Nat. Hazards Earth Syst. Sci. Discuss., 2, 5603–5645, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/5603/2014/
doi:10.5194/nhessd-2-5603-2014

Natural Hazards and Earth System Sciences
Discussions
Open Access

UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning

J. Fernandez Galarreta, N. Kerle, and M. Gerke

University of Twente – Faculty of Geo-Information Science and Earth Observation (ITC), Enschede, the Netherlands

Received: 15 July 2014 – Accepted: 11 August 2014 – Published: 2 September 2014

Correspondence to: J. Fernandez Galarreta (jorfgal@gmail.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Structural damage assessment is critical after disasters but remains a challenge. Many studies have explored the potential of remote sensing data, but limitations of vertical data persist. Oblique imagery has been identified as more useful, though the multi-angle imagery also adds a new dimension of complexity. This paper addresses damage assessment based on multi-perspective, overlapping, very high resolution oblique images obtained with unmanned aerial vehicles (UAVs). 3-D point-cloud assessment for the entire building is combined with detailed object-based image analysis (OBIA) of façades and roofs. This research focuses not on automatic damage assessment, but on creating a methodology that supports the often ambiguous classification of intermediate damage levels, aiming at producing comprehensive per-building damage scores. We identify completely damaged structures in the 3-D point cloud, and for all other cases provide the OBIA-based damage indicators to be used as auxiliary information by damage analysts. The results demonstrate the usability of the 3-D point-cloud data to identify major damage features. Also the UAV-derived and OBIA-processed oblique images are shown to be a suitable basis for the identification of detailed damage features on façades and roofs. Finally, we also demonstrate the possibility of aggregating the multi-perspective damage information at building level.

1 Introduction

The challenges and importance of structural damage assessment, in particular its critical role in efficient post-disaster response, have placed this discipline in the spotlight of the remote sensing community (Rastiveis et al., 2013). The information generated is primarily used by Search and Rescue (SAR) teams but is also valuable for many other stakeholders engaged in post disaster activities, such as those dealing with estimation of economic loses, recovery, or reconstruction (Barrington et al., 2011).

NHESSD
2, 5603–5645, 2014

For rapid damage assessment remote sensing has been found to be very useful, as it can cover large areas, and image-based assessments are realised more rapidly than through ground deployment of appropriately skilled surveyors. However, so far it has not reached the level of detail and accuracy of ground-based surveys, a target our research aims at helping to reach. The limitations of image-based damage assessment are only partly related to the spatial resolution of the sensors. The primary problem is the vertical perspective of most operational sensors that largely limits the building information to the roofs. This roof information is well suited for the identification of extreme damage states, i.e. completely destroyed structures or, to a lesser extent, undamaged buildings. However, damage is a complex 3-dimensional phenomenon, and important damage indicators expressed on building façades, such as cracks or inclined walls, are largely missed, preventing an effective assessment of intermediate damage states.

Oblique color imagery, which shows both roof and façades, was already identified as a potential solution by Mitomi et al. (2001), who attempted to use oblique TV footage to map structural damage. Commercial oblique color data acquired by Pictometry© of post-earthquake Port-au-Prince (Haiti) were tested by Gerke and Kerle (2011a) and Cambridge Architectural Research Ltd. (CAR), among others, and were found to be more useful than conventional vertical images. However, such data also lead to challenges resulting from the multi-perspective nature of the data, such as how to create single damage scores when multiple façades are imaged. Part of the solution to these challenges lies in modern oblique data that are acquired as multi-perspective stereo pairs, which allow the generation of 3-D point clouds. These exceed standard LiDAR point clouds in terms of detail, especially at façades, and provide a rich geometric environment that favours the identification of more subtle damage features, such as inclined walls, that otherwise would not be visible, and that in combination with detailed façade and roof imagery have not been studied yet.

Nevertheless, commercial oblique imagery is typically difficult to obtain in disaster situations, and control over data acquisition with piloted aircraft (e.g., Pictometry©) tends to be limited for researchers or disaster responders. Unmanned aerial vehicles (UAVs)

NHESSD
2, 5603–5645, 2014

appear to be an alternative, especially because of their ability to obtain data at higher spatial resolution, but also because they afford more flexible data acquisition that improves the quality of the point clouds that can be derived.

The image interpretation process still typically relies on expert-based visual assessment because of the complexity of the task. Most operational post-disaster damage mapping, such as the processing of satellite data acquired through the International Charter "Space and Major Disasters", remains based on visual interpretation (e.g., Kerle, 2010; Voigt et al., 2011). While oblique airborne data should in principle allow an easier and more accurate damage assessment, owing to their comparatively high spatial resolution and more complete representation of a building, the data richness itself actually hinders more automated analysis procedures. However, there seems to be an inherent limitation of remote sensing imagery for damage assessment, regardless of type and quality: visual analysis of the Pictometry© data of Port-au-Prince by CAR also only achieved accuracy rates of 63 % when compared with ground assessment (Corbane et al., 2011; K. Saito, personal communication, 2011). Nevertheless, also expert-based visual assessment of complex data only relies on directly visible spectral indicators and relatively coarse geometric information. Combining those indicators that form the basis for visual assessment with more subtle geometric features from 3-D Point clouds may lead to better performance.

Automatic image analysis techniques for building damage assessment (BDA) can be broadly grouped into pixel- and object-based methods. In a variety of domains object-based techniques have shown advantages over pixel-based approaches (Yamazaki and Matsuoka, 2007). This tendency has to do with the spatial resolution of modern remote sensing images, where target features are clusters of pixels that are better captured by objects rather than pixels (Johnson and Xie, 2011). Additionally, object-based image analysis (OBIA, in the literature also referred to as object-oriented image analysis, OOA) adds a cognitive dimension that is expected to help in a detailed object classification.

NHESSD
2, 5603–5645, 2014

In this study we thus aimed at maximizing the potential of modern multi-perspective oblique imagery captured from UAVs, using both the high-resolution image data and derived 3-D point clouds, resulting in a detailed representation of all parts of a building. This comprehensive appraisal that approaches ground-based damage assessment in terms of complexity and completeness was coupled with a semi-automatic extraction of a range of damage indicators using OBIA. This allowed a complete characterization of the images, especially by using OBIA's cognitive dimension for the features extraction. In this study we did not yet aim at an automatic classification into per-building damage scores. Instead, our assumption was that severe damage could be determined directly from the 3-D point cloud data, while for the distinguishing of lower damage levels structural engineering expertise remains necessary. Therefore, in an earlier study (Fernandez Galarreta, 2014) we created a set of experiments to enhance the UAV images by annotating them with the OBIA-extracted damage features. Those were given to experts in ground-based damage assessment to assess the added value of the OBIA information, but also to study scoring variability and uncertainty amongs the experts.

Therefore, in the final part of this study we addressed the multi-perspective dimension of the dataset, taking into account all information collected from the façades and roofs, and aggregating it at a building level by mimicking the cognitive assessment process of ground surveyors.

2 State of the art in image-based damage assessment

Remote sensing for BDA has undergone tremendous changes over time. Its roots go back to George Lawrence and his 49 pound camera attached to a set of kites over earthquake-ravaged San Francisco in 1907, and today, companies such as Skybox (2013) can deliver HD videos from satellites. However, the challenges of BDA are only partly rooted in image type and spatial image resolution: viewing angle, understanding of the damage features, subjectivity, amongst others, are factors that also play a role in the complexity of this kind of study.

NHESSD
2, 5603–5645, 2014

The utility of almost every platform and sensor, in their multiple combinations, has been assessed for BDA. There are many examples of successful studies where the results obtained have been satisfactory and useful, e.g., Li et al. (2010) using VHR satellite imagery, Ehrlich et al. (2009) processing VHR radar imagery, or Khoshelham et al. (2013) employing aerial LiDAR datasets. For a deeper review of platforms and data types used for damage mapping see reviews by Kerle et al. (2008), Zhang and Kerle (2008), and Dell'Acqua and Gamba (2012).

For the above mentioned studies, regardless of their different sensor/platform combinations, the perspective constraint applies: the typically near-vertical perspective of sensors effectively limits the damage signature to the roofs (Gerke and Kerle, 2011a), resulting in a high dependence on proxies, e.g., changes in shadows, or evidence of blow-out debris (Kerle and Hoffman, 2013). In reality, structural damage is a phenomena expressed in all parts of the building and, in particular, the intermediate damage levels tend to display damage evidences in their façades, absence of which in vertical data constitutes a several limitation for complete BDA.

To solve this constraint, color images have been acquired from an oblique perspective to allow the evaluation of building façades. Mitomi et al. (2001) and Rasika et al. (2006) were examples of early use of this type of non-conventional imagery. However, despite studies such as by Weindorf et al. (1999) that tried to overcome low image quality issues, challenges continued to persist. Recent, more sophisticated and controlled image acquisition systems, such as Pictometry or multi-head mid-format camera systems offered by Microsoft or Hexagon, have allowed data processing based on advanced photogrammetry and machine learning principles (Gerke and Kerle, 2011b). However, besides the improvements offered by oblique imagery acquired from piloted platforms, UAVs provide additional advantages (Nonami et al., 2010): fully controlled flight, VHR imagery of up to 2 cm resolution that allows detection of fine cracks, and the large degree of image overlap that supports the generation of very detailed point clouds. However, UAVs are still in development and have to overcome a variety of issues, such as short battery life, and thus limited area of coverage, unforeseen

NHESSD
2, 5603–5645, 2014

behaviour in variable atmospheric conditions, typically limited pilot training of the user, and legislation that severely limits the use of UAVs in most countries.

As stated before, image interpretation for BDA is not trivial, especially in complex urban areas. Manual approaches constitute an easy and direct method, though with a number of constraints compared with automatic approaches. Conversely, they are capable of addressing damage holistically, i.e. in its entirety, as expert knowledge can be well matched to a given level of ambiguity and uncertainty (Rastiveis et al., 2013). Automatic approaches developed to date have struggled to deal with uncertainty inherent in damage assessment, although approaches such as by Rastiveis et al. (2013), who explored fuzzy decision making approaches, or by Li et al. (2010), who studied urban damage detection incorporating support vector machines and spatial relations, have been trying to overcome this limitation.

Within the class of automatic approaches, OBIA techniques frequently outperform pixel-based methods for reasons given above. In particular recent work aiming at automatic identification of optimal segmentation settings, e.g., Drãgut et al. (2014)'s ESP 2.0 tool, the plateau objective function of Martha et al. (2011), or research on the use of machine learning for better identification of suitable object features and for threshold/parameterization (e.g., Stumpf and Kerle, 2011) have increased the utility of OBIA for more complex automated procedures.

Besides the object-based approach, one of the most interesting advantages of OBIA is its cognitive dimension. This has already been exploited in other fields, such as landslide mapping (Lu et al., 2011), but to date has not been used for detailed BDA. This cognitive dimension aims at supporting a damage feature extraction that frequently is more conceptual than physical (Kerle and Hoffman, 2013). Damage features, due to their complexity and variability, are frequently hard to reduce to a number of parameters to describe them as a feature.

BDA conventionally makes use of a damage scale. The European Macroseismic Scale of 1998 (EMS-98; Grünthal, 1998) is a damage scale that classifies buildings from D1 (negligible damage) to D5 (total collapse). Even though it is the most

NHESSD
2, 5603–5645, 2014

commonly used damage scale for image-based BDA, the EMS-98 was originally created for ground surveys, leading to several drawbacks such as vague description of damage features and a scale based on features that do not add up linearly to a per-building damage score. Examples that illustrate the challenges of using a scale that might require a new approach in the near future.

3 Methods and data used

This study aimed at generating per-building damage scores based on oblique, multi-perspective, highly overlapping and very high resolution imagery. Those were primarily acquired with a UAV, and partly with a camera attached to a pole, (details on image acquisition are given in Sect. 3.1). From the multi-view imagery, 3-D point clouds were generated to allow visual identification of the most affected building: D4–D5 (Sect. 3.2). Subsequently, the façade and roof images of the buildings that were still standing were analysed with OBIA, where damage features were extracted (Sect. 3.3). In a separate experiment by Fernandez Galarreta (2014) the image data of buildings for which the 3-D point clouds did not reveal extensive damage, or the damage features were not visually identified, were subjected to expert assessment. Each image with overlaid information from the OBIA feature extraction being assigned an EMS-98 score and a certainty measurement. The process of aggregating the individual scores at building level, and thereby simulating the understanding of the expert surveyors on the ground, is described in detail in Sect. 3.4. Figure 1 provides an overview of the methodology.

3.1 Data used

The data for this study were collected with an Aibot X6 V.1 UAV (Fig. 2a), and with a camera attached to a 7 m pole (Fig. 2b). Several acquisition campaigns were made: Gronau (Germany), Enschede (the Netherlands) and several locations near Bologna (Italy), where an earthquake in 2012 caused extensive structural damage.

NHESSD
2, 5603–5645, 2014

The buildings mapped were used independently throughout the study for the different research elements.

The UAV flights were planned beforehand using the waypoint capability of modern UAV systems, and included both vertical and oblique image acquisition. The former was defined in a stripwise manner to achieve 80 % endlap and 30 % sidelap, using a Canon 600D with a 40 mm fixed zoom Voigtländer lens. Flying at 70 m altitude resulted in image footprint of approx. 40 m × 25 m and a nominal pixel resolution of 7 mm. The oblique flight was realized using a circular setup, i.e. to fly a circle with a radius of 70 m and a camera nick angle of 45°.

For the camera attached to the pole, a simple Canon Power Shot S100 was used to simulate an UAV flight. The camera was moved around the building at an approximate distance of 15 m to the façade using three different camera heights (3, 5 and 7 m). This resulted in images with pixel resolutions of better than 1 cm.

3.2 3-D point cloud assessment

The aim of this step was to visually identify in the 3-D point cloud a number of damage features that are related to D4 and D5: total collapse, collapsed roof, rubble piles and inclined façades. It was also meant to limit the more detailed assessment to those building without clear D4–D5 damage features expressed in their 3-D point clouds.

The test dataset used to identify the damage features comprised four 3-D point clouds (Fig. 3) generated from the oblique overlapping images as explained below.

Image processing started with the computation of camera parameters, such as intrinsic and orientation information, using a structure-from-motion approach. Musialski et al. (2013) give a comprehensive overview of state-of-the-art algorithms, such as implemented in the software Autodesk 123D Catch (Autodesk-123D, 2013). The scale of the sparsely reconstructed scene and the placement of the local coordinate system is generally arbitrary, hence subsequently a local coordinate system was defined where the z axis was chosen to point upwards. In those cases where GPS was available (for the UAV, not for the pole images), the scale and coordinate layout were defined

NHESSD
2, 5603–5645, 2014

through GPS information. Through the subsequent dense image matching (Furukawa and Ponce, 2010), the initial point cloud was substantially densified. In case of well-textured areas one 3-D point for each image pixel is achievable. The accuracy of the points depends mainly on the image configuration. In our case the standard deviation was estimated to be in the range of the pixel resolution.

Following the construction of the 3-D point cloud, for each point, a local tangent plane was computed from adjacent points. In particular the z component of the normal of this plane was of interest. It is the smallest eigenvector computed from the co-variance matrix of the neighbourhood points. The z component takes values from 0 (vertical) to 1 (horizontal) and it was converted to degrees by calculating its arcsine, it scaled the parameter from 0° (vertical) to 90° (horizontal) for better user understanding. The expected outcome was a number of D4 and D5 damage features identified in the visualization of the 3-D point cloud's z component.

More automatic approaches for BDA with LiDAR point clouds have previously been attempted (Khoshelham et al., 2013; Oude Elberink et al., 2011). However, approaches for the denser point clouds feature extraction are still being developed (Weinmann et al., 2013).

3.3 OBIA-based damage feature extraction

The goal of this step was to proceed with a more detailed façade and roof analysis of the buildings that did not show any D4–D5 damage feature in the previous step. Several algorithms were created in eCognition™ (Trimble, 2013) to extract from the images several damage features that can be expected in those façades and roofs. The importance of this section relied on three aspects: the detail of the damage assessment that was similar to that of ground-based surveys, the focus on the façade damage features that tend to be excluded in the conventional remote sensing based BDAs, and the use of OBIA to bring the cognitive dimension into the BDA framework, which helped to simulate expert-based assessment.

NHESSD
2, 5603–5645, 2014

The dataset used in this section comprised 11 selected VHR images that represented roofs and two different types of façades: concrete and brick (Fig. 4). For each of the three types, a ruleset was created to extract the damage features.

The feature extraction can be subdivided into 2 steps: image segmentation and object classification, followed by results export and accuracy assessment. But first useful damage features to be extracted from the images had to be identified. In this case the same features as in Fernandez Galarreta (2014) were selected: cracks, holes, intersection of cracks with load-carrying elements and dislocated tiles. These features are the characteristic of intermediate damage in façades and roofs. In addition, non-damage related features also had to be classified as part of the process: façade, window, column and intact roof.

1. *Image segmentation*. The aim of image segmentation was to generate meaningful damage related objects that could be easily characterized. In order to achieve that a 2-step segmentation approach was implemented. This was chosen over automatic segmentation approaches such as the Estimation of Scale Parameter (ESP 2.0; Drăguţ et al., 2014), because of the possibility of objectively using such a parameter selection process as an extra tool to capture the target objects. The 2-step segmentation started with a multiresolution segmentation algorithm, using a small scale factor (Table 1). This resulted in a desired over-segmentation meant to capture every small detail in the image, such as individual bricks, tiles and sections of cracks. The secondary parameters were modified accordingly fit the requirements of the features that wanted to be captured, individual contrasted features (Table 1). Subsequently, a spectral difference segmentation was applied on the previously generated objects. The goal was to merge the more homogeneous objects (façade and intact roof objects) into larger ones, whilst retaining the heterogeneous, damage related objects (cracks and dislocated tiles) as smaller contrasted objects for easy characterization in the next step. To achieve that different maximum spectral difference (MSD) threshold values were tested, but also here parameter sensitivity was low. Especially for the concrete façades, MSD values

NHESSD
2, 5603–5645, 2014

from 10 to 40 gave very similar results, hence transferring this 2-step approach to similar building images was expected to be straightforward. The selected parameters for the different scenarios are summarized in Table 1.

2. *Object classification*. The overall strategy to classify both façades and roofs attempted to emulate the approach of surveyors in the field. It started by classifying the largest objects first: intact roof and façade objects. Once those were classified, the rest of the classes (windows, columns, cracks, holes and dislocated tiles), more geometrically differentiable, were identified based on a number of object features (Table 1). With the basic features classified, their topological relationships were subsequently used to define their semantic dimension and, hence, identify crossing cracks (cracks crossing columns) and connecting cracks (cracks touching windows or holes).

For a more detailed explanation of the segmentation approach followed in this study and for a deeper description of the created rulesets, see Sect. 4.2 in Fernandez Galarreta (2014).

3. *Export.* The classified objects were exported to ArcGIS 10.1 (ESRI, 2013). The objects were stored as vectors with two of their features attached: area in m^2 and length in m. These stored vectors created a damage inventory with a very detailed geometric description of the extracted features.

4. *Accuracy assessment.* The accuracy assessment was based on a set of statistical measurements that compared the areas of the extracted damage features with the area of reference features digitized in ArcGIS by creating individual polygons for each damage feature found. Comparing these datasets two accuracy measurements, correctness and completeness, were derived (Fig. 5). To calculate them, three indicators were needed (Fig. 5): False Positive (FP), False Negative (FN), and True Positive (TP).

NHESSD
2, 5603–5645, 2014

The overall workflow resulted in one set of extracted damage features for each of the images, which, together with their associated information, were meant to facilitate image-based visual damage assessment. Besides the extracted objects themselves, this step also produced a number of statistical indicators used within this paper to assess the quality of the extraction. As a way of providing the damage information to the damage evaluator in your tests we experimented with a 3-D wire-mesh construct on which different damage types can be interactively switched on when needed.

3.4 Aggregation of multi-perspective damage information

In the final step an approach to aggregate multi-perspective damage information, resulting from the expert-based damage classification of roof and façades images carried out in Fernandez Galarreta (2014), was developed.

6 experts in BDA analysed different façade and roof images enhanced with the OBIA-extracted damage features, and assigned EMS-98 scores to each image. In addition they were asked to rate their classification confidence (from uncertain (1) to very certain (3)). For more information about how the experiment was set up, see the results section "Interface design and testing" in Fernandez Galarreta (2014).

This section can be subdivided into a number of steps:

1. Collection of the per-façade/roof expert-based damage classification: out of the experiment carried out in Fernandez Galarreta (2014) a table with expert-based per-façade/roof damage information was obtained.

2. Aggregation algorithms development: according to the information obtained from several field guides (Baggio et al., 2007; ATC, 2005) and interviews carried out with experts in the field, two aggregation algorithms were created to generate per-building damage scores.

NHESSD
2, 5603–5645, 2014

3. Aggregated outcome assessment: the two algorithms were applied to the expert-based damage classification table to generate aggregated damage scores and certainty measurements.

4 Results

4.1 3-D point cloud assessment

The visual assessment of the 3-D point cloud's *z* component allowed the identification of the previously listed damage features to classify a building as D4–D5 (total collapse, collapsed roof, rubble pile and inclined façade; Fig. 6).

Total collapse (Fig. 6a) can be easily identified by the absence of planar building sections. For the partially collapsed roof (Fig. 6b) the indicator was the shift towards a more vertical value of the collapsed section of the roof. Rubble piles (Fig. 6c) are recognizable from the *z* component variation, as well as the positive elevation anomaly. Finally, the *z* component deviation from the vertical readily signals inclined façades (Fig. 6d).

4.2 OBIA-based damage feature extraction

For the 11 test images, three rulesets were created. The obtained results were: roofs (Fig. 7), concrete façades (Fig. 8), and brick façades (Fig. 9).

The images shown in Fig. 8b and c were edited to remove visible damage features. The rulesets were re-run on those images to test their performance in a damage-free environment. The results are shown in Fig. 8b′ and c′.

After the classification, the results were exported to ArcGIS 10.1 as vectors with the associated attributes: area and length. These vectors represented the damage inventory environment where the experts could gain more inside about the damage features that were extracted.

NHESSD
2, 5603–5645, 2014

The results of the accuracy assessment described in Sect. 3.3 (4) are shown in Table 2.

4.3 Aggregation of multi-perspective damage information

The outcome of the experiment carried out in Fernandez Galarreta (2014), were 6 experts assessed 5 images representing a real case scenario, is summarized in Table 3.

Together with this table the experts also provided feedback on the usability of the information provided. For more detailed information on this feedback, see Fernandez Galarreta (2014) Sect. 4.3.5 "Summary of the received feedback".

An aggregation algorithm was created for the damage scores (Table 4) and the certainty measurements were simply scaled to a percentage following Eq. (1).

$$\left(\sum \text{Certainty measurements}\right) \cdot 100/15 \tag{1}$$

The result of applying the previously presented algorithm (Table 4 and Eq. 1) on the individual per-façade/roof classification (Table 3) is presented in Table 5. A total of 6 final per-building damage scores and certainty measurements were generated.

5 Discussion

Structural damage assessment is a priority after a disaster event, and the potential of remote sensing has already been demonstrated in many studies. However, the lack of methods to achieve a comprehensive damage evaluation based on all external components of a building motivated our work.

5.1 3-D point cloud assessment

We aimed at assessing whether a 3-D point cloud allows the identification of damage features indicative of D4 or D5 damage. Except for the identification of rubble piles

NHESSD
2, 5603–5645, 2014

(Fig. 6c) in situations where the grass around a building partially masked rubble presence, the results demonstrated that the visual assessment of the point cloud's *z* component was very useful to identify those features. In addition it allows experts to identify subtle damage signatures, such as inclined walls (Fig. 6d), that are difficult to recognize in traditional BDA approaches. Nevertheless, our work focused on the 3-D point could processing, with the actual damage detection still requiring manual assessment. Proper characterization of the target features in the detailed 3-D point cloud remains needed to develop more automatic approaches.

5.2 OBIA-based damage feature extraction

Previous research on BDA suggests that, regardless of the data type and quality used, the detection of intermediate damage scales remains ambiguous, being strongly influenced by the expertise and experience of the assessor. Consequently, we opted for an OBIA approach to identify damage features to assess if those can meaningfully support visual damage mapping by experts. The damage detection was largely successful, achieving acceptable correctness and completeness rates (Table 2). However, several problems were found during the segmentation and the classification, in addition to problems related to the accuracy assessment, all of them addressed more in detail below.

5.2.1 Segmentation

The 2-step segmentation aimed at generating large homogeneous non-damage objects, whilst highlighting smaller damage features, and was largely successful throughout the different scenarios (e.g., Fig. 10).

The majority of published OBIA studies suffered from limited transferability, due to the need for trial-and-error segmentation parameter adjustment. In our study the 2-step approach effectively reduces the parameter sensitivity, especially for concrete façades where a relatively large MSD threshold range led to comparable results, although more

NHESSD
2, 5603–5645, 2014

research in this direction would be needed to improve this fact. Despite the overall very satisfactory performance of the 2-step segmentation, problems remained where damage features approached non-damage background with similar spectral characteristics or patterns, such as cracks in brick walls (Fig. 11). Resulting segmentation errors propagated into the analysis stage, leading to misclassifications.

5.2.2 Classification

The classification part of the rulesets had to deal with complex scenarios, many target features and a range of different images, leading to limited, but unavoidable errors. They included both false negatives (Fig. 12a) and false positives (Fig. 12b and c). In particular brick façades, due to the noisy nature that hindered accurate segmentation, were affected. Nevertheless, in general, the feature classification was found to be very satisfactory, especially in the concrete façades. We aimed at a compromise of reaching acceptable accuracy values whilst maximising ruleset transferability. Three rulesets were applied to 11 images to test that flexibility.

5.2.3 Accuracy assessment

This research focused on finding and extracting damage features from façade and roof images to support subsequent expert-based damage classification. The aim was not an automatic delineation and extraction of those features. For this reason, an accuracy assessment of the detected features based on digitized reference objects is only partially appropriate. This is because, to our knowledge, the significance of different types of misclassification has not yet been addressed in the literature. Clearly, errors in terms of absolute length of a feature, falsely identified connectivity to specific structural building elements, or number of identified dislocated tiles, still need to be assessed from a structural engineering perspective. However, it is important to notice that although the images seem to be well classified (e.g., Fig. 8a–d), the completeness parameters are still rather low. This was due to the actual extraction process that consistently

NHESSD
2, 5603–5645, 2014

missed parts of the crack borders, which consistently led to false positives around those cracks. It is also important to notice that the extraction in Fig. 8b′ and c′ reached 100 % correctness and completeness, which is explained by the absence of damage features, which the ruleset processed correctly.

Further, traditional accuracy assessment approaches do not address the semantic dimension of the extracted features. For example, errors such as the one in Fig. 12b are flagged as FP, yet to an expert analysing damage based on the OBIA damage features this type of misclassification posed no problem, according to the feedback obtained after the expert-based per-façade/roof classification (Sect. 4.3.5 in Fernandez Galarreta, 2014).

5.3 Aggregation of multi-perspective damage information

In the final step of the methodology the damage information generated at the façade/roof level had to be aggregated at the building level. This section was successful; however, many challenges and constraints were found that raise questions concerning essential parts of this research.

5.3.1 Per-façade/roof expert-based individual damage classification

Referring to Table 3, the most important conclusion was the obvious presence of subjectivity in the classification. The 6 experts assessed the exact same simulated scenario and none of them agreed for the individual damage scores. Most agreement was found for the image of the intact façade. On the other hand, the experts tended to provide more variable damage scores for the roof image, which may indicate that the surveyors typically do not have access to the roofs, hence have limited experience in roof damage assessment.

For the certainty measurements more homogeneous tendencies were found within each expert. In general experts tended to be more for images that contained some form of damage feature, and were more uncertain when no damage features were present.

NHESSD
2, 5603–5645, 2014

This was an interesting point, because the image showing an intact façade was the one where the experts agreed the most, yet they felt less certain about it. This could be related to their experience, which would tell them that even in the absence of visible damage features, the façade might still be somehow compromised.

In addition to the table (Table 3) obtained from Fernandez Galarreta (2014), feedback was sought from the experts about the usability of the overlaid OBIA-derived damage information, using a wire-mesh construct (Fig. 13). Our assumption had been that such information was going to aid the expert's classification, reducing the ambiguity of the intermediate damage levels. However, the feedback showed that such information, because it was mainly based on spectral information, was not considered to be useful, since the same features can be readily identified by an experienced analyst in the raw images. This conclusion summarized from the experts' feedback affected directly the scope of the study. However, it must be recognized that this study not only has the potential to show the experts the features based on spectral information. It is also capable of providing information that otherwise would be invisible, such as inclined façades in the 3-D point cloud. In addition we also experimented with the possibility of identifying those damage features that affect adjacent façades, by first classifying cracks in separate façade views, and then identifying those that connect. This process mimics the holistic analysis of ground-based damage assessment, whilst eliminating the risk associated with such ground work.

5.3.2 Aggregation algorithms

Two functions were used to aggregate the individual damage and certainty scores at the building level. This asks for several assumptions to be made, and semantic rules to be defined. For the aggregation of damage scores (Table 4), in general, significant damage, even if only affecting parts of a structure, has a disproportionate significance for the performance of the entire building, also because it suggests further invisible damage. Therefore, in our study we gave priority to D4 damage elements, meaning

NHESSD
2, 5603–5645, 2014

that the presence of this score in any façade or roof determined the score for the entire building in an attempt to not underestimate the overall damage.

For the rest of damage levels a more symmetric approach was followed. Field-based BDA relies on a holistic, expert-based information integration. However, also other studies (e.g., Kerle and Hoffman, 2013) emphasised that damage evidence does not add up linearly, hence mathematic integration rules are ultimately poorly suited. To our knowledge there has been no research yet on the significance of damage indicators on adjacent or opposite façades for the overall structural integrity of the building, or to what extent observed damage pattern can be extrapolated to occluded façades. Such studies based on structural engineering principles are needed for better semantic integration of image-derived damage features to be possible.

The aggregation of the certainty measurements (Eq. 1) had to represent the expert's certainty that led to the final per-building damage score; hence, all certainty measurements had to be averaged to represent the reality of the expert assessment for that building.

5.3.3 Aggregated outcome assessment

Table 5 shows the results of the damage and certainty measurement aggregation of the expert analysis results presented in Table 3. The principal conclusion was that the algorithms were not able to reduce the subjectivity effect associated with the per-façade/roof scores as expected. The final aggregated damage scores ranged from D1 to D3. A similar effect can be seen for the certainty measurements that ranged from 40 to 80 %. Nevertheless, the certainty measurements are an excellent indicator of the source of this subjectivity effect. It can be seen in the Table 5 how different experts showed different self-confidence when tagging an image with an EMS-98 score.

Nevertheless, the goal of this study of generating more comprehensive per-building damage scores was reached. The produced scores of this study not only take into account the overall structure of the building; they also aggregate the information collected

NHESSD
2, 5603–5645, 2014

from each one of the façades and roofs of the building to provide an individual per-building damage score.

6 Conclusions and further work

In this paper we addressed a number of problems, starting with the identification of a number of principal gaps in the existing literature: (i) remote sensing-based BDA does not reach the ground-based BDA level of detail, (ii) façade assessments tend to be missed, (iii) the multi-perspective dimension of BDA has so far been relatively unexplored, (iv) UAVs, as a very detailed source of information, have not been used in this field, and (v) OBIA's cognitive dimension has not previously been exploited for BDA at such a level of detail.

We successfully used 3-D point clouds to identify D4–D5 building damage, and exploited the cognitive dimension of OBIA to assess at a detailed level damage on both façades and roofs, which is largely lacking in traditional BDA. However, in our understanding, the main constraint of this study is the actual aggregation of the damage information collected from the different parts of the building. The approach of dealing with individual façades and roofs not only failed to reduce the subjectivity of the classification. It actually increased complexity by adding the topological relationships of the damage features in the buildings. Besides, it requires the creation of aggregation algorithms to bring the information to building level, which mimics the cognitive process followed by ground surveyors.

A solution may be a building damage classification directly performed in a 3-D environment, where experts can analyse the entire building using both geometric information from the 3-D point cloud and the OBIA-based damage feature simultaneously. Nevertheless, this would still suffer from the subjectivity that characterises expert-based image analysis. In summary, more research is needed to extract automatically damage features from point clouds, combine those with spectral and pattern indicators of

NHESSD
2, 5603–5645, 2014

damage, and to couple this with engineering understanding of the significance of connected or occluded damage indictors for the overall structural integrity of a building.

References

ArcGIS: Mapping & analysis for understanding our world, available at: http://www.esri.com/software/arcgis, last access: 18 December 2013.

ATC 20-2 Appendix A: Guidelines for owners and occupants of damaged buildings, available at: https://www.atcouncil.org/pdfs/ATC202appendixA.pdf (last access: 5 August 2013), 2005.

Autodesk-123D: 3D models from photos, available at: http://www.123dapp.com/catch, last access: 19 August 2013.

Baggio, C., Bernardini, A., Colozza, R., Corazza, L., Della-Bella, M., Di-Pasquale, G., Dolce, M., Goretti, A., Martinelli, A., Orsini, G., Papa, F., and Zuccaro, G.: Field Manual for Post-Earthquake Damage and Safety Assessment and Short Term Countermeasures (AeDES), Office for official publications of the european communities, Italy, 018-5593, 2007

Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., Lin, A. Y. M., and Huyck, C.: Crowdsourcing earthquake damage assessment using remote sensing imagery, Ann. Geophys.-Italy, 54, 680–687, doi:10.4401/ag-5324, 2011.

Corbane, C., Saito, K., Dell'Oro, L., Bjorgo, E., Gill, S. P. D., Piard, B. E., Huyck, C. K., Kemper, T., Lemoine, G., Spence, R. J. S., Shankar, R., Senegas, O., Ghesquiere, F., Lallemant, D., Evans, G. B., Gartley, R. A., Toro, J., Ghosh, S., Svekla, W. D., Adams, B. J., and Eguchi, R. T.: A comprehensive analysis of building damage in the 12 January 2010 M_w7 Haiti earthquake using high-resolution satelliteand aerial imagery, Photogramm. Eng. Rem. S., 77, 997–1009, 2011.

Dell'Acqua, F. and Gamba, P.: Remote sensing and rarthquake damage assessment: experiences, limits, and perspectives, Proceedings of the IEEE, 100, 2876–2890, doi:10.1109/jproc.2012.2196404, 2012.

Drăguţ, L., Csillik, O., Eisank, C., and Tiede, D.: Automated parameterisation for multi-scale image segmentation on multiple layers, ISPRS J. Photogramm., 88, 119–127, doi:10.1016/j.isprsjprs.2013.11.018, 2014.

eCognition: eCognition. A development environment for object-based image analysis, available at: http://www.ecognition.com/, last access: 19 August 2013.

NHESSD
2, 5603–5645, 2014

Ehrlich, D., Guo, H. D., Molch, K., Ma, J. W., and Pesaresi, M.: Identifying damage caused by the 2008 Wenchuan earthquake from VHR remote sensing data, Int. J. Digit. Earth, 2, 309–326, doi:10.1080/17538940902767401, 2009.

Fernandez Galarreta, J.: Urban Structural Damage Assessment using Object-Oriented Analysis and Semantic Reasoning, M.S. thesis, University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede, 93 pp., 2014.

Furukawa, Y. and Ponce, J.: Accurate, dense, and robust multiview stereopsis, IEEE T. Pattern. Anal., 32, 1362–1376, doi:10.1109/tpami.2009.161, 2010.

Gerke, M. and Kerle, N.: Automatic structural seismic damage assessment with airborne oblique pictometry (c) imagery, Photogramm. Eng. Rem. S., 77, 885–898, 2011a.

Gerke, M. and Kerle, N.: Graph matching in 3D space for structural seismic damage assessment, IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 6–13 November 2011, Barcelona, 204–211, 2011b.

Grünthal, G.: European Macroseismic Scale 1998 (EMS-98), Cahiers du Centre Européen de Géodynamique et de Séismologie, Centre Européen de Géodynamique et de Séismologie, Luxembourg, 99, 1998.

Johnson, B. and Xie, Z.: Unsupervised image segmentation evaluation and refinement using a multi-scale approach, ISPRS J. Photogramm., 66, 473–483, doi:10.1016/j.isprsjprs.2011.02.006, 2011.

Joshi, J. R.: Improving the quality of digital surface model generated from very high resolution satellite stereo imagery by using object oriented image analysis technique, University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede, 106 pp., 2010.

Kerle, N.: Satellite-based damage mapping following the 2006 Indonesia earthquake – how accurate was it?, Int. J. Appl. Earth Obs., 12, 466–476, doi:10.1016/j.jag.2010.07.004, 2010.

Kerle, N. and Hoffman, R. R.: Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering, Nat. Hazards Earth Syst. Sci., 13, 97–113, doi:10.5194/nhess-13-97-2013, 2013.

Kerle, N., Heuel, S., and Pfeifer, N.: Real-time data collection and information generation using airborne sensors, in: Geospatial Information Technology for Emergency Response, edited by: Zlatanova, S. A. L. J., Taylor & Francis, London, 43–74, 2008.

Khoshelham, K., Oude Elberink, S., and Sudan, X.: Segment-based classification of damaged building roofs in aerial laser scanning data, IEEE T. Geosci. Remote, 10, 1258–1262, doi:10.1109/lgrs.2013.2257676, 2013.

NHESSD

2, 5603–5645, 2014

Li, P. J., Xu, H. Q., and Guo, J. C.: Urban building damage detection from very high resolution imagery using OCSVM and spatial features, Int. J. Remote Sens., 31, 3393–3409, doi:10.1080/01431161003727705, 2010.

Lu, P., Stumpf, A., Kerle, N., and Casagli, N.: Object-oriented change detection for landslide rapid mapping, IEEE T. Geosci. Remote, 8, 701–705, doi:10.1109/lgrs.2010.2101045, 2011.

Martha, T. R., Kerle, N., van Westen, C. J., Jetten, V., and Kumar, K. V.: Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis, IEEE T. Geosci. Remote, 49, 4928–4943, doi:10.1109/tgrs.2011.2151866, 2011.

Mitomi, H., Saita, J., Matsuoka, M., and Yamazaki, F.: Automated damage detection of buildings from aerial television images of the 2001 Gujarat, India earthquake, Geoscience and Remote Sensing Symposium, 2001, IGARSS '01, IEEE 2001 International, 141, 147–149, 2001.

Musialski, P., Wonka, P., Aliaga, D. G., Wimmer, M., van Gool, L., and Purgathofer, W.: A survey of urban reconstruction, Comput. Graph. Forum, 32, 146–177, doi:10.1111/cgf.12077, 2013.

Nonami, K., Kendoul, F., Suzuki, S., Wang, W., and Daisuke, N.: Autonomous Flying Robots, Springer Verlag, Dordrecht, Heidelberg, London, New York, 2010.

Oude Elberink, S. J., Shoko, M., Fathi, S. A. M., and Rutzinger, M.: Detection of Collapsed Buildings by Classifying Segmented Airborne Laser Scanner Data, ISPRS Workshop Laser Scanning 2011, Calgary, Canada, 29–31 August 2011, edited by: Lichti, D. D. and Habib, A. F., International Society for Photogrammetry and Remote Sensing (ISPRS), 2011, International Archives of Photogrammetry and Remote Sensing: IAPRS: ISPRS; XXXVIII-5/W12, 6 pp., 2011.

Rasika, A. K., Kerle, N., and Heuel, S. R. K. B.: Multi-Scale Texture and Color Segmentation of Oblique Airborne Video Data for Damage Classification, ISPRS 2006: ISPRS Mid-Term Symposium 2006 Remote Sensing: from Pixels to Processes, ITC, Enschede, the Netherlands, 8–11 May 2006, 8 pp., 2006.

Rastiveis, H., Samadzadegan, F., and Reinartz, P.: A fuzzy decision making system for building damage map creation using high resolution satellite imagery, Nat. Hazards Earth Syst. Sci., 13, 455–472, doi:10.5194/nhess-13-455-2013, 2013.

Skybox imaging: Skybox imaging. Sub-meter satellite imagery and high-definition video, available at: http://www.skyboximaging.com/ (last access: 28 December 2013), 2013.

Stumpf, A. and Kerle, N.: Object-oriented mapping of landslides using random forests, Remote Sens. Environ., 115, 2564–2577, doi:10.1016/j.rse.2011.05.013, 2011.

NHESSD

2, 5603–5645, 2014

Trimble: eCognition 8.7 Reference Book, Trimble Germany, München, Germany, 2011.
Voigt, S., Scheneiderhan, T., Twele, A., Gahler, M., Stein, E., and Mehl, H.: Rapid damage assessment and situation mapping: learning from the 2010 Haiti earthquake, Photogramm. Eng. Rem. S., 77, 923–931, 2011.
Weindorf, M., Vögtle, T., and Bähr, H. P.: An approach for the detection of damages in buildings from digital aerial information, in: Vrancea Earthquakes: Tectonics, Hazard and Risk Mitigation, edited by: Wenzel, F., Lungu, D., and Novak, O., Advances in Natural and Technological Hazards Research, Springer Netherlands, 341–348, 1999.
Weinmann, M., Jutzi, B., and Mallet, C.: Feature relevance assessment for semantic interpretation of 3D point cloud data, ISPRS Journal of Photogrammetry, Remote Sens. Spatial Inform. Sci., II-5/W2, 313–318, 2013.
Yamazaki, F. and Matsuoka, M.: Remote sensing technologies in post-disaster damage assessment, J. Earthq. Tsunami, 1, 193–210, doi:10.1142/S1793431107000122, 2007.
Zhang, Y. and Kerle, N.: Satellite remote sensing for near-real time data collection, in: Geospatial Information Technology for Emergency Response, Taylor & Francis, London, 75–102, 2008.

NHESSD
2, 5603–5645, 2014

Table 1. Segmentation parameters and classification features for the three developed rulesets.

	Parameters	Roof	Concrete façade	Brick façade
1st segmentation: Multiresolution algorithm	Scale factor	75	100	25
	Shape	0.2	0.8	0.2
	Compactness	0.2	0.5	0.2
2nd segmentation: Spectral difference segmentation	Maximum spectral difference	20	20	35
Classification	Features	Area	Area, max. difference, relative border to –, compactness, rectangular fit and length/width ratio (Trimble, 2011)	

NHESSD
2, 5603–5645, 2014

Table 2. Results of the accuracy assessment.

Accuracy parameter	Fig. 7a Dislocated tiles	Fig. 7b Intact façade	Fig. 8a Cracks	Fig. 8b Cracks	Fig. 8c Cracks	Fig. 8d Cracks
Correctness	66.6 %	100 %	65.7 %	25.9 %	87.2 %	84.8 %
Completeness	32.9 %	100 %	62.4 %	50.6 %	64.2 %	38.2 %

Accuracy parameter	Fig. 8b′ Intact	Fig. 8c′ Intact	Fig. 9a Cracks	Fig. 9b Cracks and holes	Fig. 9c Cracks and holes
Correctness	100 %	100 %	57.0 %	53.6 %	72.4 %
Completeness	100 %	100 %	45.5 %	79.1 %	89.8 %

NHESSD
2, 5603–5645, 2014

Table 3. The collected damage scores and certainty measurements from the 6 experts that carried out the interface test in Fernandez Galarreta (2014).

					
Damage scores					
Expert	**Façade 1**	**Façade 2**	**Façade 3**	**Façade 4**	**Roof**
Expert 1	3	3	2	2	1
Expert 2	2	2	1	1	1
Expert 3	3	3	3	1	2
Expert 4	3	2	2	1	3
Expert 5	3	3	3	1	1
Expert 6	2	3	2	1	2
Legend	D4: very heavy damage	D3: heavy damage	D2: moderate damage	D1: negligible damage	
Certainty measurements					
Expert	**Façade 1**	**Façade 2**	**Façade 3**	**Façade 4**	**Roof**
Expert 1	3	2	2	3	3
Expert 2	2	2	2	3	2
Expert 3	2	2	2	2	2
Expert 4	3	3	2	2	1
Expert 5	2	2	2	1	1
Expert 6	1	1	1	2	1
Legend	3: very certain	2: quite certain	1: uncertain		

NHESSD
2, 5603–5645, 2014

Table 4. Description of the algorithm created to aggregate the per-façade/roof damage score at the building level.

Per-façade/roof damage score	Per-building damage score
One or more D4 evidences	D4
> 50% of the façades are D3	D3
< 50 % of the façades are D3	D2
> 50% of the façades are D2	D2
< 50% of the façades are D2	D1

NHESSD
2, 5603–5645, 2014

Table 5. Results of the aggregation of the individual expert-based per-façade/roof classifications.

Expert	Aggregated damage score	Aggregated certainty measurements
Expert 1	2	86 %
Expert 2	1	73 %
Expert 3	3	66 %
Expert 4	2	73 %
Expert 5	3	53 %
Expert 6	2	40 %

NHESSD
2, 5603–5645, 2014

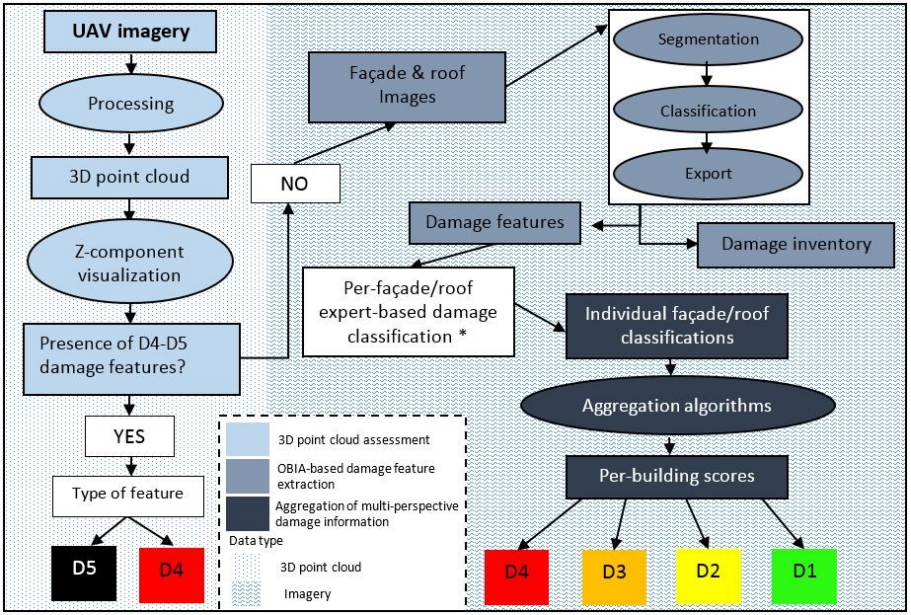

Figure 1. Overview of the methodology. * Fernandez Galarreta (2014).

NHESSD
2, 5603–5645, 2014

UAV-based urban structural damage assessment

J. Fernandez Galarreta et al.

Figure 2. The two platforms used to collect data. **(A)** Aibot X6 V1 UAV, **(B)** camera attached to a pole.

NHESSD
2, 5603–5645, 2014

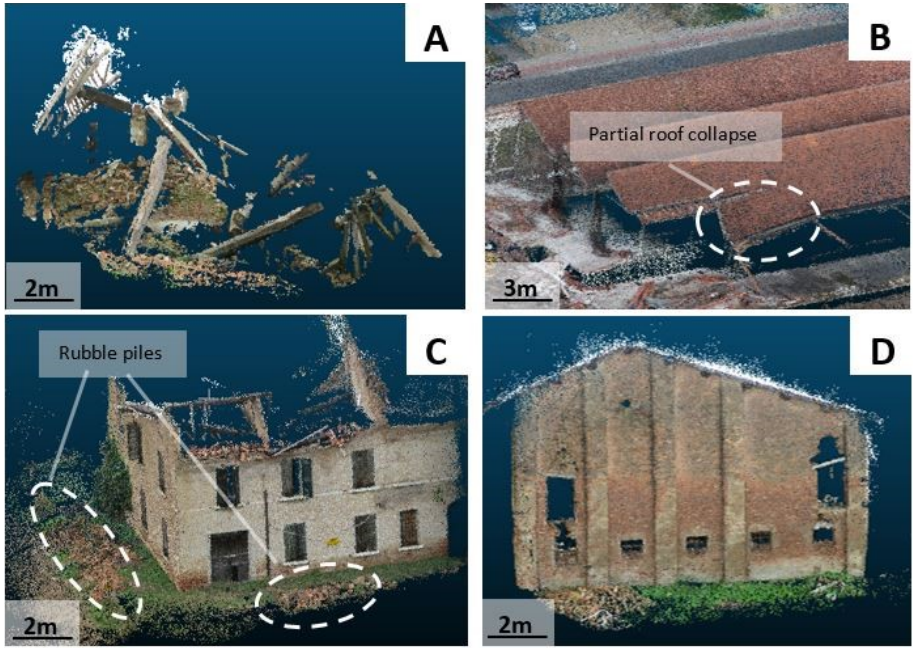

Figure 3. The 3-D point clouds used to identify the mentioned damage features. **(A)** Total collapse (Italy: pole-based), **(B)** partly collapsed roof (Germany: UAV-based), **(C)** rubble pile (Italy: pole-based), and **(D)** frontal view of an inclined wall (Italy: pole-based). Scale approx.

NHESSD
2, 5603–5645, 2014

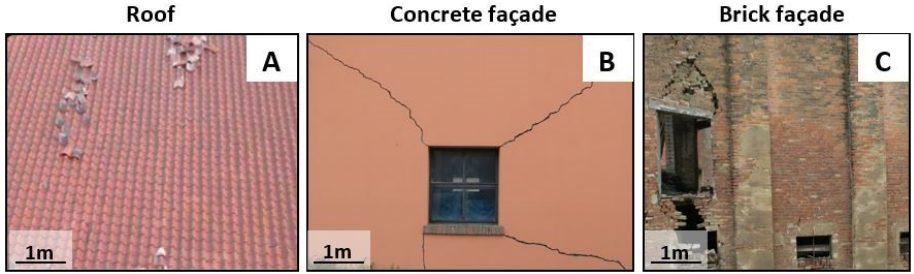

Figure 4. Examples of the different images used in this section. **(A)** Roof with dislocated tiles, **(B)** cracks in concrete façade, and **(C)** cracks and hole in brick façade. Scale approx.

NHESSD
2, 5603–5645, 2014

$$Correctness = \frac{TP}{TP + FP}$$

$$Completeness = \frac{TP}{TP + FN}$$

Figure 5. Equations for the correctness and completeness accuracy measurements based on the accuracy indicator: false positive (FP, red), false negative (FN, blue), and true positive (TP, green) (Joshi, 2010).

NHESSD
2, 5603–5645, 2014

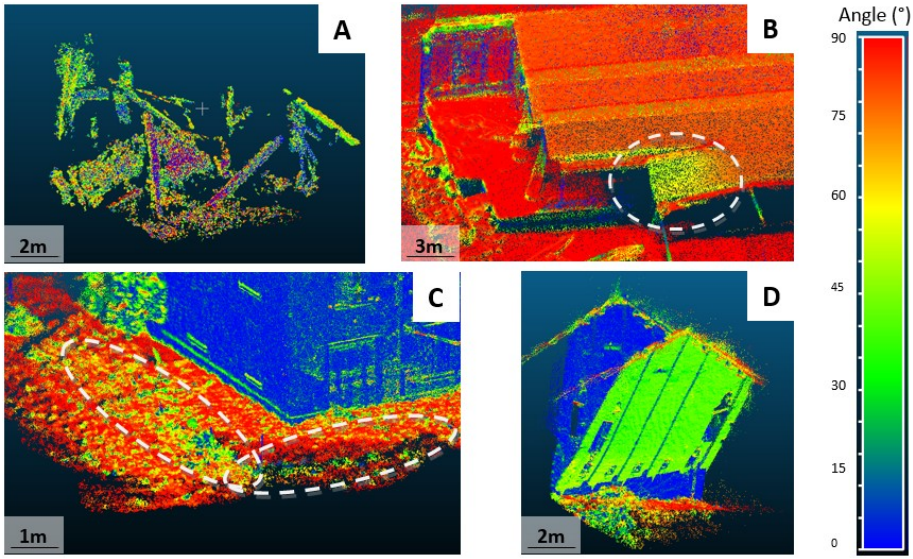

Figure 6. *z* component visualization of target damage features: **(A)** collapsed building, **(B)** partially collapsed roof (outlined in white), **(C)** rubble pile (outlined in white), and **(D)** vertical façade (blue) compared to a simulated inclined façade (green). Scale approx.

NHESSD
2, 5603–5645, 2014

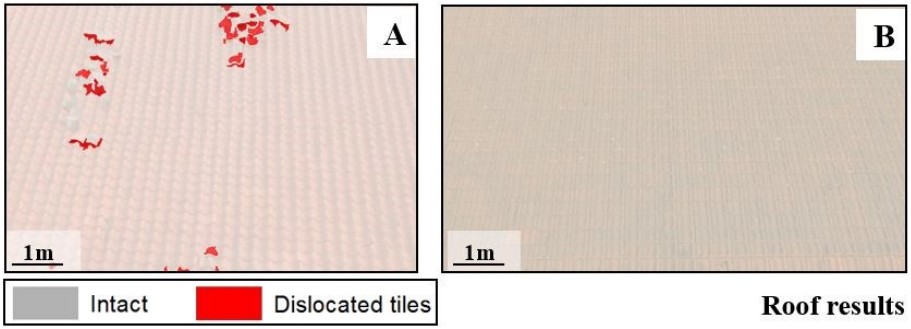

Figure 7. Result of applying the roof ruleset on 2 roof images. Scale approx.

NHESSD
2, 5603–5645, 2014

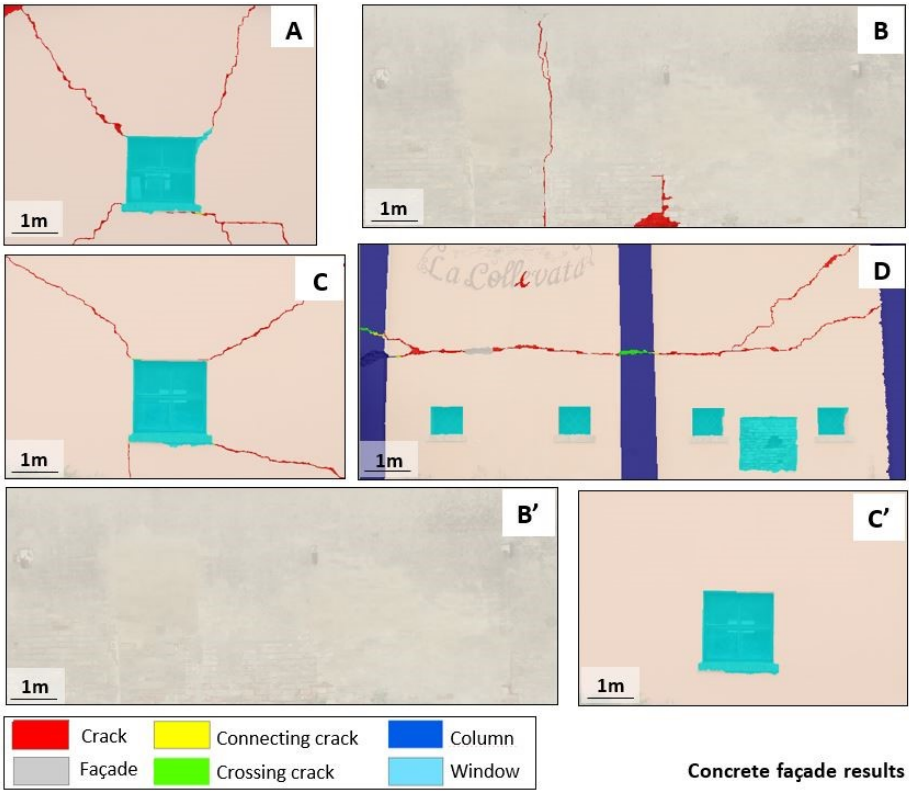

Figure 8. Results of applying the concrete ruleset on 6 concrete façade images. **(B′)** and **(C′)** were edited in order to remove the damage feature. Scale approx.

NHESSD
2, 5603–5645, 2014

UAV-based urban structural damage assessment

J. Fernandez Galarreta et al.

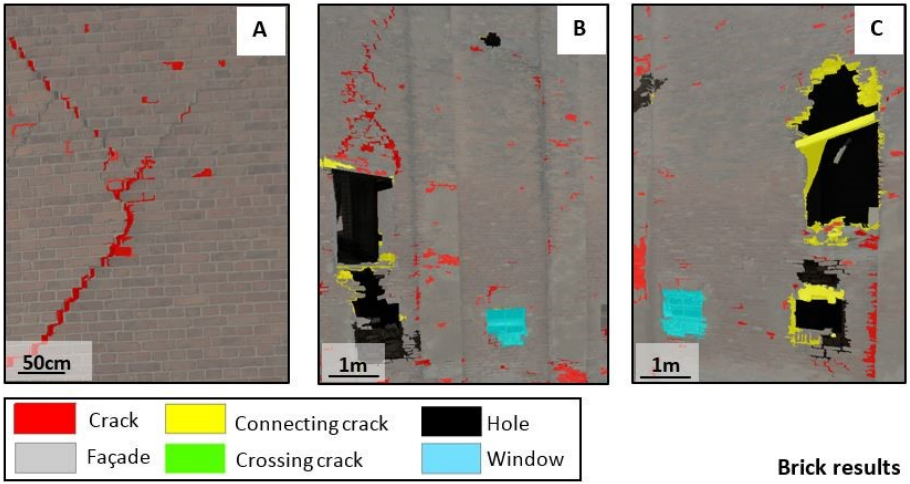

Figure 9. Results of applying the brick ruleset on 3 brick façade images. Scale approx.

NHESSD
2, 5603–5645, 2014

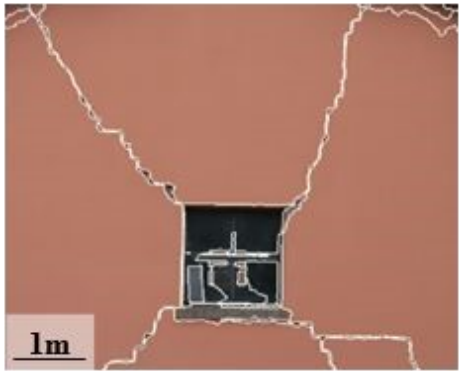

Figure 10. Result of the 2-step segmentation of a concrete façade. Scale approx.

NHESSD
2, 5603–5645, 2014

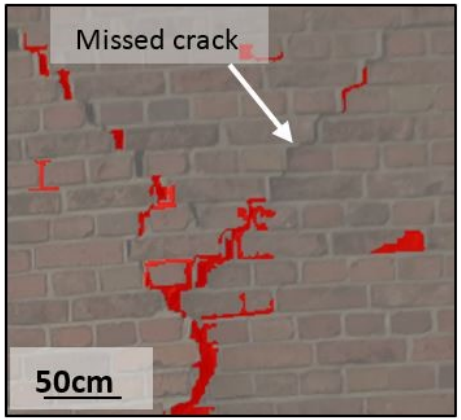

Figure 11. Example of misclassified crack due to segmentation problems. Scale approx.

NHESSD
2, 5603–5645, 2014

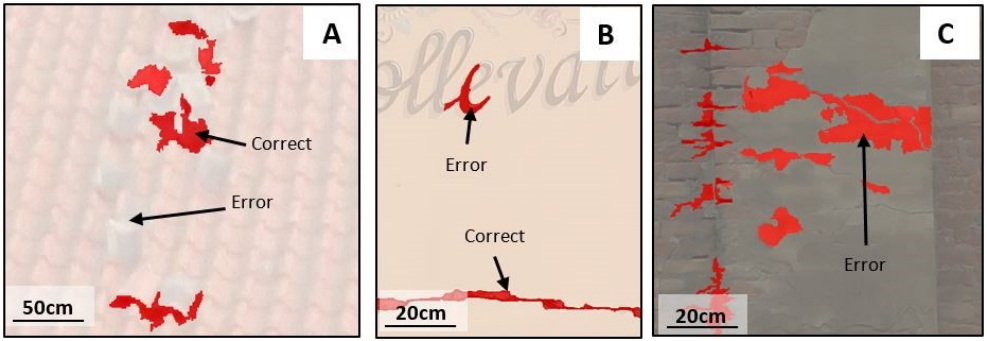

Figure 12. (A) Example of misclassified roof tiles (false negative), **(B)** example of a letter classified as a crack in a concrete façade (false positive), and **(C)** example of non-related objects classified as cracks in a brick façade (false positive). Scale approx.

NHESSD
2, 5603–5645, 2014

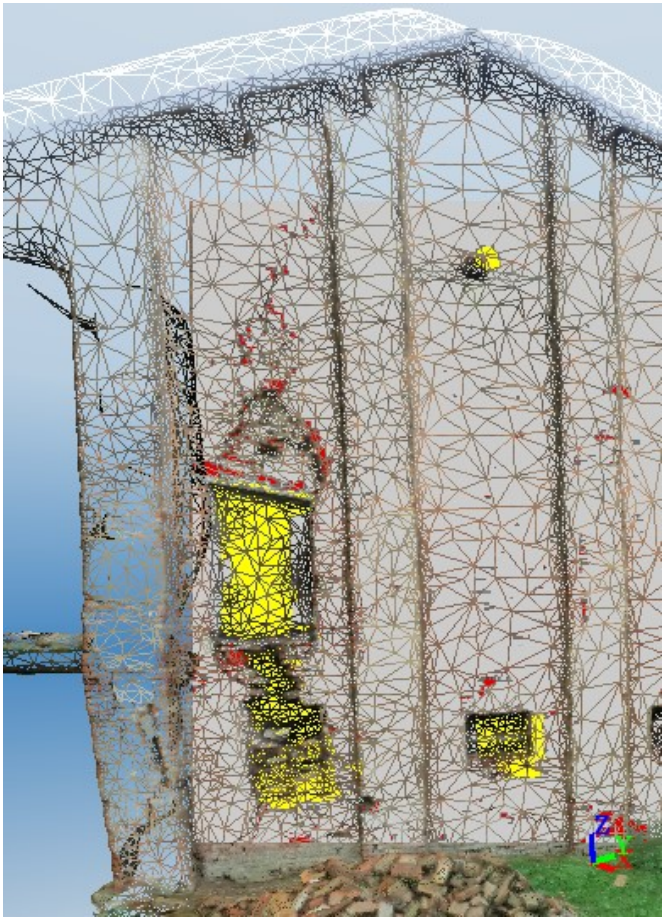

Figure 13. Example of the wire-mesh with OBIA extracted information overlaid. Yellow indicated holes and red cracks in the façade (for better contrast the color scheme was changed from the previous examples).

NHESSD
2, 5603–5645, 2014