continuously, so the spectrum may also be adequate for describing the shape of buildings and subsequently classifying them.

Different discretisations of the LBO exist, such as the linear finite element method operator (Reuter et al., 2006). This discretisation can be used on a triangular surface mesh $M(V, E, F)$ with vertices $V = \{1, ..., N\}$.
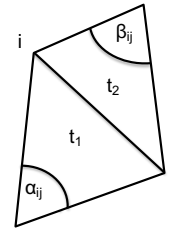

Figure 3. Example of a polygon mesh

The mesh has to be 2-manifold, meaning each interior edge $(i, j) \in E$ is shared by exactly two triangular faces t_1 and $t_2 \in F$, and boundary edges belong to exactly one triangular face (see Figure 3). In this case the LBO is given as an $N \times N$ matrix $\Delta = A^{-1}B$, where the stiffness matrix

$$A(i,j) = \begin{cases} \frac{cot\alpha_{ij} + cot\beta_{ij}}{2} & (i,j) \text{ edge,} \\ -\sum_{k \in N(i)} A(i,k) & i = j, \end{cases} \quad (3)$$

and the mass matrix

$$B(i,j) = \begin{cases} \frac{|t_1| + |t_2|}{12} & (i,j) \text{ edge,} \\ \frac{\sum_{k \in N(i)} |t_k|}{6} & i = j, \end{cases} \quad (4)$$

with $|t_i|$ being the area of the triangle t_i (Reuter et al., 2009a). The first $n \leq N$ eigenvalues of the LBO can be computed by performing the generalised eigendecomposition.

$$A\mathbf{f} = -\lambda B\mathbf{f}, \quad (5)$$

where $\mathbf{f} = (\mathbf{f}_1, ..., \mathbf{f}_n)$ is an $N \times n$ matrix containing as columns the first n discretised eigenfunctions and $\lambda = diag(\lambda_1, ..., \lambda_n)$ is the diagonal matrix of the corresponding eigenvalues (Masci et al., 2015). A truncated version of λ can be used as a feature vector that describes the geometric shape of the underlying mesh.

It is yet unclear as to what number of eigenvalues, i.e. n, should be used to form the Shape-DNA. Different publications used 11 (Reuter et al., 2009a), 20 (Niethammer et al., 2007) or 10-15 (Lian et al., 2011) eigenvalues for shape analysis and retrieval. (Arteaga, 2014) defines the LBO directly on a point cloud and reports about using 50 eigenvalues for accurate shape matching. (Gao et al., 2014) also discuss this topic and use at most 100 eigenvalues for shape description and (Reuter et al., 2009b) mentions that 500 eigenvalues had to be computed for extracting important information from eigenvalues. In view of signal processing, more eigenvalues contain more information of detail and can describe the shape more accurately (Gao et al., 2014). However,

more eigenvalues can also carry information about non-isometric deformations which might make the detection of shape similarities difficult, especially in view of building classification. We will initially make use of 50 eigenvalues simply because this seems like a good compromise between the above mentioned numbers and a feasible amount of eigenvalues to extract in view of data handling. We will not investigate deeply into the amount of eigenvalues to be used, however, we do not expect a great influence when altering this number.

Another problem is that a mesh can only approximate the underlying manifold. Simply put, the more dense a mesh, the better it can approximate the shape. Especially in areas with concavities or with high surface curvature, most eigenfunctions will need dense mesh representations (Reuter et al., 2006). However, globally dense meshes may have a large number of vertices and thus make the solution of equation 5 difficult or at least very time consuming. How dense a mesh needs to be for efficient shape analysis, such as shape classification, is problem dependent and cannot be said in general. For the building models gained from ALS point clouds and footprint data, the mesh density poses a significant problem. As we do not reconstruct the geometry with parameterised segments, the mesh density is initially limited by the point density of the ALS point cloud. Additionally, we don't have any points on the walls of the building (see section 3.1.2). As such, we end up with big triangles on the wall segments of the meshes. This is especially undesirable as the roof-wall and wall-footprint areas of a building exhibit high surface curvature. Thus, we refine and smoothen the mesh before extracting Shape-DNA.

Besides being an isometry invariant, Shape-DNA has another nice property: by normalising it, the spectrum can also be made scale-invariant. There are different methods for normalising the spectrum, such as dividing it by its first non-zero eigenvalue or multiplying it with the surface area of the underlying manifold. For identifying the roof type of the synthetic building models we make use of the first method for simplicity reasons. However, because two similarly shaped buildings with different dimensions may certainly have different SBSTs scale invariance may not always be desired.

3.3 Feature extraction

In the following two sections we report on the creation of feature vectors. We consider all features to be equally important. Thus, before we use them in a classification, we apply feature standardisation (Pedregosa et al., 2011), by subtracting the mean of all feature values and dividing by their corresponding standard deviation.

3.3.1 Synthetic building models The footprint area and footprint perimeter can simply be extracted from the length l and width w of the building models, also regarding extensions. Next, we extract the gutter height h and the surface area of the models. Because we constructed the synthetic building models with piecewise planar segments we could also easily extract the roof segment count and corresponding roof angle. However, due to the characteristics of our models, even simple rules would allow to classify the buildings with these features. Instead we aim to classify the models only by using Shape-DNA and not using any explicit information about the roof structure in the classification. Thus, we extract the first 50 normalised eigenvalues of the meshes according to the method presented in section 3.2.

3.3.2 Building models from remote sensing and cadastral data From the Groningen building models we extract the same features as for the synthetic models. Because, we have the building footprint as a polygon, the extraction of footprint area and perimeter is trivial. We approximate the gutter height of the building as the mean Z coordinate of the roof points. By applying a region growing segmentation (Grilli et al., 2017) to the building point cloud we extract the number of roof segments and the average segment angle weighted by the number of points in a segment (Figure 4).
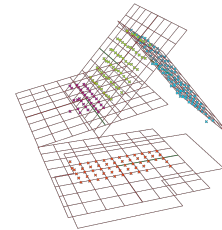

Figure 4. Region growing segmentation applied to building point cloud to extract roof segments

Furthermore, we approximate the buildings surface area as the sum of all triangle areas of the mesh. Last, we extract Shape-DNA from the building mesh. The only non-geometric feature we use for the SBST classification is the year of construction, stemming from a cadastral dataset.

3.4 Labeling according to Global Earthquake Model

Within the context of the Groningen Earthquakes Structural Upgrading project, Arup developed a methodology to gain six out of the 13 GEM attributes for around 20.000 buildings. These are (Brzev et al., 2013):

(1) The lateral load-resisting system (LLRS) describing the structural system that provides resistance against horizontal earthquake forces through vertical and horizontal components. Examples are wall based lateral load resisting systems (LWAL) or hybrid lateral load resisting systems (LH). In the latter case, different types of LLRS are combined in one structure. (2) The material of the LLRS. Examples may be concrete (CR), wood (W) or unreinforced masonry (MUR). (3,4) Both the LLRS and its material are given for the two main directions of the building. A common example are terraced houses that may have a wall based lateral load resisting system (LWAL) in the direction parallel to the street, but have no lateral load resisting system (LN) in the direction orthogonal to that. (5) The floor attribute describing floor material, floor system type, and floor-wall connection. In this paper we only describe the material of the building floor, such as wooden floor (FW), concrete floor (FC), masonry floor (FM) or other floor (FO) material. (6) The material of exterior walls describing the building enclosure. In this paper we use this parameter to denote the presence of an outer leaf wall as follows: outer leaf wall present (EW) or no outer leaf wall present (EWN).

By first expressing the material and LLRS in the main direction, material and LLRS in the second direction, floor and exterior wall type, a SBST can be represented such as the following: $MUR-LWAL-MUR-LN-FW-EW$. This is a common SBST for a terraced building unit. For the Groningen buildings these six attributes reflect the most influential parameters for assessing their seismic vulnerability. For several reasons it is not always possible to identify each of the six attributes for every

building in the labeling process. We consider the distinct combinations of the attributes as one class label and therefore only use buildings for the training set that are fully labeled with all six attributes. Otherwise there would be classes such as $unknown-LH-unknown-LH-FW-EWN$. A way to circumvent this problem and increase the labeled training set could be to predict each of the six GEM attributes separately. Although this is possible in a so-called multi-label classification, we decided to not use this technique in this paper. The reason for this is that patterns formed by classes represented with the combination of several attributes may be easier to detected than those from classes grouped by single GEM attributes. As an example, a building with concrete walls can still occur in many different shapes, while the combination with other GEM attributes describes the building more precisely. The distinct combination of the attributes results in more than 30 classes present in our dataset. However, most of these classes are only represented with around 10 to 100 samples.

The input of the labeling process stems from different sources, including detailed inspections of building plans and in situ inspections. Furthermore, a dataset containing every agricultural building in the area is included. Stemming from the dataset including only agricultural buildings, an educated guess is made towards the SBST of these buildings. All buildings in this dataset are labeled with the GEM combination $MUR-LH-MUR-LH-FW-EWN$. This guess is not always correct and may therefore also influence the measured classification accuracy.

In a previous classification all buildings that are directly adjacent to a similar building were identified. These are terraced houses which are assigned to have no lateral load resisting system (LN) in the second direction. Because we are interested in expected classification performance also for many other types, we do not include an adjacency feature in our experiments as it would immediately identify six out of the eight SBST classes as terraced houses.

4. EXPERIMENTS

For the following experiments we first use the design application Rhinoceros 3D and the visual scripting add-on Grasshopper to construct the synthetic as well as real-world building models as described in section 3.1. The data for the real-world building models, namely the cadastral dataset Basisregistratie Adressen en Gebouwen[5] and the ALS point cloud Actueel Hoogtebestand Nederland[6], is openly available at the Dutch national geoportal Publieke Dienstverlening Op de Kaart[7]. We remesh some of the models and extract the first 50 eigenvalues of the LBO of the meshes. Extracting a smaller number of eigenvalues does not significantly change the outcomes. Additionally, we normalise the eigenvalues in the first two experiments. For the last three steps we use the 'ShapeDNA-tria' software[8].

4.1 Influence of meshing on Shape-DNA

With this experiment we aim to answer the question, how density and quality of a building mesh influence Shape-DNA. We use the extracted Shape-DNA as feature vectors of the building models and plot the results using multidimensional scaling (Figure 5) - a way to visualise high dimensional feature spaces. The

---

[5] https://www.kadaster.nl/bag
[6] http://www.ahn.nl/index.html
[7] https://www.pdok.nl/
[8] http://reuter.mit.edu/software/shapedna

---

red markers represent the building models that were meshed with $0.2m \times 0.2m$ grid spacing. These meshes represent the underlying manifold most accurately. The blue markers represent models that resemble the meshes gained from remote sensing data. We can see that the Shape-DNA extracted from these models is far from the red markers and therefore considered to be imprecise. By applying mesh refinement and smoothing to these models, we can increase the precision of the Shape-DNA greatly (cyan markers). Thus, we conclude that it is possible to extract reasonably precise Shape-DNA from smoothened building meshes, that can also be gained from remote sensing data. However, the cyan markers still seam to produce a slightly different pattern then the red and green markers. In this case it also has the effect that the distance of the cyan markers in this embedding between gabled and flat roofed buildings is slightly smaller. If this poses an issue for classification will be answered in the next experiment.
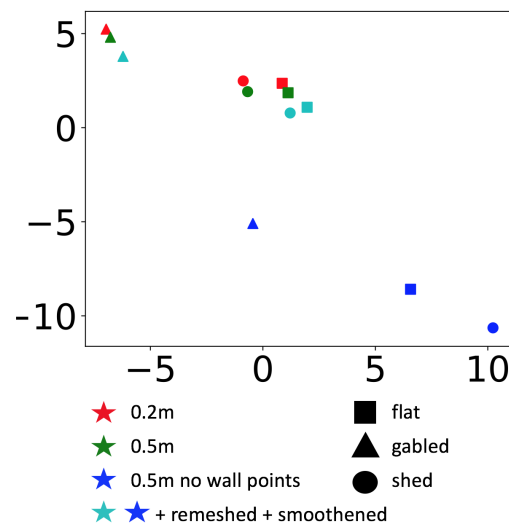
Figure 5. MDS plot of Shape-DNA of synthetic building models. The legend shows the grid spacing of the meshing and different roof types of the models. It is clearly visible that remeshing and smoothing of the models without wall points increases the accuracy of the extracted Shape-DNA.

4.2 Roof type prediction of synthetic building models

With this experiment we aim to answer the question, how Shape-DNA performs at predicting the roof type of a building. Again, we use the synthetic models that resemble the meshes gained from remote sensing data. For the classification, we use a SVM with Gaussian radial basis function kernel, as implemented in the Python machine learning library Scikit-learn (Pedregosa et al., 2011). Table 1 shows an overview of all the carried out classifications. The average classification accuracy is measured using 10 fold cross-validation.

Samples per class	Extension in m^3	Accuracy in %
100	0	93.2
100	1	90.1
100	4	79.7
100	9	75.0
100	{ 0, 1, 4, 9}	66.3
200	{ 0, 1, 4, 9}	74.3
300	{ 0, 1, 4, 9}	77.6
400	{ 0, 1, 4, 9}	79.6

Table 1. Results of roof type classification

For buildings without an extension we can predict the roof type with 93.2% average classification accuracy. With an increasing size of the extension the predictive accuracy decreases. While an extension of $1m \times 1m \times 1m$, e.g. representing a chimney, does not affect the performance significantly, bigger extensions do. For models with an extension of $3m \times 3m \times 3m$, e.g. representing a small garage, the average classification accuracy goes down to 75.0%. This is mostly due to buildings with smaller dimensions, e.g. $4m \times 5m \times 3m$, where a big extension significantly changes the building shape. Thus, the Shape-DNA features are less efficient in capturing local changes on the roof. A local shape descriptor only describing the (main) roof shape would not be affected by an extension. For four more experiments we randomly draw 100, 200, 300 and 400 buildings per roof type with all possible extension types. We can see that the classification significantly benefits from more training data. With 400 buildings per class we achieve an average accuracy of 79.6%. This proves that the SVM classifier can become partially invariant to the building extensions. Such a behaviour is desirable when classifying buildings according to their shape. Even for buildings with different dimensions and extensions the roof type can still be predicted.

In a different experiment we also classify the models according to the presence and size of their extension. Here we reach an average accuracy of 92.3% over all four classes. This proves again that Shape-DNA is a relevant shape descriptor for building shape classification. Additionally, we also conduct some of the previous experiments with meshes with a $0.2m \times 0.2m$ grid spacing, where we often gain around 3% predictive accuracy. In general, we conclude that it is possible to predicted the roof type and other shape features, such as the presence of an extension, by using Shape-DNA on 3D building models that can also be gained from remote sensing data.

4.3 SBST prediction for Groningen building models

With this experiment we aim to answer the question, how the features extracted in section 3.3.2 perform at predicting GEM attributes of real-world buildings. From the dataset described in section 3.4 we only use the eight most common classes to get a reliable measure of the classification accuracy. To provide an equal number of training samples for each class we randomly undersample the classes to 100 buildings each. In this experiment, we make use of a RF to classify the buildings, as it performs better compared to a SVM. Again, we measure average classification accuracy with a 10 fold cross-validation using different feature combinations. The RF also allows to inspect the feature importance score in the SBST classification. According to this, the year of construction and footprint area are the most important features. We will further discuss this when inspecting a confusion matrix of the classification. Table 2 shows an overview of the average classification accuracy of different feature sets.

2D		3D					Semantic	
fparea	peri-meter	gutter height	rcount	rangle	sarea	Shape DNA	yoc	Accuracy in %
•	•							60.1
•	•						•	75.5
•	•				•		•	78.3
•	•	•	•	•	•		•	79.6
•	•					•	•	75.3

Table 2. Results of SBST classification

We achieve 75.5% average classification accuracy only using footprint area (fparea), perimeter and year of construction (yoc)

features in the classification. These three features are all explicitly included in the cadastral dataset. Adding 3D features does not significantly improve the classification accuracy. Adding Shape-DNA even results in a slightly worse performance. This may be due to the fact that we only have 100 training samples per class, but potentially a variety of small shape variations inside a class. It would be interesting to see if the performance of Shape-DNA improves with more training data, similar to the behaviour we so on the synthetic models.

We inspect a confusion matrix of the best performing feature set to gain further inside (Figure 6). We can see that the masonry terraced houses ($MUR - LWAL - MUR - LN - X - EW$) achieve the lowest accuracy. The three classes with this type only differ in their floor material, which may not form patterns that are visible in the shape or age of the buildings. We still achieve good accuracies, most likely due to patterns formed by common building practice. Terraced houses with concrete walls ($CR - LWAL - CR - LN - X - X$) achieve the highest classification accuracy. These buildings can mainly be identified because they were constructed more recently compared to the masonry terraced units. Agricultural ($MUR - LH - MUR - LH - FW - EWN$) and single houses ($MUR - LWAL - MUR - LWAL - FW - EWN$) can often be identified by their larger footprint areas and thus also achieve a high classification accuracy.
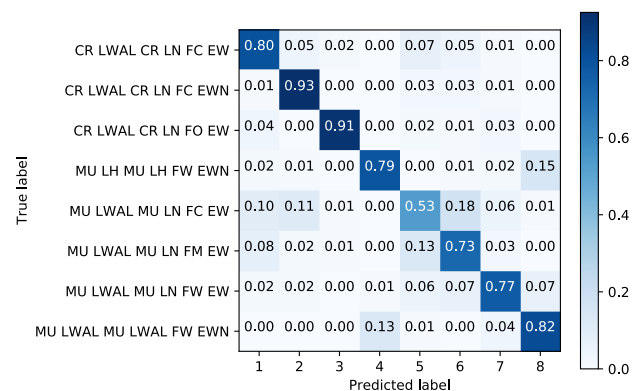

Figure 6. Confusion matrix of GEM classification.

5. CONCLUSIONS AND FUTURE WORK

In this paper we apply a new method to describe and classify buildings by using Shape-DNA. On synthetic building data, we show that it is possible to predict the roof type of buildings and the presence of an extension only by using Shape-DNA. For the real-world data from the Groningen region we predict SBSTs defined by the GEM building taxonomy. We achieve high accuracies, only using simple features, such as the footprint area and year of construction from a 2D cadastral dataset. We thereby validate previous research in SBST prediction for a sample of the Groningen building stock. Although we only predict attributes for less than 1% of the building stock in this paper, our method can in theory be applied for the entire region and possibly also for different regions. However, we can only predict SBSTs that are present in our training set. In the future we want to investigate into predicting all the classes of this dataset and improve the classification accuracy for classes that do not achieve a satisfying performance yet. We expect to achieve both by using the remaining training data and potentially adding different feature types, such as visual or geographic features to the classification. We

also tried to incorporate the shape features gained by Shape-DNA in the presented SBST classification. However, Shape-DNA was not able to improve the classification performance in our experiments. Local shape features, such as the footprint area or the number of roof segments led to better results, provided these features can be extracted. Shape-DNA may require a larger number of training samples, which is often problematic and particularly so in SBST classification. In situations where a large amount of training data is available and the aim is to classify buildings into uniform geometric groups, Shape-DNA can still be useful. Furthermore, Shape-DNA could be useful in shape retrieval problems, by using the spectrum of real-world buildings to find similar building models, e.g. in a 3D model database.

In general our paper shows that it is possible to predict building attributes according to the GEM taxonomy by using simple classification models and, thus, provide important input for rapid and precise large scale seismic assessment.

ACKNOWLEDGEMENTS

REFERENCES

Arteaga, R. J., 2014. Laplace-beltrami spectra as 'shape-dna' of surfaces using the closest point method. Master's thesis, Simon Fraser University.

Borfecchia, F., Pollino, M., Cecco, L. D., Lugari, R., Martini, R., Porta, L. L., Ristoratore, E. and Pascale, C., 2010. Active and passive remote sensing for supporting the evaluation of the urban seismic vulnerability. *Italian Journal of Remote Sensing* pp. 129–141.

Borzi, B., Dell'Acqua, F., Faravelli, M., Gamba, P., Lisini, G., Onida, M. and Polli, D., 2011. Vulnerability study on a large industrial area using satellite remotely sensed images. *Bulletin of Earthquake Engineering* 9(2), pp. 675–690.

Brzev, S., Scawthorn, C., Charleson, A. W., Allen, L., Greene, M., Jaiswal, K. and Silva, V., 2013. Gem building taxonomy (version 2.0). Technical report, GEM Foundation.

Christodoulou, A., Kokkos, A. and Palmieri, M., 2017. Automated building stock data mining and classification using open source data. In: *IASS 2017, At Hamburg, Germany*.

Gao, Z., Yu, Z. and Pang, X., 2014. A compact shape descriptor for triangular surface meshes. *Computer-Aided Design* 53, pp. 62–69.

Geiß, C., 2015. Seismic vulnerability assessment of built environments with remote sensing. http://dx.doi.org/10.18452/17104, Humboldt-Universität zu Berlin.

Geiß, C., Aravena Pelizari, P., Marconcini, M., Sengara, W., Edwards, M., Lakes, T. and Taubenboeck, H., 2015. Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *ISPRS Journal of Photogrammetry and Remote Sensing* 104, pp. 175–188.

Grilli, E., Menna, F. and Remondino, F., 2017. *A REVIEW OF POINT CLOUDS SEGMENTATION AND CLASSIFICATION ALGORITHMS*. Vol. XLII-2/W3.

Haala, N. and Kada, M., 2010. An update on automatic 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(6), pp. 570–580.

Lian, Z., Godil, A., Bustos, B., Daoudi, M., Hermans, J., Kawamura, S., Kurita, Y., Lavoué, G., Nguyen, H. V., Ohbuchi, R., Ohkita, Y., Ohishi, Y., Porikli, F., Reuter, M., Sipiran, I., Smeets, D., Suetens, P., Tabia, H. and Vandermeulen, D., 2011. SHREC '11 Track: Shape Retrieval on Non-rigid 3D Watertight Meshes. In: H. Laga, T. Schreck, A. Ferreira, A. Godil, I. Pratikakis and R. Veltkamp (eds), *Eurographics Workshop on 3D Object Retrieval*, The Eurographics Association.

Lugari, A., 2014. Active and passive remote sensing techniques and artificial neural networks in support of buildings seismic vulnerability assessment. PhD thesis, TOR VERGATA UNIVERSITY, ROME, ITALY.

Masci, J., Boscaini, D., Bronstein, M. M. and Vandergheynst, P., 2015. Geodesic convolutional neural networks on riemannian manifolds. In: *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 832–840.

Muntendam-Bos, A. G., Roest, J. P. A. and De Waal, H., 2015. *A guideline for assessing seismic risk induced by gas extraction in the Netherlands*. Vol. 34, Society of Exploration Geophysicists.

Nederlandse Aardolie Maatschappij BV, 2015. Hazard and risk assessment for induced seismicity groningen. interim update 7 november 2015. Technical report.

Niethammer, M., Reuter, M., Wolter, F.-E., Bouix, S., Peinecke, N., Koo, M.-S. and Shenton, M. E., 2007. *Global Medical Shape Analysis Using the Laplace-Beltrami Spectrum*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 850–857.

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, É., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*.

Pittore, M. and Wieland, M., 2013. Toward a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing. *Natural Hazards* 68(1), pp. 115–145.

Reuter, M., Biasotti, S., Giorgi, D., Patanè, G. and Spagnuolo, M., 2009a. Discrete laplace–beltrami operators for shape analysis and segmentation. *Computers and Graphics* 33(3), pp. 381 – 390. IEEE International Conference on Shape Modelling and Applications 2009.

Reuter, M., Wolter, F.-E. and Peinecke, N., 2006. Laplace-beltrami spectra as 'shape-dna' of surfaces and solids. *Comput. Aided Des.* 38(4), pp. 342–366.

Reuter, M., Wolter, F.-E., Shenton, M. and Niethammer, M., 2009b. Laplace–beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. *Computer-Aided Design* 41(10), pp. 739 – 755. Selected Papers from the 2007 New Advances in Shape Analysis and Geometric Modeling Workshop.

Sarabandi, P., 2007. Development of algorithms for building inventory compilation through remote sensing and statistical inferencing. PhD thesis, Stanford University.

Vollmer, J., Mencl, R. and Muller, H., 1999. Improved laplacian smoothing of noisy surface meshes. *Computer Graphics Forum* 18(3), pp. 131–138.

Wieland, M., Pittore, M., Parolai, S., Zschau, J., Moldobekov, B. and Begaliev, U., 2012. Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging. *Soil Dynamics and Earthquake Engineering* 36(Supplement C), pp. 70 – 83.