Contextual classification using photometry and elevation data for damage detection after an earthquake event

Ewelina Rupnik\(^a\), Francesco Nex\(^b\), Isabella Toschi\(^c\) and Fabio Remondino\(^c\)

\(^a\)Ecole Nationale des Sciences Géographiques (ENSG), Marne la Vallée, Saint Mandé, France; \(^b\)Faculty of Geo-Information Science and Earth Observation ITC, University of Twente, Enschede, The Netherlands; \(^c\)3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy

**ABSTRACT**
This research presents a processing workflow to automatically find damaged building areas in an urban context. The input data requirements are high-resolution multi-view images, acquired from airborne platform. The elevations are derived from a dense surface model generated with photogrammetric methods. With the principal objective of rapid response in emergency situations, two different processing roadmaps are proposed, semi-supervised and unsupervised. Both of them follow a two-step workflow of building detection and building health estimation. Optionally, cadastral layers may serve as a-priori knowledge on building location. The semi-supervised approach involves a data training step, while the unsupervised approach exploits the similarities and dissimilarities between sets of features calculated over the detected buildings. The change detection task is formulated as a classification task defined over a conditional random field. The algorithms are evaluated using two datasets (Vexcel and Midas cameras) and results are compared with ground truth data and specific metrics.

**Introduction**

**Overview**
In recent years, Earth Observation (EO) systems (satellite as well as aerial acquisitions) became a valuable asset at the disposal of emergency response teams in the case of natural disasters and hazards (earthquake, tsunami, landslide, oil spills, flooding, hurricane, volcanic eruption, etc.). It has been recognized and endorsed by the public, voluntary and private sector (e.g. European Union, United Nations offices and programmes, The International Charter on Space and Major Disasters, DLR-ZKI, GRSS, ITHACA, INSPIRE, GSDI, DigitalGlobe, MDA, Airbus) (Boccardo & Tonolo, 2014; Clark, Holliday, Chau, Eisenberg, & Chau, 2010; Dong & Shan, 2013; Mahmood, Bessis, Bequignon, Lauritson, & Venkatachary, 2002; Roche, Propeck-Zimmermann, & Mericskay, 2013) in a number of worldwide activities [e.g. HORIZON2020 Secure Societies, Disaster Management and Emergency Response (UN-SPIDER), UNITAR’s Operational Satellite Applications Programme (UNOSAT), Eye on Disaster Management)] to promote geospatial research and innovation as an instrument to aid the post – disaster operations (Altan et al. 2013).

The present research is the outcome of an international cooperation between Europe and Japan within the RAPIDMAP (RAPID MAPping) project funded under the Capacities Programme of the EU FP7 "Resilience against Disasters" (Cho et al., 2014). This publication is focused on the processing of high-resolution airborne optical imagery and shows how to exploit such datasets for automatically deriving contextual information about collapsed buildings in urban areas. Besides the image photometry, the algorithms strongly rely on the elevations derived from photogrammetric processing (see “Experiments” section). With respect to existing approaches (see “Related works” section), the presented methods (i) automatically detect structures destroyed by earthquake events using aerial imagery, (ii) improve the contextual classification by combining geometric and photometric information in a Higher Order Conditional Random Field Framework, (iii) limit the number of used features to decrease the computational effort needed to deliver a map (without reducing the quality of the achieved results) and (iv) consider different input data scenarios to simulate real cases.

This paper is organized as follows. The section “Related work” reports on existing approaches for change detection after some natural disaster using optical data. In the “Developed processing framework” section, the theoretical base for the employed classification method, an extensive introduction to the feature extraction part, as well as the developed processing workflows are discussed. The “Experiments” section

**CONTACT** Ewelina Rupnik ewelina.rupnik@ign.fr

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
demonstrates the performance of the algorithms in operation: two post-earthquake datasets are evaluated: San Felice (Italy, 2012) and Aquila (Italy, 2009). Finally, the discussion of the results and the final conclusions and outlook of the presented work.

Related works

Optical remotely sensed datasets are more straightforward for visual interpretation, and therefore are suitable for the crowd-sourcing campaigns and prevailing in rapid mapping (Kerle, 2010). Satellite imagery is preferred in change detection applications when a large area should be surveyed. Automated satellite-based approaches perform mostly by comparing the multi-epoch data in a supervised or unsupervised way. The supervised scenarios classified on the base of image-to-image analyses (Rosu, Pierrot-Deseilligny, Delorme, Binet, & Klinger, 2015; Voigt et al., 2007), detect changes by incorporating prior knowledge, e.g. in the form of cadastral vector layers, or subtract respective nDSMs to extract objects (Guerin, Binet, & Pierrot-Deseilligny, 2015; Poli & Caravaggi, 2013; Zhou, Parsons, Elliott, Barisin, & Walker, 2015). Samadzadegan and Rastiveisi (2008) experimented high-resolution satellite images from QuickBird satellite over the city of Bam, Iran. The data were available in two time epochs: before and after the earthquake occurrence. The building locations were known from a vector map and indicated where the texture features shall be extracted. The building condition was finally assessed comparing the 2D features in pre- and post-event imagery. When no previous cases are available, unsupervised change detection based on pre- and post-event data is possible as well (Dalla Mura, Benediktsson, Bovolo, & Bruzzone, 2008; Pesaresi, Gerhardinger, & Haag, 2007). Guerin et al. (2015) worked with Worldview-I and IKONOS imagery collected over the city of Phoenix, USA, and Sendai, Japan, respectively. In either case, two epoch datasets were simultaneously bundle adjusted, and two separate DSM were obtained. The change detection was defined as a labelling problem on a differentiated DSM, and solved with dynamic programming, considering the contextual information in pixel neighbourhood.

As far as aerial platforms and UAVs are concerned, the benefit is the more flexible data acquisition, both in terms of time and flight pattern. The on-board sensors have a much higher spatial resolution than satellite sensors, making them more suitable for detailed inventory. Rezaeian and Gruen (2011) proposed a methodology to detect man-made damages using two-epoch analogue aerial stereo pairs. Image intensity values and 3D information from the generated DSM served directly and indirectly as the damage assessment evidence. Partial and total collapses were recognized. Guo, Lu, Ma, Pesaresi, and Yuan (2009) identified collapsed building after the 2009 earthquake in Wenchuan China, with only post-event data captured by an aerial push-broom camera. They employed image processing techniques based on a morphological operator to treat the RGB images channel by channel. The result was a composite image with the colours corresponding to the delineated classes (including the damaged building). Gerke and Kerle (2011) exploited the added value of aerial oblique photographs, acquired over Port-Au-Prince (Haiti) in 2010. The traditional nadir images accompanied with the slanted angles allowed a more complete assessment of the building’s condition. Following a two-step supervised classification, every detected building gained a score whether damaged or not. 2D and 3D features were extracted from rooftops and façades mapped from oblique views. Similarly, Vetrivel, Gerke, Kerle, and Vosselman (2015) addressed the question of building change detection in point clouds derived from oblique cameras. Having delineated building candidates, the gaps in the 3D data were interpreted on the basis of a set of photometric features. Airborne and UAV imagery after the Mirabello (Italy) earthquake in 2012 was incorporated in the study. Rapid mapping conducted from lower altitudes with UAVs was also deployed for reconnaissance purposes, for instance after the Katrina hurricane in 2005, L’Aquila earthquake in 2009, Morakot typhoon in 2009, Haiti earthquake in 2011, Sendai earthquake in 2011 and Hurricane Sandy in 2012 (Adams & Friedland, 2011; Baltsavias, Cho, Remondino, Soergel, & Wakabayashi, 2013; Shanley, Burns, Bastian, & Robson, 2013).

Refer to Kerle, Heuel, and Pfeifer (2008) for a review of different airborne sensors used in mapping hazards and natural disasters, and Joyce, Belliss, Samsonov, McNeill, and Glassey (2009) for available space-borne solution.

Developed processing framework

The approach focuses on damaged building identification and it is based on three successive steps: (i) pre-processing, (ii) building candidate recovery (cf. “Buildings candidates recovery” section) and (iii) ranking of the building’s health (cf. section “Building health ranking”) (Figure 1). The general framework was primarily inspired by the work of Montoya-Zegarra, Wegner, Ladický, & Schindler (2014), who proposed a “recover-and-select” strategy to infer buildings and roads from high-resolution aerial images.

We propose two different strategies: semi-supervised and unsupervised. The first method assumes a training phase in the building recovery step (cf. the first pre-classification in Figure 1), followed by the health ranking in unsupervised manner. As half of the workflow is supervised, and the other half unsupervised, we refer to it as semi-supervised. The second one delineates building candidates on the basis of one,
site-specific parameter, to be set by the user – the mean building height (cf. the first pre-classification in Figure 1). Finally, if a cadastral map containing the building layer is available (delivered by e.g. a national mapping agency), the approach is purely unsupervised. Irrespective of the approach to recover the building candidates, ultimately the building health classification problem is formulated in a probabilistic graphical model, defined as Conditional Random Fields (CRFs). In all three scenarios, the workflow initiates with the data pre-processing, i.e. the generation of a normalized Digital Surface Model (nDSM) and an orthophoto with their derived information: the 3D and 2D features. See the workflow illustrated in Figure 1.

All adopted algorithms’ implementations were available from open-source codes: Random Forests (RFs) and K-means were part of the Open source computer vision library (OpenCV) (Bradski & Kaehler, 2008), α – expansion for robust PN potentials as described in Kohli, Ladicky, and Torr (2008) and the max – flow code from Boykov and Kolmogorov (2004).

Rationale for building damage detection

The earthquake intensity (magnitude and wave typology) as well as building materials and structures (i.e. concrete, wood, breaks or metals) are two main factors that condition the extent and the typology of building damages. As a consequence, the appearance of building damages is case-dependent and has a huge variability. When damage’s evidence from nadir aerial views is considered, two main typologies of buildings can be detected (Figure 2): (i) the holes/gaps on the rooftops and (ii) the building debris (clutters) around the building and in correspondence of severely damaged parts of the construction. These two typologies of damages appear very different: holes/gaps are usually dark, untextured, with both regular and irregular contours and relatively limited extent. On the other hand, building debris are brighter regions, with irregular shapes and contours as well as randomly distributed texture patterns. The detection of these two typologies of regions is conducted using a dual strategy (Figure 2) combining the complementary information provided by 2D and 3D features. Untextured regions can be hardly modelled by image matching, while bright and well textured regions can be reliably reconstructed. Shadowed areas can be easily detected from images and they can be differentiated from the other untextured surfaces.

Feature selection

The adopted approach is based on the hand-crafted feature selection as coined in Tokarczyk, Wegner, Walk, and Schindler (2013), being features produced from an “educated guess”, given the data and the problem at hand. Alternatively, one could adopt abundant and “blind” feature sets, leaving the selection of the significant features to the classifier, e.g. AdaBoost. Our philosophy is in part determined by the specificity of the data and practicality. At this scale, the variability of the image content in urban landscape is enormous, and providing the classifier with a training sample big enough to model all classes of interest becomes a cumbersome task. Then, there is no means to pre-define the type of damaged buildings. The practical objective of the work is the rapid mapping of urban areas after catastrophic events: the use of a limited number of features allows fulfilling the time requirements as well. Considering all the above, we have knowingly devised a limited number

---

**Figure 1.** The three-step processing workflow. The building candidates (building mask) are provided from the building cadastral map or generated from data. The likelihood of every object point to be a particular class is calculated based on learned data (supervised) or parameters set by the user (unsupervised). To select the candidates, one thresholds the building likelihood. In the last stage, the features corresponding to building candidates are grouped into clusters that shall delineate intact building, damaged building and other. The dashed and dotted boxes refer to first and second pre-classification, respectively. *If unsupervised HMEAN is employed, a vegetation mask is applied on all features.*
of significant spectral and geometrical features to explain the data.

In the “Feature selection” section, the spectral features extracted from the true orthophoto as well as the geometric features extracted from the generated photogrammetric DSM are described in detail. These features support the detection of holes and debris regions within the defined modelling approach.

**Spectral features**

**Image segmentation.** A two-step approach has been developed, (i) a super-pixel segmentation algorithm and (ii) a clustering procedure, in order to aggregate adjacent regions together and reduce the over-segmentation problems.

The SLIC (Simple Linear Iterative Clustering) super-pixel segmentation algorithm (Achanta et al., 2012) has been adopted as it has shown to overcome most of the existing methods. The clustering procedure is performed according to a two-step approach: the DBSCAN (Ester, Kriegel, Sander, & Xu, 1996) algorithm and a refined aggregation step. DBSCAN performs a first aggregation in order to reduce the number of segments. The successive refined aggregation takes advantage from previous similar implementations (Shakelford & Devis, 2003), defining a homogeneity index between adjacent regions where both spectral and regions’ shape information are aggregated together.

**Repetitive pattern feature (RP).** Macro-texture elements on urban elements cannot be detected when satellite or low-resolution aerial images are considered. On the other hand, when high-resolution images (Ground Sampling Distance GSD < 15 cm) are considered, image patterns can be generally divided in three main classes: (i) repetitive, (ii) untextured and (iii) randomly distributed. Repetitive and untextured patterns can be present on intact building and they depend on the material

---

**Figure 2.** Recognized damage characteristics (a). In image space (spectrally) the damage is represented as bright (c) or dark (b), and considered to always carry non-repetitive texture. In object space (geometrically), it is represented by roof-face planarity, the height above the ground (collapsed/standing) and the regularity of the borders. Both bright and dark damages may take either of the geometrical representations.
and typology of roof. Random patterns are instead concentrated in cluttered areas, where blobs of different colours are chaotically distributed (Figure 4).

The repetitive pattern index detects and quantifies the presence of repeated textons (Julesz, 1981) in the scene. The higher will be the percentage of repetitive patterns in each pixel neighbourhood, the higher will be the feature value. The detected structures are characterized by regular sequences of brighter and darker pixels (Figure 4, first row). The index is computed according to an algorithm very similar to the one presented in Yalniz and Aksoy (2010). Compared to this implementation, an additional averaging step is performed in order to reduce the variability of the achieved results on neighbouring image regions.

In order to preserve regions boundaries, the image segmentation will be used as constraints and the sliding windows will be stopped once the borders are reached.

**Texture intensity feature (TI).** The repetitive pattern feature can distinguish irregular spectral variations from repetitive patterns, but very few information can be retrieved when untextured regions are considered. To distinguish among untextured and randomly textured regions, a texture intensity index has been developed. It is defined as the mean value of the absolute gradients over a sliding window. Its values are normalized between 0 and 1 to manage the variability of textures intensities in different images. The values are generally higher in correspondence of repetitive and chaotic textures, while they will be lower on untextured regions (Figure 4, fourth row).

**Morphological feature (Morpho).** Dark regions can be detected using morphological operators. In particular, the morphological closing by reconstruction (Dalla Mura, Benediktsson, Waske, & Bruzzone, 2010) provided satisfactory results for our task. This algorithm performs a geodesic reconstruction by erosion of the dilation of the input image using the original image as a mask. Defining the minimum and maximum sizes of the regions to be detected, it is possible to extract dark regions of this size in the orthophoto. In our tests, dark regions of size comprised between 0.5 and 5 m could be consistently detected (Figure 4, fifth row).

**GRVI index (GRVI).** Vegetation can be reliably detected using the well-known NDVI index if the orthophoto’s infrared (IR) channel is available. If it is not available, the Green–Red Vegetation Index (GRVI; Motohka, Nasahara, Oguma, & Tsuchida, 2010) can be adopted for the detection of vegetated areas. The index is comprised between −1 and 1 and values higher than 0 are assumed to be in correspondence of vegetated areas. To remove the vegetation, a multiplication with binary mask has been performed. The mask holds true for all positive GRVI indices and false otherwise.

**Spectral feature integration.** Within the defined model, a damage can be seen in the images either as bright or dark region. With this hypothesis, various spectral filters deliver varied responses for what corresponds to a single class – the damage. To recognize both, by successfully establishing the decision boundary within the clustering algorithm – the damages and nothing more grouped into a unique cluster – the two responses must be pre-processed to bring them close in the feature space. The goal is accomplished by creating a so-called spectral feature that integrates the RP, TI and Morpho features depending on the response of the texture and the size of a segmented region (Figure 4). The TI feature per se allows to discern between flat (untextured) and rough textures. A flat surface (in the spectral sense) showing no repeating structure assimilates the appearance of an (i) industrial site (Figure 4(a)), (ii) plain residential roof-faces (Figure 4(d)) or (iii) the dark roof damages (Figure 4(b)). With the no-repetitiveness condition maintained, but changing to a rough surface it will resemble a bright damaged area (Figure 4(c)), or less probably a building roof with an unordered pattern.

Still, within the group of buildings with flat texture, the black roof and a collapsed building area pictured as black are both considered by the morphological index the same (Figure 4, fifth row). To separate them, we incorporate a further branch into the decision tree, stating that the black spots corresponding to collapsed areas must be small (area below 20 m²). Employing the three features plus the image segmentation into a set of rules helps to select the appropriate feature to explain a pixel, whether interpreted as belonging to a healthy or damaged building. Figure 3 illustrates the applied rules in form of a decision tree, and Figure 4 presents the responses of given features on the modelled rooftop types.

**Geometric features**

3D geometry can be a valuable instrument to discriminate different parts of an urban scene completing the information provided by spectral data. From a Photogrammetric DSM the roofs and road regions can be easily distinguished thanks to their medium/large size and their local flat and regular shape. On the other hand, the debris is usually characterized by irregularities in shape and size (Khoshelham, Oude Elberink, & Xu, 2013; Oude-Elberink, Shoko, Fathi, & Rutzinger, 2011). The employed 3D geometric features are shortly described in the following.
Point cloud segmentation. The point cloud segmentation is implemented according to a region growing algorithm. The aggregation criteria are given by a distance threshold in Z (i.e. 1 GSD) between adjacent points. Damaged areas can be often segmented in smaller regions but their aggregation on wider segments is needed for a more efficient feature computation. Smaller areas (<0.5 m²) are therefore selected and aggregated to other small regions according to a topological aggregation rule (see Figure 5(b)).

Figure 3. Spectral feature pre-processing. The spectral feature takes either the morpho or the RP amplitudes. A non-damaged site is said to manifest high repetitiveness. In places of low repetitiveness, low texture intensity, the morpho feature is adopted. It exhibits low amplitudes for blackish (damaged), and high amplitudes for luminous (non-damaged) areas. Thanks to the segment size, black non-damaged roofs and black damaged roof parts are separated. High TI and RP are a signature of damaged sites (such as in Figure 4); thus, the RP response is adopted.

Figure 4. Representation of different rooftops and their spectral feature responses. Left: intact (no repeating structure or texture), damaged (bright and dark collapsed areas), damaged (bright, collapsed area) and intact buildings, respectively. Rows: original orthophoto, segmented orthophoto, repetitiveness, texture intensity and morpho features. The grey, orange and blue rectangles correspond with the colours of the decision tree in Figure 3 and indicate how features are integrated into the final spectral feature.
Normalized DSM extraction (nDSM). The normalized DSM generation is performed by subtracting DSM and DTM. The DTM is extracted according to an iterative filtering process similar to the one presented in Beumier and Idrissa (2016).

Non-Planarity (NP) and 3D roof segments size (nS). The off-ground regions are initially selected for the computation of these features. The normal vector is estimated for each point and points are grouped in different classes according to their horizontal component. Points belonging to the same class are then used as seeds of a planar segmentation algorithm. A plane fitting approach similar to the one suggested in Vosselmanm (2012) is implemented.

Main planes are iteratively defined on the building roofs. Small regions are aggregated together using a proximity criterion. The nS index is computed considering the number of points belonging to each region. For each planar segment, the plane equation is then determined. The signed distance of each point from the fitted plane is finally computed to estimate the Nonplanarity index (Figure 5(d)). Both indexes give an indication of the segment status: high Nonplanarity values and small nS are usually in correspondence of damage regions (Figure 5(e)). Due to the extreme variability of each region size nS is normalized in the interval [0 1].

Contour irregularity (CI). This feature is computed on the off-ground regions. It analyses the shape of each planar segment that has been extracted in former step (3D roof segments). The rationale is that buildings should have a regular shape distinguished by long nearly linear segments and few corners (i.e. change of direction of the contour), while contours on damaged and vegetated regions are more frequently curved and irregular. The developed algorithm analyses points on the region contours and counts the number of pixels in correspondence of the curved areas of the boundaries (Figure 5(f)).

Geometric feature integration. Digital Surface Models (DSMs) generated with dense image matching techniques suffer from the data inherent characteristics. Surfaces with no texture (e.g. plain rooftops, shadows, façades, etc.) are the source of noise, while occlusions cause incomplete reconstructions. As geometric features are derived from the DSM, its unreliability is propagated to the extracted features. Incorporating the knowledge of where the DSM fails, offers the possibility for such unreliable areas to be specially treated, alternatively permits to eliminate them from further analyses and thus get rid of false positives.

The identification of these areas is realized by looking at the amplitudes of the texture and subsequently the morpho filters (see Figure 5). Small texture amplitudes correspond to texture-less surfaces where the matching fails: behind the small amplitudes there may hide a damage re-projected to the image as a black spot, or a non-damaged but texture-less building area. The feature correction shall be applied to the latter, otherwise the classification algorithm will see both areas as related. To discriminate the areas the morpho index and the
segmented region size are employed (see Figure 6), by analogy to the formerly described spectral feature. Pixels with high morpho indices (dark) and belonging to large segments are then assigned low nonplanarity and the boundary irregularity values, equivalent to non-damaged sites.

**Building candidate recovery**

This processing step extracts from the image and elevation data areas that correspond to buildings. Sampling of the building candidates takes place on the images of unary potentials, i.e. on the likelihoods of the pixels to take a given class. In the supervised learning, the authors adopt potentials that are a product of a discriminative classifier – the RFs (Breiman, 2001). The composition of the feature vector shall be agreed upon in conjunction with the set of classes, so that their discriminative character is maximized while effectively modelling the scene. Accordingly, we single out three classes: building, vegetation and ground. Following the strategy proposed by Niemeyer, Rottensteiner, and Soergel (2014), the unary potentials become

\[ \varphi_{i,RF}(x_i = l, z) = \exp(1 - w_i/N) \]  

For a particular sample \(i\), its attributed feature vector \(z\), a target class \(l\), and \(N\) decision trees, the energy is inversely proportional to the number of votes \(n_l\). The exponential expression ascertains that the function never takes the value zero.

If the supervised learning cannot be afforded, the unary potential are created on the basis of a single evidence – the mean building height. The discrimination occurs between the building and non-building class (two classes). This reduction in dimensionality can be the cause of a problematic representation of the building candidates (i.e. over-representation), especially in the presence of high vegetation. To overcome this issue, in the pre-processing phase, the vegetation mask provided by GRVI index is applied. Similar to Lafarge and Mallet (2012) and Gerke and Xiao (2014), the unary potentials are the normalized heights above the terrain, linearly projected within the range of \([0, 1]\):

\[ \varphi_{i,h}(x_i = l, h_i) = \min(1, \exp(1 - h_i/\sigma_h)) \]  

where \(h\) is the height derived from the normalized nDSM and the \(\sigma_h\) is the mean height building parameter set by the user. Once the unary potentials are retrieved, the definite building candidates are defined after a sequence of image processing routines that aim at removing small objects and closing holes in the foreground. These include bilateral filter, morphological opening and closing. Lastly, the candidates are separated by simple thresholding.

If only post-event data exist, and the buildings are entirely collapsed, they will be invisible to the algorithm outlined above. Partially collapsed buildings are taken into account by employing the low thresholding values.

**Building health ranking**

The ranking is formulated as a CRF complemented with high-order terms for image segments. The

![Figure 6. Geometric feature enhancement. Marked in bold is the decision path a pixel must take to be eligible for the correction.](image)
objective of the ranking is to tell apart intact from damaged buildings. The third class – other – is allowed to explain everything that is outside the building mask validity zone, as well as to cover its inaccuracies. The mask acts as a decision maker: in the false regions, in order to reduce the energy in the graph, it fixes the node’s unary potential to low and high values for the other and intact/damaged building class, respectively; for the true regions, the unary potentials are furnished by the k-means clustering (cf. the second pre-classification in Figure 1), as is described below.

**Contextual classification with graphical models**

A per-pixel classification precedes the regularization with Probabilistic Graphical Models defined over CRF, which adds the contextual information to the model. In the context of images, pixels are the nodes of the graph and have associated hidden variables, being e.g. classes. The optimal image classification is obtained by maximizing the a posteriori probability of the total assignment of classes to given observations (e.g. pixel intensities), written as the energy:

$$E(x) = \sum_{i \in V} \varphi_i(x_i, z) + \sum_{i,j \in \mathcal{E}} \psi_{ij}(x_i, x_j, z)$$

(3)

where $V$ is the number of image pixels, $\mathcal{E}$ is the set of edges connecting the pixels and $z$ is the observed data. The first terms $\varphi$ in Equation (3) is the unary potential and the second term $\psi$ is the pairwise potential that encourages the model to assign similar classes to neighbouring pixels (Boykov & Jolly, 2001; Kumar & Hebert, 2006; Lafferty, McCallum, & Pereira, 2001).

Pairwise formulation fails to represent complex scene structures. The result is often oversmoothed and does not follow the object’s sharp boundary. One way to overcome these artefacts is to incorporate higher order features into the model, and this is the approach adopted in this study. The high-level priors are embedded in the energy function as an additional term defined over a so-called multi-pixel clique:

$$E(x) = \sum_{i \in V} \varphi_i(x_i, z) + \sum_{i,j \in \mathcal{E}} \psi_{ij}(x_i, x_j, z)$$

$$+ \sum_{c \in S} \psi_c(x_c)$$

(4)

where $S$ is the number of regions and $\psi_c$ represents the higher order potentials.

**Unary and pairwise potentials**

The unary potentials are created for the building mask true regions, within the second pre-classification step (“Building candidates recovery” section). It is presumed that the mask is contaminated with a small percentage of error, so that the clustering of the local appearances encoded in the feature vectors is plausible.

The $k$-means clustering is an iterative algorithm that searches for clumps within the data (Bradski & Kaehler, 2008; Lloyd, 1982). In the proposed workflow, the individual features are normalized to a range of $[0, 1]$. Equation (5) expresses the unary potential: in the nominator – the L2 norm of the difference between a vector feature being the centroid of a cluster $l$ ($f_{i,l}$) and the current sample’s feature vector ($f_i$); in the denominator – normalization factor that is the largest distance of the current sample from all clusters.

The pairwise terms realize the classical Potts model (see Equation (6)). In our case, this term is of less importance in the model, as was pointed by Montoya-Zegarra et al. (2014), and it could be as well omitted as the context information provided by segments and encoded in the higher order term, already ascertains the smoothing effect.

$$\varphi_{i,k-m}(x_i = l, z) = \frac{f_{i,l} - f_i}{\max_{l \in L} f_{i,l} - f_i}$$

(5)

$$\psi_{ij}(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j \\ 1 & \text{otherwise} \end{cases}$$

(6)

where $x_i$ is a classification of pixel $i$ and the $z$ is the pixel’s local appearance.

**Higher order potentials**

In Kohli and Torr (2009), the pairwise energy equation is extended with higher order potentials specified on segments, and expressed for a list of classes energy as

$$\psi_c(x_c) = \min\{\min_{l \in L} (|c| - n_l(x_c))\theta_l + \gamma_l), \gamma_{\max}\}$$

(7)

$$\theta_l = \frac{\gamma_{\max} - \gamma_l}{Q}, \gamma_{\max} > \gamma_l$$

(8)

The maximal clique $c$ is a segment out of a collection of image segments $S$. There are $c$ pixels in a clique (i.e. a segment), $n_l(x_c)$ of which take the current class $l$, $\gamma_P$, $\gamma_{\max}$, $\theta_l$, $Q$ are the potential’s parameters, with the last being the truncation parameter. The so-called higher order potential function encourages that all pixels within a clique take the same class and assigns a cost to mixed configurations that is a function of the number of inconsistent pixels, truncated by the Q, thereby making the model robust. The lowest energy class is found with the st-mincut algorithm (Kohli & Torr, 2009; Kolmogorov & Zabih, 2004).

**Experiments**

The experiments evaluate the proposed methodology on two datasets: high-resolution (San Felice dataset, Table 3) and medium-resolution (Aquila dataset, Table 2) aerial imagery. In each case, three strategies are adopted, i.e. when building candidates (i) are furnished by the cadastre, (ii) originate from sampling the
class likelihood generated by RFs, or (iii) from the class likelihood generated by the sheer mean height parameter (H_MEAN). To give a sense of the sensibility to the threshold parameters isolating the building candidates, two distinct height values (σ₁ = 10 m and σ₂ = 15 m) are applied and assessed. In both cases, the buildings’ candidates are ousted from the likelihoods using a 20% threshold value. Areas below 20 m² were regarded small (nS). Table 1 outlines the processing strategies and the set of features adopted. In the RF approach, 300 trees are built and trained using 1000 samples. The following high-order potential parameter were selected: γ₁ = 1, γ_max₁ = 10, γ_max₂ = 1000 Q = 100. Two different γ_max values were tested to verify their influence on the outcome.

In order to evaluate the performance of the developed methodology, damaged areas are manually identified on the cadastral layer by visual inspection, thus producing a ground truth (GT) map for the datasets. The adopted assessment approach is a compromise between purely pixel-based and object-based strategies, i.e. both corresponding pixels and objects are considered as entities to be compared. First, building objects are automatically extracted from the GT map. Then, for each building, corresponding pixels are compared and three quality measures are derived, corresponding to the proportion of the damaged area of a building that is correctly classified (overlap), missed (underlap), or incorrectly detected as damaged (extralap). If no damaged area is observed in the GT nor in the classification map, a 100% overlap is assigned to the building. On the contrary, an intact building that includes damage-classified pixels in the classification map is defined as a 100% extralap. These metrics are inspired by the work presented in Shan and Lee (2005) and they are here adopted as quality figures to determine if a building and its damages should be considered as a True Positive (TP), False Positive (FP) or False Negative (FN). A building is accepted as a TP if featuring an overlap percentage greater than (or equal to) a given threshold (10%, 25% and 50%); otherwise, it is counted as FN. An extralap percentage exceeding 100% produces an FP building. Different thresholds are set up and, for each of them, the following metrics are finally computed:

\[
\text{Accuracy} = \frac{TP}{TP + FN + FP}; \\
\text{Precision} = \frac{TP}{TP + FP}; \\
\text{Completeness} = \frac{TP}{TP + FN}
\]

Given the final aim of the RAPIDMAP project, that is to address the issue of rapid mapping of post-disaster scenarios, small damaged areas (in most cases wrongly classified) are filtered out from the GT and classification maps before their evaluation. An area threshold of about 10 m² is considered as a reasonable value and adopted for all cases.

**Results**

The first dataset (San Felice, Italy) includes 21 images that were collected with a Vexcel Ultracam XP system after an earthquake in 2012. The images feature an average ground sample distance (GSD) of 5 cm and an overlap of about 80/50 (along-/across-track). The region comprises historical and residential buildings, together with warehouses and factories. Buildings are mainly low-storey and detached, surrounded by trees and bare ground.

The second dataset (L’Aquila, Italy) was acquired by a Pictometry imaging system, a five-lens camera system that incorporates vertical and slanted views. For the present work, only the nadir-looking images were used, consisting of 103 images with a mean GSD of 11 cm and a 70/40 overlap (along-/across-track). The flight surveyed the city of Aquila in 2009, after the massive seismic events that severely damaged most of the buildings. The area is characterized by a dense development of historic buildings having rather complex shapes, narrow roads and trees.

For both datasets, images were photogrammetrically processed using Apero/MicMac (Pierrot-Deseilligny & Clery, 2011, March; Pierrot-Deseilligny & Paparoditis, 2006). After the image block triangulation, 2.5D DSMs and true ortho-photos were extracted at full resolution. The developed methodology was then tested on a subset of the surveyed areas, corresponding to the most densely-built region (770 m by 560 m in San Felice; 616 m by 330 m in L’Aquila). Both subsets include different types of damages, ranging from completely collapsed buildings to partially damaged ones.

The datasets were processed on a PC with i7 (1st generation) 1.87 GHz, 8 GB RAM. The features (spectral and geometric) calculation times for a tile of 8MPixel took approximately an hour, while the classification part took a few minutes. The algorithms were implemented in C++, without parallelization.

---

**Table 1.** The three developed strategies and a set of features adopted in the pre-classification stage (aiming at building candidates recovery) and the final classification (aiming at detecting the damages).

| Strategies     | Pre-classification (candidates recovery) | Regularized classification (health rank) |
|----------------|------------------------------------------|------------------------------------------|
| Cadastre       | H₁₀, H₁₅                                   | nDSM                                     |
| RF             | nDSM, R                                    | Spectral (morpho, RP, TI, R*, B*)         |
|                |                                          | geometric (NP, CI)                       |

*Used only in the low-resolution dataset (Aquila).*
Using parallel computing and GPU, the estimated computation time could drop to a few minutes which is compatible with the end-user needs. It shall be underlined that the set of algorithms was conceived in a proof of concept framework.

**High-resolution imagery – San Felice dataset**

Some of the classification results are shown in Figure 7, whereas Table 2 presents the numerical results of the evaluation. Various combinations of features were adopted throughout the tests, and the reported figures correspond to the feature sets discussed in the paper. The appearance of rooftops in the aerial images is detailed and very clear due to the high-resolution data. For that reason, the red and blue band layers were not taken into account as independent features (see Table 1).

The algorithm performs best when building candidates are furnished by the cadastre. The cadastral mask was very precise as generated from accurate digital vector maps. Accuracy/precision/completeness place between 80 and 90%, and the ratio of detected to all damages is equivalent to ~87%. It should not be ignored that cadastre was also used as the base for creating the GT; thus, this scenario is additionally privileged. In all scenarios but the RF, there is a clear trend towards overrepresentation of the damaged areas, rather than underrepresentation. In other words, the results are complete as they do not predict false negatives; nonetheless, they are less precise, i.e. there are a certain number of false positives. The RF gives vice versa results – good precision measures, and the completeness that is by 10% points below. It indicates that our training did not represent

![Image](image_url)

**Figure 7.** San Felice dataset: orthophoto of a part of the investigated area (a), manually produced GT (b) and automatically produced classification maps (c–f). Three classes are visible: damaged buildings (red), intact buildings (blue/white) and other (grey/black).

**Table 2.** San Felice dataset. Accuracy, precision and completeness values computed for different overlap thresholds (10%, 25% and 50%).

| Method           | Accuracy (%) | Precision (%) | Completeness (%) |
|------------------|--------------|---------------|------------------|
| Cadastre         | 87.78        | 83.33         | 81.11            | 92.94 | 92.59 | 92.41 | 94.05 | 89.29 | 86.90 | 86.67 |
| $H_{\text{MEAN}}_{Y_{\text{RF}}}$ | 77.78 | 74.44 | 73.33 | 83.33 | 82.72 | 82.50 | 92.11 | 88.16 | 86.84 | 94.65 |
| $H_{\text{MEAN}}_{Y_{\text{RF}}}$ | 73.33 | 68.89 | 67.78 | 77.65 | 76.54 | 76.25 | 92.96 | 87.32 | 85.92 | 80.00 |
| $H_{\text{MEAN}}_{Y_{\text{RF}}}$ | 72.22 | 68.89 | 67.78 | 78.31 | 77.50 | 77.22 | 90.28 | 86.11 | 84.72 | 73.33 |
| $H_{\text{MEAN}}_{Y_{\text{RF}}}$ | 72.22 | 68.89 | 68.89 | 78.31 | 77.50 | 77.50 | 90.28 | 86.11 | 86.11 | 73.33 |
| RF               | 83.33 | 81.11 | 80.00 | 93.75 | 93.59 | 93.51 | 88.24 | 85.88 | 84.71 | 60.00 |

*Note. The last column lists the number of detected damages against the total number of damages.*
holistically all the possible damages. The average building height in the dataset is closer to 10 m than 15 m because the first assumption \( H_{\text{MEAN},10} \) reports all four quality measures at higher levels. The influence of higher order term proves significant. For both \( H_{\text{MEAN},10} \) and \( H_{\text{MEAN},15} \), giving more importance to the higher order term \( (\gamma_{\text{max},2}) \) improves the accuracy/precision/completeness. The detected regions are usually better delineated as the segmentation gains its importance. On the other hand, small and just partially detected often disappear: this usually reflects on a lower percentage of detected damages.

**Medium-resolution imagery – L’Aquila dataset**

Figure 8 and Table 3 present the classification results. The lower resolution images render the geometric features less significant, as opposed to the San Felice images. The spectral features, however, turned much more homogenous and showing less detail, hence more relevant. Consequently, the red and blue bands were included in the classification (see Table 1).

The inferior quality of DSM is caused by both poor image overlap and larger GSD. In scenarios \( H_{\text{MEAN},10} \) and \( H_{\text{MEAN},15} \) the two contribute to poor building delineations as narrow streets and courtyards were often merged, while non-planar structures smoothed. It also explains the presence of false negatives (i.e. worse completeness).

The cadastral layer was obtained from digital vector map that had been digitized from a small-scale analogue map. The accuracy potential and the actuality were limited, as shown in Figure 8(b, f) where the courtyards is represented as a building part, and the building borders are outdated. The RF lends better precision than the completeness scores which indicates again poorly modelled existing damage types and the incapability of the method to extrapolate its actions.

**Conclusions**

In this paper, we present a methodology for the extraction of building damages from a set of airborne nadir images. The developed methodology has been designed within the RAPIDMAP project framework. Therefore, the human interaction and time to process the data must be limited in order to provide as soon as possible the needed information for the rescue work planning.

The methodology relies on two different approaches: the semi-supervised and unsupervised approach. The selected approach depends on the availability of prior information (i.e. cadastral maps) to provide the building layer. A reduced number of

---

**Table 3.** L’Aquila dataset. Accuracy, precision and completeness values computed for different overlap thresholds (10%, 25% and 50%).

|                     | Accuracy (%) | Precision (%) | Completeness (%) |
|---------------------|--------------|---------------|------------------|
|                     | 10% 25% 50%  | 10% 25% 50%   | 10% 25% 50%      |
| **Detected damages/total damages (%)**\( \gamma_{\text{max},1} \) | Cadastre: 61.86 59.32 50.00 | 71.57 70.71 67.05 | 82.02 78.65 66.29 | 73.91 |
|                     | \( H_{\text{MEAN},10} \) \( \gamma_{\text{max},1} \) | 63.56 55.93 48.31 | 79.79 77.65 75.00 | 75.76 66.67 57.58 | 71.74 |
|                     | \( H_{\text{MEAN},15} \) \( \gamma_{\text{max},1} \) | 63.56 55.08 46.61 | 78.95 76.47 73.33 | 76.53 66.33 56.12 | 78.26 |
|                     | RF           | 64.41 60.17 53.39 | 86.36 85.54 84.00 | 71.70 66.98 59.43 | 50.00 |

The last column lists the number of detected damages against the total number of damages.
specifically designed features are then used as input in a higher order CRF algorithm. Some parameters can be tuned in the whole workflow in order to reduce the dependency of results from the initial set up. The algorithm has been tested on two different datasets in order to assess its reliability in different urban context and with different data resolutions. The high-resolution and high image overlap of the first dataset represent the standard of nowadays image acquisitions. On the other hand, the second test represents a sub-optimal data due to the poor image resolution and reduced image overlap. In both cases, the results have been encouraging, showing quite good performances in the detection of all the damage typologies. As expected, the results are strongly conditioned by the image resolution and the quality of the DSM and the orthophoto: results are satisfying only when features are meaningful and not affected by poor input data quality. The size of the damage together with the resolution of the data plays an important role in the detection process. Delineation of the damages is often imprecise. It is particularly evident in the second dataset, where most of the damages are only partially detected as stated by the variability of completeness according to the used threshold. However, the underestimation of damages often does not affect the practical usability of the results (i.e. percentage of detected damages), as rescuers often need a first map of the damages location without caring of their actual extension. Field inspections or simply higher resolution data (Vetrivel et al., 2015) could provide better results. The use of higher $γ_{\text{max}}$ improves the accuracy and the precision of the results, showing that the classification process greatly benefits from the higher order terms (i.e. object analysis). Damages are always better delineated. It is reflected in the increased classification accuracy and precision, and the completeness being rather steady. The value of $γ_{\text{max}}$ should be carefully selected because it may lead to the deletion of little and incomplete detected damages. The higher order by default promotes a region to have (more likely) the same label that translates to very small and fragmented regions being deleted. This in turn reduces the completeness and the overall percentage of correctly detected damages. The availability of a prior knowledge such as cadastral maps can greatly help the damage detection task. These maps help to remove problems due to incorrect 3D reconstructions (i.e. using only height values) or due to incorrect pre-classifications (i.e. removal of green roofs, classified as vegetation. In these cases, problems on the borders of the buildings or in the courtyards can arise. The use of RF or height information can provide generally good results in the detection of candidate building regions, yet they remain susceptible to wrongful 3D reconstructions (DSM quality). Moreover, the RF building delineation is also dependent on the training – incomplete results caused by the incomplete or inappropriate learning phase.

The algorithm computation time has not been assessed in this paper as most of the code is still not optimized. A more efficient release of the code will be delivered in the future.

Acknowledgments

The authors thank Mauro Dalla Mura (INP–Grenoble) for providing the morphological filter code and Andrea Baraldi (University of Maryland) for his fruitful suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by RAPIDMAP (http://rapidmap.fbk.eu), a CONCERT-Japan project, i.e. a European Union (EU) funded project in the International Cooperation Activities under the Capacities Programme the 7th Framework Programme for Research and Technology Development.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274–2282.

Adams, S.M., & Friedland, C.J., 2011. A survey of unmanned aerial vehicle (UAV) usage for imagery collection in disaster research and management. 9th International Workshop on Remote Sensing for Disaster Response, Stanford University, California.

Altan, O., Backhaus, R., Boccardo, P., Giulio Tonolo, F., Trinder, J., Van Manen, N., & Zlatanova, S., 2013. The Value of Geoinformation for Disaster and Risk Management (VALID) (Joint Board of Geospatial Information Societies Report).

Baltisavias, E., Cho, K., Remondino, F., Soergel, U., & Wakabayashi, H. (2013). Rapidmap – rapid mapping and information dissemination for disasters using remote sensing and geoinformation. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40, 31–35. 7/W2

Beumier, C., & Idrissa, M. (2016). Digital terrain models derived from digital surface model uniform regions in urban areas. International Journal Remote Sens., 1–17. doi:10.1080/01431161.2016.1182666

Boccardo, P., & Tonolo, F. (2014). Remote sensing role in emergency mapping for disaster response. Engineering Geology for Society and Territory, 5(1), 17–24.

Boykov, Y., & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(9), 1124–1137.
Poli, D., & Caravaggi, I. (2013). 3D information extraction from stereo VHR imagery on large urban areas: Lessons learned. *Natural Hazards, 68*(1), 53–78.

Rezaeian, M., & Gruen, A. (2011). Automatic 3D building extraction from aerial and space images for earthquake risk management. *Georisk, 5*(1), 77–96.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2013). GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal, 78*(1), 21–40.

Rosu, A.M., Pierrot-Deseilligny, M., Delorme, A., Binet, R., & Klinger, Y. (2015). Measurement of ground displacement from optical satellite image correlation using the free open-source software MicMac. *ISPRS Journal of Photogrammetry and Remote Sensing, 100*, 48–59.

Samadzadegan, F., & Rastiveisi, H. (2008). Automatic detection and classification of damaged buildings, using high resolution satellite imagery and vector data. *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37*, 415–420.

Shackelford, A. K., & Davis, C. H. (2003). A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Transactions on Geoscience and Remote Sensing, 41*(10), 2354–2365.

Shan, J., & Lee, S.D. (2005). Quality of building extraction from IKONOS imagery. *Journal of Surveying Engineering, 131*(1), 27–32.

Shanley, L., Burns, R., Bastian, Z., & Robson, E. (2013). Tweeting up a storm: The promise and perils of crisis mapping. *Photogrammetric Engineering and Remote Sensing, 79*(10), 865–879.

Tokarczyk, P., Wegner, J.D., Walk, S., & Schindler, K. (2013). Beyond hand-crafted features in remote sensing. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 1*(1), 35–40.

Vetrivel, A., Gerke, M., Kerle, N., & Vosselman, G. (2015). Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 61–78.

Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., & Mehl, H. (2007). Satellite image analysis for disaster and crises-management support. *IEEE Transaction On Geoscience and Remote Sensing, 45*(6), 1520–1528.

Vosselman, G. (2012). Automated planimetric quality control in high accuracy airborne laser scanning survey. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*, 90–100.

Yalniz, I.Z., & Aksoy, S. (2010). Unsupervised detection and localization of structural textures using projection profiles. *Pattern Recognition, 43*(10), 3324–3337.

Zhou, Y., Parsons, B., Elliott, J.R., Barisin, I., & Walker, R.T. (2015). Assessing the ability of pleiades stereo imagery to determine height changes in earthquakes: A case study for the El Mayor-Cucapah epicentral area. *Journal of Geophysical Research: Solid Earth, 120*(12), 8793–8808.