AUTOMATIC INFORMATION EXTRACTION FROM REMOTE SENSING IMAGES AND 3D POINT CLOUDS FOR BUILDING DAMAGE ASSESSMENT

Anand Vetrivel

Graduation committee:

Chairman/Secretary

Prof.dr.ir. A. Veldkamp	University of Twente, ITC

Supervisors

Prof.dr.ir. M.G. Vosselman	University of Twente, ITC
Prof.dr. N. Kerle	University of Twente, ITC

Co-supervisor

Prof.dr.-ing. M. Gerke	Technische Universität Braunschweig

Members

Prof.dr. V.G. Jetten	University of Twente, ITC
Prof.dr.ir. A. Stein	University of Twente, ITC
Prof.dr. N. Paparoditis	Université Paris-Est
Prof.dr. S.M. de Jong	Universiteit Utrecht

ITC dissertation number 316
ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN: 978-90-365-4507-5
DOI: 10.3990/1.9789036545075
Printed by: ITC Printing Department, Enschede, The Netherlands

UNIVERSITY OF TWENTE.
ITC
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

AUTOMATIC INFORMATION EXTRACTION FROM REMOTE SENSING IMAGES AND 3D POINT CLOUDS FOR BUILDING DAMAGE ASSESSMENT

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr. T.T.M. Palstra,
on account of the decision of the graduation committee,
to be publicly defended
on Friday, March 9, 2018 at 12.45 hrs

by

Anand Vetrivel
born on September 18, 1986
in Chennai, India

This dissertation is approved by:

Prof.dr.ir. M.G. Vosselman (supervisor)
Prof.dr. N. Kerle (supervisor)
Prof.dr.-ing. M. Gerke (co-supervisor)

To my father Vetrivel and my friend Judge Raja

தாமின் புறுவது உலகின் புறக்கண்டு
காமுறுவர் கற்றறிந் தார் *(Thirukkural 399)*

The educated will desire to lean more when they see that the world delights in learning as much as they delight in it.

Summary

Rapid automated information generation about the damages to the buildings after a destructive disaster event such as an earthquake is crucial to carry out speedy response and recovery actions. Remote sensing is the most suitable technology to provide data for automatic extraction of damage information for such spatially extensive events. In particular oblique airborne images from manned and unmanned aerial platforms have been recognized as a potentially more useful data source for building damage assessment than conventional vertical images, due to the following specific reasons: these images are generally captured with (i) multiple camera views (coverage of top and sides of the buildings) which is crucial for holistic building damage assessment; (ii) high spatial resolution (rich radiometric features); (iii) high frame overlap making it suitable to generate 3D point clouds (rich geometric features). Data with these characteristics are an important prerequisite for the automatic extraction of fundamental information for building damage assessment as described below. Although oblique airborne images and derived 3D point clouds are desirable for damage assessment, reliable, robust and operational methods for automated extraction of damage information from such data are rare. Thus the objective of this research was to design and develop methods for automatic extraction of information needed for damage assessment, specifically from the oblique airborne images and the 3D point clouds derived from them. Pertaining to this, several methods were developed as summarized below:

1) Automated building detection: Identification of individual buildings is the first step in building damage assessment process. Photogrammetric point clouds have been considered for building detection process, but they are often noisy and error prone. Moreover, damaged scenes appear cluttered, which makes object recognition complex. Hence, a building detection method particularly suitable for this kind of data and scenario was developed and tested. It achieved an overall accuracy of 96%.

2) Automated damage detection using supervised approach: Debris, rubble piles and spalling are strong evidences to identify heavily damaged structures. The texture features have often been reported to be superior for the identification of these damage evidences from images. Based on preliminary analysis it was anticipated that texture features such as Gabor

and Histogram of Oriented Gradients (HoG) would be more useful compared to other widely reported features for damaged region detection. Supervised classification models based on Support Vector Machine (SVM) and Random Forests (RF) were trained using the above features independently. The classifier based on Gabor features with RF performed best, identifying 95% of the damaged regions. However, the developed method suffered from a generalization problem, where the accuracy dropped by around 30% when tested on another independent data set. To address this generalization issue, a method using visual bag-of-words (BoW) was developed using the above features–HoG and Gabor. The developed method was tested using four different data sets that varied substantially in terms of data and scene characteristics. The overall accuracy improved by 14% (i.e., from 77% to 91%) when applying the BoW approach with the Gabor features on the most complex dataset, which was used to test the generalization capability. It was observed that these texture features fail in specific urban settings with complex radiometric characteristics e.g., L'Aquila city in Italy. In the past few years, deep learning features have been reported as being superior to conventional handcrafted features for many applications in remote sensing and related domains. Hence, a method based on deep learning features was developed for damage detection, and achieved an accuracy around 90% for the areas where the above mentioned conventional textures features failed. Also, patch level 3D point cloud features were proposed and used in addition to image-based deep learning features and achieved an accuracy improvement of 3% to 7% under different settings. In contrast to the above mentioned supervised methods based on the batch learning setting, an incremental (online) learning based method using deep learning features was developed for damage detection. This was attempted to demonstrate how the streaming information about the damaged locations (possibly available at different point in time by various sources during disaster event) can be used to dynamically frame the training samples to incrementally build a reliable and robust classifier for damage detection. The obtained results show that the developed incremental classifier performed on par with the classifiers based on batch learning approach when deep learning features were used.

3) Damaged region identification using unsupervised approaches: The aforementioned supervised approaches work based on aprioristic assumptions of damaged building's shapes and textures, which sometimes lead to uncertainties and misdetections. The usage of pre-event data as reference could be of help to resolve these disputes. Towards this, a methodology based on an unsupervised approach for comparing the 3D point clouds and airborne oblique images of pre- and post-event was developed for damage detection. The developed methodology detected 87% of damaged elements based on element-wise comparison in pre- and post-event. The missed detections were mainly due to varying noise levels within the point cloud, which hindered the recognition of some structural elements. Also, the method for identifying openings in the building created due to damage using post-event image and point cloud was developed based on an unsupervised approach. In this method, the gaps in the 3D point cloud were detected in an unsupervised manner. Subsequently, the gaps that result from damages were classified based on a set of rules. The developed approach detected all gaps due to damage in the considered study area.

4) Accurate roof segment delineation for 3D reconstruction: For the holistic and reliable damage assessment, 3D modelling of buildings was realized as the desirable product, particularly for assessing the building with intact roofs and damaged façades. Automated and accurate delineation of roof faces of buildings is a minimal requirement for automatic 3D reconstruction of buildings using 3D point clouds. Image segmentation methods, incorporating 3D features in images (at pixel and super-pixel levels), were developed for delineating independent roof faces of the building. Based on these roof faces the 3D reconstruction was done using the existing approach. The quality of the 3D models depends on the accuracy of the roof face delineation. The 3D model of the buildings obtained for the chosen study area was visually close enough to the shape of the original buildings, thus depicting the accuracy of the roof face delineation method.

The methods developed through this research were integrated to build an automated damage assessment system which was thoroughly demonstrated using the data obtained through EU-FP7 project RECONASS (www.reconass.eu, a pilot project in the field of near real-time damage assessment). This demonstration yielded promising results, thereby highlighting the potentials of the developed system to scale well to applications in the real-world setting with minimal enhancements.

Samenvatting

Snelle geautomatiseerde informatieverstrekking over de schade aan de gebouwen na een verwoestende ramp zoals een aardbeving is cruciaal om snelle reddings- en herstelacties uit te voeren. Remote sensing is de geschiktste technologie om gegevens te verschaffen voor automatische extractie van schade-informatie voor dergelijke ruimtelijk omvangrijke gebeurtenissen. Met name oblieke luchtbeelden van bemande en onbemande vliegtuigen zijn erkend als een potentieel nuttiger gegevensbron voor het beoordelen van schade aan gebouwen dan conventionele verticale opnamen, vanwege de volgende specifieke redenen: deze beelden worden meestal vastgelegd met (i) meerdere camera's (opname van de bovenkant en zijkanten van gebouwen) die cruciaal zijn voor een holistische beoordeling van schade aan gebouwen; (ii) hoge ruimtelijke resolutie (duidelijke radiometrische kenmerken); (iii) de hoge mate van overlap maakt de beelden geschikt om 3D-puntenwolken te genereren (duidelijke geometrische kenmerken). Beelden met deze kenmerken zijn een belangrijke voorwaarde voor de automatische extractie van fundamentele informatie voor schadebeoordeling van gebouwen, zoals hieronder beschreven. Hoewel oblieke luchtfoto's en afgeleide 3D-puntenwolken wenselijk zijn voor het beoordelen van schade, zijn betrouwbare, robuuste en operationele methoden voor geautomatiseerde extractie van schade-informatie uit dergelijke beelden zeldzaam. Het doel van dit onderzoek was om methoden te ontwerpen en te ontwikkelen voor automatische extractie van informatie die nodig is voor schadebepaling, met name uit oblieke luchtfoto's en de daaruit afgeleide 3D-puntenwolken. Hiervoor zijn verschillende methoden ontwikkeld zoals hieronder samengevat:

1) Geautomatiseerde gebouwdetectie: Identificatie van individuele gebouwen is de eerste stap in het proces van schadebeoordeling. Overwogen is om fotogrammetrische puntwolken te gebruiken voor het detectieproces van gebouwen, maar deze bevatten vaak te veel ruis en grove fouten. Bovendien lijken beschadigde scènes rommelig, waardoor de objectherkenning complex wordt. Daarom is een gebouwdetectiemethode ontwikkeld en getest, die bijzonder geschikt is voor dit soort gegevens en scenario's. Hiermee is een algehele nauwkeurigheid van 96% bereikt.

2) Geautomatiseerde schadeherkenning met behulp van een gecontroleerde aanpak: puin, puinhopen en afbrokkelende gevels zijn sterke bewijzen om zwaar beschadigde constructies te identificeren. Van textuurkenmerken wordt vaak gezegd dat ze superieur zijn voor de identificatie van dergelijke schade in afbeeldingen. Op basis van een voorlopige analyse werd verwacht dat textuurkenmerken zoals Gabor en Histogram of Oriented Gradients (HoG) nuttiger zouden zijn in vergelijking met andere veel gerapporteerde functies voor detectie van beschadigingen. Gecontroleerde classificatoren gebaseerd op Support Vector Machine (SVM) en Random Forests (RF) zijn onafhankelijk van elkaar met behulp van de bovengenoemde functies getraind. De classificator op basis van Gaborkenmerken met RF presteerde het beste en identificeerde 95% van de beschadigde gebieden. De ontwikkelde methode had echter te lijden onder een generalisatieprobleem, waarbij de nauwkeurigheid met ongeveer 30% daalde bij testen op andere onafhankelijke beelden. Om dit generaliseringsprobleem aan te pakken, werd een methode ontwikkeld met behulp van visuele Bag-of-Words (BoW) op basis van de bovenstaande HoG en Gaborkenmerken. De ontwikkelde methode is getest met behulp van vier verschillende datasets die aanzienlijk varieerden in de eigenschappen van zowel de beelden als de scènes. De algehele nauwkeurigheid verbeterde met 14% (d.w.z. van 77% tot 91%) bij het toepassen van de BoW-methode met de Gabor-kenmerken op de meest complexe gegevensreeks, die werd gebruikt om het vermogen tot generaliseren te testen. Vastgesteld werd dat deze textuurkenmerken falen in specifieke stedelijke omgevingen met complexe radiometrische kenmerken, bijvoorbeeld in de stad L'Aquila in Italië. In de afgelopen paar jaar zijn deep learning-kenmerken gepubliceerd, die voor veel toepassingen in de aardobservatie en aanverwante domeinen superieur zijn aan de conventionele handgemaakte kenmerken. Daarom werd een methode voor schadeherkenning op basis van deep learning-kenmerken ontwikkeld. Hiermee werd een nauwkeurigheid van ongeveer 90% bereikt voor de gebieden waar de bovengenoemde conventionele textuurkenmerken faalden. Ook zijn 3D-puntenwolkkenmerken voorgesteld en gebruikt naast de op beelden gebaseerde kenmerken voor deep learning. Hiermee werd een nauwkeurigheidsverbetering van 3% tot 7% onder verschillende instellingen bereikt. In tegenstelling tot de hierboven beschreven gecontroleerde methoden op basis van een batch-leeromgeving, werd een incrementele (online) op leren gebaseerde methode ontwikkeld met behulp van deep learning-kenmerken voor schadeherkenning. Hiermee is geprobeerd aan te tonen hoe de informatie over de beschadigde locaties (mogelijk beschikbaar op verschillende tijdstippen uit verschillende bronnen tijdens een rampgebeurtenis) kan worden gebruikt om de trainingsdata dynamisch aan te leveren om zo incrementeel een betrouwbare en robuuste classificator voor schadeherkenning te ontwikkelen. De verkregen resultaten laten zien dat de ontwikkelde incrementele classificator op hetzelfde niveau presteerde als de classificatoren op basis van de batch-leerbenadering wanneer deep learning-kenmerken werden gebruikt.

3) Identificatie van beschadigde gebieden met behulp van een ongecontroleerde aanpak: de bovengenoemde gecontroleerde aanpak werkt op basis van a priori veronderstellingen over beschadigde gebouwvormen en -structuren, die soms leiden tot onvolledigheid en onjuistheid. Het gebruik van gegevens van voor een aardbeving als referentie zou kunnen helpen om deze problemen op te lossen. Hiertoe werd een methode ontwikkeld op basis van een ongecontroleerde aanpak voor het vergelijken van de 3D-puntwolken en oblieke luchtfoto's van voor en na de gebeurtenis. De ontwikkelde methodologie detecteerde 87% van de beschadigde elementen op basis van deze vergelijking. De onvolledigheid was voornamelijk te wijten aan het variëren van de ruis in de puntenwolk, wat de herkenning van sommige gebouwelementen bemoeilijkte. Ook is een methode ontwikkeld voor het identificeren van gaten in gebouwen die zijn ontstaan als gevolg van schade. In deze methode werden de gaten in de 3D-puntenwolk met een ongecontroleerde aanpak gedetecteerd. Vervolgens werden de gaten die het gevolg zijn van schades geclassificeerd op basis van een reeks regels. De ontwikkelde aanpak detecteerde alle gaten als gevolg van schade in het onderzochte studiegebied.

4) Nauwkeurige omlijning van dakvlakken voor 3D-reconstructie: voor de holistische en betrouwbare schadebeoordeling werd een methode voor 3D-modellering van gebouwen gerealiseerd, vooral voor de beoordeling van gebouwen met intacte daken en beschadigde gevels. Geautomatiseerde en nauwkeurige omlijning van dakvlakken van gebouwen in foto's is een minimale vereiste voor de automatische 3D-reconstructie van gebouwen uit 3D-puntenwolken. Voor het omlijnen van de verschillende dakvlakken van een gebouw zijn beeldsegmentatiemethoden ontwikkeld, die 3D-kenmerken in rasterbeelden gebruiken (op pixel- en superpixelniveau). Op basis van deze dakvlakken werd de 3D-reconstructie uitgevoerd met behulp van een al bestaande aanpak. De kwaliteit van de 3D-modellen is afhankelijk van de nauwkeurigheid van de omlijning van dakvlakken. De 3D-modellen van de gebouwen, die voor het gekozen studiegebied werden verkregen, waren visueel voldoende dicht bij de vorm van de werkelijke gebouwen en gaven daarmee de nauwkeurigheid weer van de omlijningsmethode van dakvlakken.

De methoden, die in dit onderzoek zijn ontwikkeld, zijn geïntegreerd om een geautomatiseerd schadebeoordelingssysteem op te bouwen. Dit systeem is grondig gedemonstreerd met behulp van de gegevens die zijn verkregen uit het EU-FP7-project RECONASS (www.reconass.eu, een pilotproject op het gebied van bijna-realtime schadebeoordeling). Deze demonstratie leverde veelbelovende resultaten op en benadrukte zo het potentieel van het ontwikkelde systeem om zich met minimale noodzaak tot afstelling goed aan te passen aan toepassingen in de praktijk.

Acknowledgement

This dissertation would not have been possible without the support of many people who in one way or another have contributed and extended their precious knowledge and experience in my PhD studies. It is my pleasure to express my gratitude to them.

My PhD life was fully peaceful and pleasant if few things were excluded. I would say this was only because of my two supervisors Prof. Markus Gerke and Prof. Norman Kerle. Before joining PhD, I was a Software Engineer with some background in Image Processing. When I joined my PhD in 2013, I was not proficient in Photogrammetry and had no idea about what 3D point cloud was. I had learnt all these technical stuff effortlessly because of Markus. He taught me any concept like teaching to a kid. I feel blessed and lucky to have Markus as my supervisor. Norman is a great motivator and critic. He always notices and passes comments on even minor things as he expects his PhD students to be qualified research professionals in terms of both technical and interpersonal skills. Thus, his constructive criticism encouraged my critical thinking and technical writing which helped me a lot to improve myself. Both of my supervisors spent enormous time for my PhD contributing through technical discussion, paper and thesis revision. As I got lot of support from my supervisors, I had an opportunity to explore several other interesting things during my PhD life such as learning music, exploring my skills on writing poems and to enrich my knowledge with history, political and social issues of my country. I'm always thankful to Markus and Norman for making my PhD life easier and joyful.

I'm grateful to my promotor Prof. George Vosselman for his continuous support and encouragement throughout my PhD. Certainly, it is a great pride for me to obtain this PhD degree under a legend like Prof. George Vosselman in a reputed institution like ITC. So, I would like to express my sincere gratitude to ITC and the European Commission (funding agency of RECONASS project) for providing me this PhD opportunity.

I would like to thank our department secretary Ms. Teresa for her great support on both official and personal things that made my life easier in The Netherlands.

There is one person who has always been a pillar of support throughout my thesis and personal life. That's my friend Annamalai Narayanan and I feel that he is my perfect research partner as our background, interests and thought process are almost similar. I would like to thank him for the countless intensive brainstorming sessions that always helped me to improve my work.

My heartfelt thanks to my undergraduate professors Dr. Alaguraja, Dr. Sathyabama and Dr. Kavitha Devi for their special efforts to make me succeed in every step of my career.

My deepest gratitude to my sister-in-law Geetha who stood behind me in hard times especially when I resigned my job before joining my Master's course.

Special mention goes to my fellow PhD students and other faculties in our department for being supportive and providing a friendly working environment.

Last but not the least, I thank my wife for supporting me spiritually throughout this thesis work and my life in general.

Table of Contents

List of figures

List of tables

1 Introduction

1.1 Background

Earthquakes are one of the deadly and destructive disaster events which is unpredictable and increasingly happening at least once every year in several parts of the world. Earthquakes have the potential to cause severe damages to the manmade structures particularly buildings, thereby causing spatially extensive rapid human and economic loss. The earthquake disaster statistics (cf. Figure 1-1 and Figure 1-2) emphasizes the requirement and magnitude of response and recovery actions to be carried out quickly as part of disaster management after any such event. The response processes are more related to the actions of search and rescuing victims, providing first aid and other temporary relief actions. Whereas the recovery processes refer to the actions taken to restore the affected area and people's life to its original state. For initiating the effective response and recovery actions the damage assessment which provides complete picture about the severity, extent and location of the damaged areas are crucial. For example, the approximate estimation of number of causalities is very important to plan for the rescue forces to be involved in the response actions and this information is closely related to the number of fully and partially collapsed buildings. Therefore, mapping the precise locations of highly damaged buildings is important as they are the suspected locations for victims, required to direct the rescue teams. This information also provides a rough estimate of number of people rendered homeless which can further aid to plan and allocate resources (funds, food, shelter, medical, etc.,). Therefore, rapid assessment of damage at building level is an important requisite and provides crucial information for stakeholders involved in response and recovery actions. However, the required level of damage information varies for each stakeholder. For example, those involved in the search and rescue process require only the information about the collapsed/heavily damaged buildings very rapidly. Other actors particularly related to recovery phases, such as insurance companies, require a less time critical but very detailed damage assessment, down to the level of cracks on a building. Conventional manual field assessment can provide very detailed and accurate damage assessment. However, it is cost, time and labor intensive, and cannot access all parts of the building. For example, the roof tops and even upper sections of facades of very tall buildings are difficult to assess based on field inspection. In practice, remote sensing images are increasingly being used as an alternative to field survey, and they constitute the best data source for initial and large-area damage assessment (Dell'Acqua and Gamba, 2012; Dong and Shan, 2013). Moreover, remote sensing has a long history in damage assessment, starting in early twentieth century by successfully capturing the extensive view of 1906 earthquake damage in San Francisco through a kite-borne camera (cf. Figure 1-3). Since then technology has evolved gradually from aerial (sensors mounted on the balloon, aircraft or helicopter) to space borne satellites and now it has become the prospective source for spatial information in the field of disaster management (Dong and Shan, 2013).

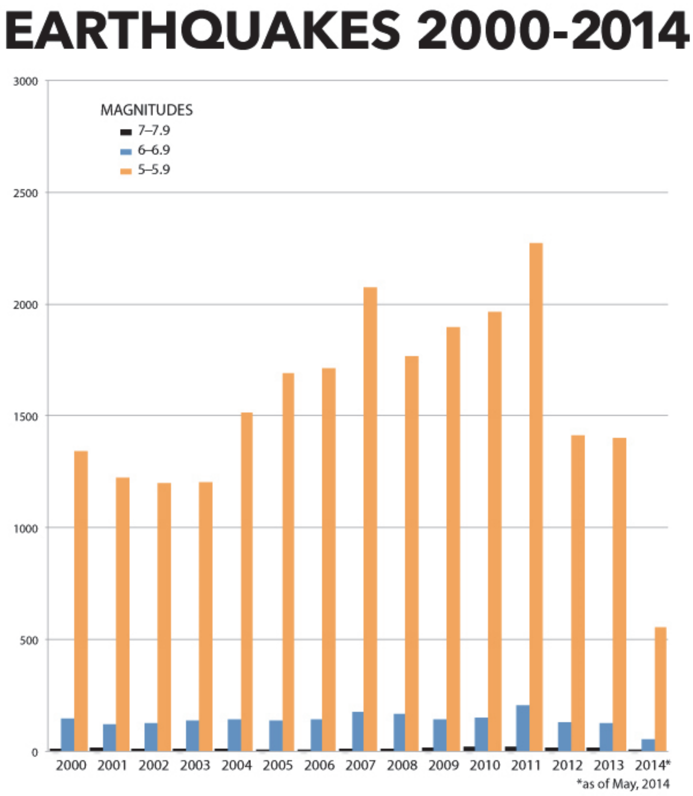


Figure 1-1 Trends in occurrence of earthquakes all around the world

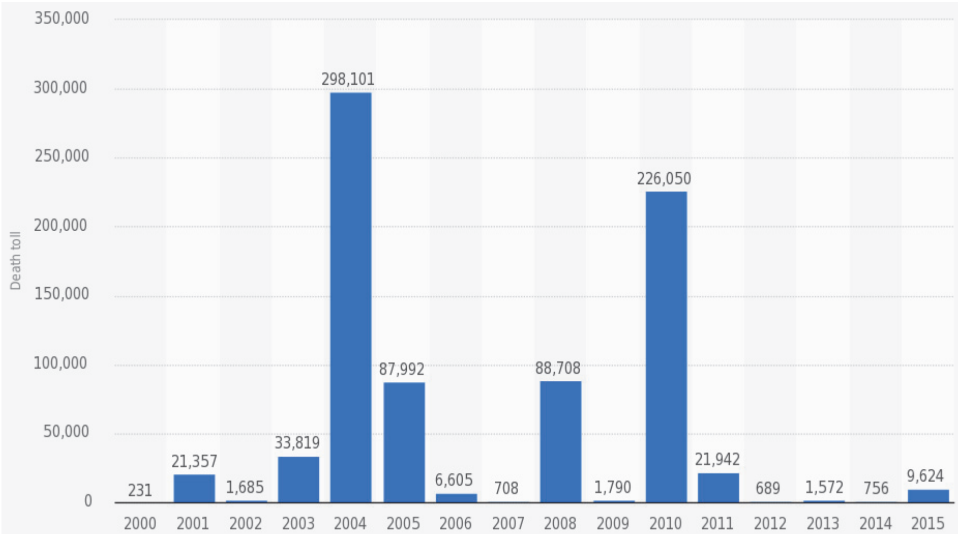


Figure 1-2 Global death toll due to earthquakes from 2000 to 2015

Figure 1-3 George Laurence's photo of the San Francisco after the earthquake in 1906

1.2 Satellite and aerial images for damage assessment

Satellite and aerial images are nowadays the conventional data source for assessing the extent and degree of damage after an event (Voigt et al., 2007). The advent of very high spatial resolution (VHR) remote sensing data like satellite images of IKONOS, Quick Bird, EROS-B, World View, Geo Eye, etc., having spatial resolutions <1m made qualitative and quantitative damage assessment on per building level feasible. For example, see Figure 1-4 image of Geo Eye-1 satellite with a spatial resolution of 0.5 m where the individual buildings can be distinguished.

Many mapping agencies such as Charter[1] are largely using this kind of high resolution satellite images for producing damage maps based on manual annotations to aid for emergency response actions. Alternative to manual mapping, from past decade, increasingly studies are being attempted to develop automated methods for mapping damage. Particularly, the temporal characteristics of satellite remote sensing provided the image of the scene before and after the event, motivated researchers to move forward from the qualitative to automated quantitative damage assessment. The damaged regions are identified based on the changes in the radiometric characteristics of the images captured before and after the event (Matsuoka et al., 2004). In addition to 2D radiometric features, the changes in 3D geometry especially the height information of pre- and post-event is recognized as a key metric in identification of damaged building (Turker and Cetinkaya, 2005). Conventionally, the high resolution stereo image pairs, for example the stereo image pairs provided by satellites like IKONOS were being used to derive information such as heights of the buildings through constructing a digital elevation model (DEM) (Tong et al., 2012a). Besides space-borne images, overlapping aerial images captured by aircrafts are also useful to generate such DEM for identifying the damages. For example, Turker and Cetinkaya (2005) mapped the damaged buildings by differencing the DEM of the pre- and post-seismic aerial images. However, this multi-temporal approach has potential limitations. For example, it can produce the overall changes in the pre- and post- event and cannot effectively differentiate between the changes caused by damage and by other kind (Li et al., 2010a). Also the availability of pre-event data of the required quality at the time of emergency cannot be guaranteed. Therefore, with the advancements in computer vision, pattern recognition, machine learning and other related fields, the research community inclined towards the development of methods for automatic identification of building damages from mono-temporal post-event images alone. Pertaining to this, several automated methods based on mono-temporal data have been reported over the past decade with results comparable to multi-temporal approach (Dong and Shan, 2013). However, assessment results have been highly variable in terms of accuracy e.g., Kerle (2010) declining to allow relying only on a remote sensing-based assessment. The major reason for the reported shortcoming is that in conventional nadir view remote sensing images, only the partial information is available i.e., the façades of the building from all four

[1] https://www.disasterscharter.org/

cardinal directions are often not largely visible in this perspective. Such partial information is useful yet insufficient for reliable building damage assessment (refer to below section 1.3).

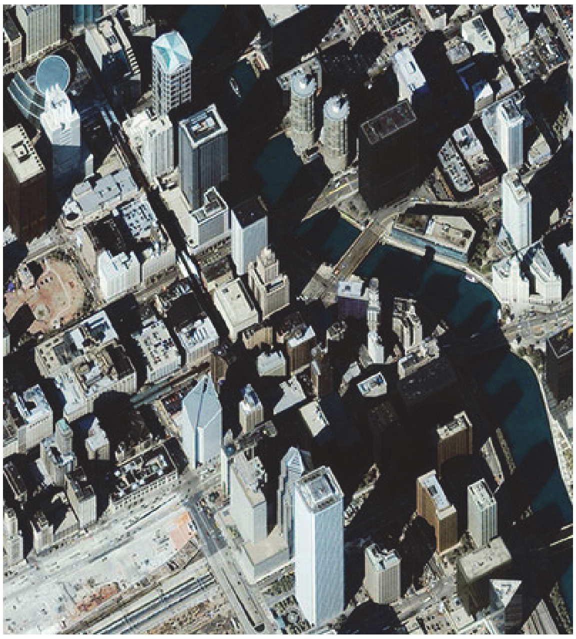


Figure 1-4 Chicago Illinois, 0.5 meter resolution image taken by the Geoeye-1 satellite
Copyright © 2010 GeoEye, Inc.

(Source: http://eoedu.belspo.be/en/satellites/geoeye.htm)

1.3 Limitations with vertical data for building damage assessment

Though, several studies reported the potentials and usefulness of conventional vertical view remote sensing data for per building-level damage assessment (Dong and Shan, 2013), they have some limitations as well. For example, recent studies have highlighted that assessments based on even very high resolution vertical imageries are often found to underestimate damage (Lemoine, 2010). This is because, nadir imagery could provide only quasi-vertical perspective of the scene in which damage along the facades and lower grade damages to buildings that do not result in building roof collapse are not largely visible; even very heavy damages such as pancake collapse with intact roofs cannot be detected (cf.

Figure 1-5) (Booth et al., 2011; Gerke and Kerle, 2011a; Saito et al., 2010). As reported by previous studies vertical images could provide better results when the focus is to differentiate between heavily collapsed and uncollapsed roofs of the buildings. However, this information is not sufficient for accurate damage assessment. In reality damage information at various levels of detail is required, such as damages along both roofs and facades, to obtain a complete and reliable damage assessment.

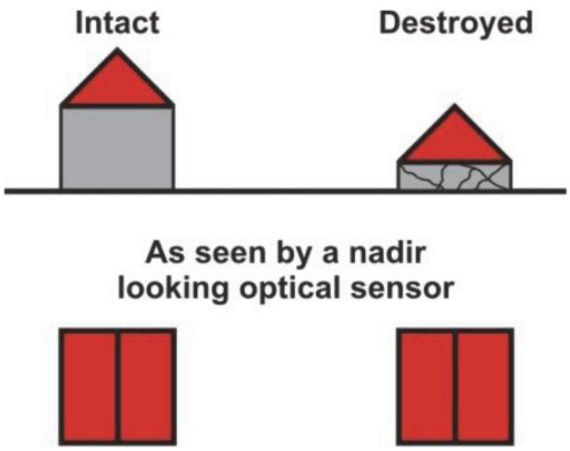


Figure 1-5 Appearance of damaged building with intact roof in nadir view (source: Plank (2014a))

1.4 Airborne oblique images for building damage assessment

Multi-view oblique airborne images particularly captured at a large tilt angle can provide more information about the façades of different sides of the buildings (Gerke and Kerle, 2011b). For instance, systems such as Pictometry (cf. Jurisch and Mountain, 2008) provide a systematically captured full set of very high resolution multi-perspective images. This includes oblique images from four cardinal directions along with one nadir image providing more information of both roof and façades of the building compared to nadir-images (cf. Figure 1-6 & Figure 1-7). These oblique images are captured with very high spatial resolution (~ 10-15 cm) where even a failure of exterior structural elements at specific story of the building is visible (Saito et al., 2010). For example, Saito et al. (2010) conducted a visual interpretation of the damage status of the individual buildings using oblique aerial imagery and reported that oblique imageries enable inferring specific damages to façades. Corbane et al. (2011)

reported that soft story failure of the building have been visually identified in airborne oblique imagery which is not visible in the nadir aerial imagery. Moreover, it is common to capture aerial images with high block overlap enabling to generate 3D point clouds based on photogrammetric processing. This provides 3D geometric information for both roof and façades in better quality than information derived from vertical images which is crucial for identifying damages to the buildings. Thus, multi-view oblique imageries are identified as the optimal alternative to vertical imagery for accurate building damage assessment regardless of its few limitations such as varying image scale and occlusions (Kerle and Hoffman, 2013a). Already, studies have demonstrated the potential of oblique aerial images and the 3D information generated from them in building damage assessment (Gerke and Kerle, 2011b).

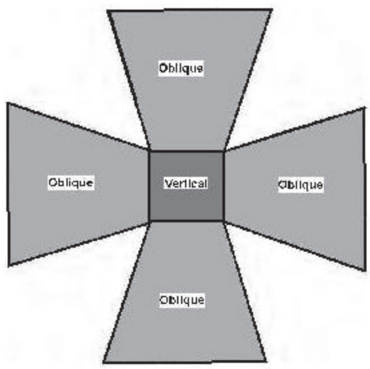


Figure 1-6 Footprints of the Pictometry camera sensor systems consisting of five cameras, one directed nadir, the others viewing forward, backward, left and right (Image Courtesy: Blom Group)

Figure 1-7 Portions of building visible in different camera views captured by Pictometry- centre image is nadir view and others are oblique view.

(Source: http://ncbc.jsums.edu/Projects/DOJ/Pictometry.aspx)

1.5 Unmanned Aerial Vehicles

Alternative to manned aircrafts, unmanned aerial vehicles (UAVs) are recognized as a new promising miniature platform for photogrammetric data acquisition. It combine the advantages of both terrestrial and aerial photogrammetry, as it can be operated at lower altitude (Colomina and Molina, 2014; Haala et al., 2013). This miniature platforms are designed to carry a variety of less weighted data acquisition systems including still or video single-lens reflex (SLR) optical and thermal cameras, LiDAR sensors, etc., and the GNSS/INS system for autonomous navigation of UAV to the planned locations and also to support the photogrammetric processing such as image orientations. The UAV system is capable of providing very high spatial resolution data as it can operate at lower altitude and can also capture data with high overlap in multi-view directions (nadir and oblique), and positions (along different sides of the building, or even inside, if structural design permits) which helps to diminish occlusion that is a major challenge in the use of oblique images acquired by traditional survey aircraft. The images captured with these flexible characteristics of the UAV facilitates generation of high quality photogrammetric products compared to images from manned aerial platforms such as dense 3D point clouds possibly (in ideal case) for the entire exterior view of individual buildings. The images in combination with such high quality 3D point clouds provide rich radiometric and geometric information essential for performing very detailed building level damage assessment. Thus UAVs have been anticipated as the most suitable platform for very detailed assessments of individual buildings. However, there are some limitations with UAV systems as well such as short flight time and small area coverage. Hence, compared to manned aircrafts, the UAV systems are more suitable particularly for local area applications (e.g., damage assessment for local municipality blocks) where repetitive, fast and cost effective data collection is desired. Moreover, UAVs can be operated remotely controlled, semi-automated or autonomous with no human pilot on-board. Thereby, it reduces the life risk of pilots for operating in highly risky disaster environments and inaccessible areas.

1.6 Research background, objectives and contributions

This PhD research is part of RECONASS (http://www.reconass.eu/), a EU_FP7 funded collaborative research project between ten research organizations across several countries. The aim of RECONASS is to develop an automated and near real time structural (building) damage assessment system to provide a stakeholders involved in response and recovery actions with continuously updated assessment of the structural condition of the monitored facilities after a natural or manmade disaster event. The RECONASS system mainly consists of two independent sub-systems for assessing damages of individual buildings based on two technologies: wireless sensor networks (WSN) and remote sensing. The sub-system based on WSN assess damages to the selected high value building based on various wireless sensors such as strain, accelerometer and temperature sensors which are equipped on the desired locations of the building. Another sub-system based on remote sensing focuses on assessing damages to the externally visible elements of the selected building equipped with sensors and also the neighbouring buildings based on the airborne oblique images and point clouds derived from them. The oblique images are preferred for aforementioned reasons. We (ITC) are the sole organization responsible for developing remote sensing sub-system within RECONASS. The primary objectives of developing this subsystem are: 1) to validate the assessments of WSN using image-based assessments; 2) to use the image-based assessments as a proxy in case of any sensor information loss; 3) to improve the sensor based assessment if any inconsistency is observed; 4) to produce local damage assessment for the RECONASS monitored and neighbouring buildings based on UAV's data; and 5) to use this local assessment to validate and calibrate the damage maps of larger areas produced by different agencies based on the satellite images. To accomplish the objectives of RECONASS, an automated extraction of damages to the building using the oblique images and the 3D point clouds derived from them is required. However, it is a challenging task and accomplishing these tasks is the focus of this PhD research. The challenges are briefly described below with necessary background, pertaining to that the research objectives and contributions of this PhD research are also presented precisely.

In existing literature, several methods have been reported for automated identification of damaged buildings using vertical view remote sensing images (Miura et al., 2013; Nex et al., 2014; Tong et al., 2012b). However, still in practice, damage maps are created based on tedious and time consuming manual visual interpretation. This is mainly because, the accuracies achieved by the reported automated approaches are not sufficient enough for practical use. There are many reasons for the limitations of the reported methods including the drawbacks mentioned earlier about the vertical imageries and the limitations with the reported methods itself which are discussed in detail in subsequent chapters. Concerning the oblique view images, the systematic capturing of these images based on manned and unmanned aerial platforms started advancing well from the beginning of this decade, since then they have been used increasingly for various remote sensing applications (Gerke and Xiao, 2013; Nyaruhuma et al., 2012; Rau et al., 2015a; Xiao et al., 2012). To this end, the research community started to exploit these oblique aerial images for automated damage assessment as they have been acknowledged to be more proficient than vertical images for this specific application (Gerke and Kerle, 2011a; Kerle and Hoffman, 2013a; Xu et al., 2014). At the beginning of this PhD research in February 2013, very few studies have been reported on the automated methods for damage detection based on oblique images from manned aircrafts which were also not matured and robust enough to consider them for practical use. At the same time the potential of oblique images based on UAVs and the point clouds derived from them was not explored for damage assessment. However, the methods reported based on vertical images could not be directly adopted for oblique images. This is because, as mentioned earlier, these methods are not robust enough and moreover, when compared to vertical images, oblique images possess complex characteristics. For example, they vary in scale and viewing directions, and often look cluttered as they capture both façades and roofs of the buildings unlike vertical images where only the roofs are largely visible. Also, the photogrammetric point clouds derived from these images are complex, since they are often noisy and contain significant gaps due to various reasons such as occlusion and inherent issues in the point cloud generation process (Vetrivel et al., 2015a). These complexities in oblique images and photogrammetric point clouds pose a number of challenges in automated extraction of useful information from them for building damage assessment. Thus, this research focuses on developing robust, reliable and practically adoptable methods and frameworks for automated damage information extraction using data with challenging characteristics i.e. both vertical and oblique images from manned and unmanned aerial platforms, in combination with photogrammetric 3D point clouds derived from them. Pertaining to this, by utilizing the advancements in various fields such as computer vision, photogrammetry, machine learning and pattern recognition, several automated methods and frameworks have been developed for performing various tasks to achieve automated and comprehensive building damage assessment: building delineation and 3D reconstruction, and automatic identification of various damage evidences such as debris, spalling, broken, inclined elements and debris quantification using both images and photogrammetric point clouds. The detailed description, significance and novelty of the proposed methods and frameworks are provided in the respective chapters.

1.7 Structure of the thesis

This thesis is organized into ten chapters. The first and last chapters are introduction and synthesis respectively, the remaining chapters are standalone scientific chapters holding specific research objectives, methodology/framework, results, discussions and conclusions. The outline of the thesis is briefly described below.

Chapter 1. Introduction: It presents the background and importance of this research, research scope and the contributions in a broader context and the outline of the thesis.

Chapter 2. Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images: For building level damage detection, the first step would be the delineation of buildings in the scene. 3D point clouds are the ideal source for building delineation. However, as mentioned earlier, the photogrammetric point clouds are often noisy and contain significant gaps. The gaps that appear in a 3D point cloud of the building are highly complex. For several reasons, these gaps can be normal and desired features (e.g., architectural elements), but they can also indicate damage. However, due to the 3D information generation process, gaps can also be created in case of partial building occlusion (e.g., by vegetation) or image matching problems. This chapter focuses on developing automated methods to identify and characterize a gap, leading to a reliable determination of openings due to structural damage. Towards this, an automated framework has been developed which includes three independent methods: 1) building delineation 2) identification of gaps in the 3D point cloud of the delineated building and 3) supervised classification model to detect damage evidences such as debris and spalling, thereby classifying the detected gaps that co-occur with these damage evidences as building openings due to damage.

Chapter 3. Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach: Automated identification of damage evidences such as debris, rubble piles and spalling is important for damage assessment as they are the significant indicators of severe damages. The methods developed in chapter 2 for detecting these evidences are not robust enough for generalization and transferability. This chapter is the extension of the damage detection work in chapter 2 where we have developed a method based on visual bag of words approach for damage detection which is more robust than the global feature representation method developed in chapter 2. The method is designed to identify the damage patterns related to rubble piles, debris and spalling, regardless of the scale and clutter of the defined region in an image.

Chapter 4. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple kernel learning: The damage detection work in chapter 3 is based on 2D texture features which works extremely well for specific study areas and is found to fail in areas with complex textures. This chapter focuses on developing methods based on deep learning features which are recently recognized to exhibit greater potential than conventional features such as textures, for damage detection process. Moreover, the 3D geometric features from point cloud are also expected to be useful for identifying these damage evidences. In this chapter, some 3D features are proposed and a framework is developed to integrate these 3D point cloud geometric features and image-based deep learning features for damage detection using multiple kernel learning approach.

Chapter 5. Potential of multi-temporal oblique airborne imagery for structural damage assessment: The damage detection methods reported in previous chapters are based on features derived from post-event images and 3D point clouds. These methods are based on the assumption that manmade undamaged areas possess more uniform radiometric and geometric pattern than damaged regions. However, this assumption fails in complex urban environments. In such cases, the pre-event reference is anticipated to be useful. Thus, this chapter focuses on developing automated methods to identify damages based on comparing pre- and post-event 3D point clouds and images.

Chapter 6. Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning: The methods developed for damaged detection in previous chapters are based on supervised learning approach where the classifier for damage detection is built based on a large number of training samples. The significance of site-specific training samples for building a robust classifier is inferred from the experiments conducted in the previous chapters. With the advancement in various technologies, it has become a common practice to make the damage information after any disaster event available online through various sources. Thus, this chapter focuses on developing a robust supervised classifier based on online learning approach where the classifier learns continuously from the aforementioned streaming site-specific training samples which are made available online by various sources.

Chapter 7. Segmentation of UAV-based images incorporating 3D point cloud information: A 3D model of buildings is realised as the best representation for comprehensive damage assessment by systematic integration of damage evidences that are detected along various sides of the building. However, it is challenging to construct an accurate 3D model even for undamaged building from the image-based 3D point clouds as it is difficult to obtain the accurate roof segments as they are often too noisy. Thus, this chapter focuses on developing methods for accurate delineation of roof segments from noisy point cloud.

Chapter 8. Automatic 3D reconstruction of buildings by synergistic use of UAV images and derived 3D point clouds for detailed structural damage assessment: The developed method in chapter 7 for accurate building roof segment delineation is based on pixel-based approach where

it possesses some limitations. Hence, as an extension of this method, a framework has been developed in this chapter for accurate roof delineation and 3D reconstruction of the building using segment-based approach from the noisy point cloud and images by addressing the inherent challenges.
Chapter 9. Evaluation of methods based on EU-FP7 project RECONASS: Several datasets are considered in this research for the evaluation of developed methods. However, among them, only specific datasets are used to evaluate a specific method as they possess interesting characteristics and best suited for evaluating that particular method or other datasets were not available at that point of time when the particular method was developed and evaluated. Hence, in this chapter, the developed methods in previous chapters pertaining to building damage assessment are evaluated collectively using a single dataset from the EU FP7 project RECONASS which satisfies the requirement for evaluating most of the methods developed in this research.
Chapter 10. Synthesis: This chapter presents the summary of this research findings and contributions, conclusions and recommendations for future research by extending or improving the methods reported in this research to achieve a reliable, operational and automated building damage assessment using remote sensing technologies.

Most chapters of this thesis are based on the published journal and conference papers. There are some repetitive information between chapters, for example, the background, significance and justification for using oblique view images and UAVs will be presented in most chapters' introduction. This is desired as this makes every chapter standalone, so the reader can directly read the interested chapters without having any dependencies on previous chapters.

1.8 References of chapter 1

Booth, E., Saito, K., Spence, R., Madabhushi, G., Eguchi, R.T., 2011. Validating assessments of seismic damage made from remote sensing. *Earthquake Spectra,* 27 (S1), pp. S157-S177.
Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 92, pp. 79-97.
Corbane, C., Saito, K., Dell'Oro, L., Gill, S., Piard, B., Huyck, C., Kemper, T., Lemoine, G., Spence, R., Krishnan, R., 2011. A comprehensive analysis of building damage in the January 12, 2010 Mw7 Haiti earthquake using high-resolution satellite and aerial imagery. *Photogrammetric Engineering and Remote Sensing (PE&RS),* 77 (10), pp. 1.

Dell'Acqua, F., Gamba, P., 2012. Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE,* 100 (10), pp. 2876-2890.

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.

Gerke, M., Kerle, N., 2011a. Automatic structural seismic damage assessment with airborne oblique pictometry imagery. *Photogrammetric Engineering & Remote Sensing,* 77 (9), pp. 885-898.

Gerke, M., Kerle, N., 2011b. Automatic structural seismic damage assessment with airborne oblique Pictometry© imagery. *Photogrammetric Engineering & Remote Sensing,* 77 (9), pp. 885-898.

Gerke, M., Xiao, J., 2013. Supervised and unsupervised MRF based 3D scene classification in multiple view airborne oblique images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 2 (3), pp. 25-30.

Haala, N., Cramer, M., Rothermel, M., 2013. Quality of 3D point clouds from highly overlapping UAV imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XL-1/W2, pp. 183-188.

Jurisch, A., Mountain, D., 2008. Evaluating the viability of Pictometry® imagery for creating models of the built environment, *International Conference on Computational Science and Its Applications*. Springer, pp. 663-677.

Kerle, N., 2010. Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it? *International Journal of Applied Earth Observation and Geoinformation,* 12 (6), pp. 466-476.

Kerle, N., Hoffman, R.R., 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural hazards and earth system sciences (NHESS),* 13 (1), pp. 97-113.

Lemoine, G., 2010. Validation of building damage assessments based on post-Haiti 2010 earthquake imagery using multi-source reference data, *Proceedings of the 2nd International Workshop on Validation of Geo-Information Products for Crisis Management*, pp. 11-13.

Li, P., Xu, H., Guo, J., 2010. Urban building damage detection from very high resolution imagery using OCSVM and spatial features. *International Journal of Remote Sensing,* 31 (13), pp. 3393-3409.

Matsuoka, M., Vu, T.T., Yamazaki, F., 2004. Automated damage detection and visualization of the 2003 Bam, Iran, earthquake using high-resolution satellite images, *Proc. 25th Asian Conf. Remote Sens*, pp. 841-845.

Miura, H., Midorikawa, S., Kerle, N., 2013. Detection of building damage areas of the 2006 Central Java, Indonesia, earthquake through digital analysis of optical satellite images. *Earthquake Spectra,* 29 (2), pp. 453-473.

Nex, F., Rupnik, E., Toschi, I., Remondino, F., 2014. Automated processing of high resolution airborne images for earthquake damage assessment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XL-1, pp. 315-321.

Nyaruhuma, A.P., Gerke, M., Vosselman, G., Mtalo, E.G., 2012. Verification of 2D building outlines using oblique airborne images. *ISPRS Journal of Photogrammetry and Remote Sensing,* 71, pp. 62-75.
Plank, S., 2014. Rapid damage assessment by means of multi-temporal SAR—A comprehensive review and outlook to Sentinel-1. *Remote Sensing,* 6 (6), pp. 4870-4906.
Rau, J.-Y., Jhan, J.-P., Hsu, Y.-C., 2015. Analysis of oblique aerial images for land cover and point cloud classification in an urban environment. *IEEE Transactions on Geoscience and Remote Sensing,* 53 (3), pp. 1304-1319.
Saito, K., Spence, R., Booth, E., Madabhushi, M., Eguchi, R., Gill, S., 2010. Damage assessment of Port-au-Prince using Pictometry, *8th international workshop on remote sensing for disaster management*. Tokyo.
Tong, X., Hong, Z., Liu, S., Zhang, X., Xie, H., Li, Z., Yang, S., Wang, W., Bao, F., 2012a. Building-damage detection using pre- and post-seismic high-resolution satellite stereo imagery: A case study of the May 2008 Wenchuan earthquake. *ISPRS Journal of Photogrammetry and Remote Sensing,* 68, pp. 13-27.
Tong, X., Hong, Z., Liu, S., Zhang, X., Xie, H., Li, Z., Yang, S., Wang, W., Bao, F., 2012b. Building-damage detection using pre-and post-seismic high-resolution satellite stereo imagery: A case study of the May 2008 Wenchuan earthquake. *ISPRS Journal of Photogrammetry and Remote Sensing,* 68, pp. 13-27.
Turker, M., Cetinkaya, B., 2005. Automatic detection of earthquake-damaged buildings using DEMs created from pre-and post-earthquake stereo aerial photographs. *International Journal of Remote Sensing,* 26 (4), pp. 823-832.
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, pp. 61-78.
Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., Mehl, H., 2007. Satellite image analysis for disaster and crisis-management support. *IEEE Transactions on Geoscience and Remote Sensing,* 45 (6), pp. 1520-1528.
Xiao, J., Gerke, M., Vosselman, G., 2012. Building extraction from oblique airborne imagery based on robust façade detection. *ISPRS Journal of Photogrammetry and Remote Sensing,* 68, pp. 56-68.
Xu, Z., Wu, L., Shen, Y., Wang, Q., Wang, R., Li, F., 2014. Extraction of damaged building's geometric features from multi-source point clouds, *IEEE Geoscience and Remote Sensing Symposium*, pp. 4764-4767.

2 Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images*

* This chapter is based on the article:
Vetrivel, A., Gerke, M., Kerle, N. and Vosselman, G. (2015) Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images. In: ISPRS Journal of Photogrammetry and Remote Sensing, 105 (2015) pp. 61-78.

Abstract

Point clouds generated from airborne oblique images have become a suitable source for detailed building damage assessment after a disaster event, since they provide the essential geometric and radiometric features of both roof and façades of the building. However, they often contain gaps that result either from physical damage or from a range of image artefacts or data acquisition conditions. A clear understanding of those reasons, and accurate classification of gap-type, are critical for 3D geometry-based damage assessment. In this study, a methodology was developed to delineate buildings from a point cloud and classify the present gaps. The building delineation process was carried out by identifying and merging the roof segments of single buildings from the pre-segmented 3D point cloud. This approach detected 96% of the buildings from a point cloud generated using airborne oblique images. The gap detection and classification methods were tested using two other data sets obtained with Unmanned Aerial Vehicle (UAV) images with a ground resolution of around 1-2 cm. The methods detected all significant gaps and correctly identified the gaps due to damage. The gaps due to damage were identified based on the surrounding damage pattern, applying Gabor wavelets and a histogram of gradient orientation features. Two learning algorithms – SVM and Random Forests were tested for mapping the damaged regions based on radiometric descriptors. The learning model based on Gabor features with Random Forests performed best, identifying 95% of the damaged regions. The generalisation performance of the supervised model, however, was less successful: quality measures decreased by around 15 to 30%.

2.1 Introduction

Rapid and detailed damage assessment on a per building level after disaster events such as earthquakes has become imperative for initiating effective emergency response and recovery actions. However, manual assessment is expensive, time consuming and may be hindered by site accessibility problems. Remote sensing technology has been recognized as a potential alternative to manual damage assessment. Satellite and aerial images are the conventional data source for assessing the extent and degree of damage after an event (Voigt et al., 2007). Damage assessment on a per-building scale has become possible with the availability of remote sensing images with very high spatial resolution up to sub-meter precision (Ehrlich et al., 2009). Numerous semi-automatic and automatic methods have been reported for image-based building damage assessment (Chen et al., 2011; Rathje et al., 2005). However, assessment results have been highly variable e.g., Kerle (2010) and generally not accurate enough to allow relying only on a remote sensing-based assessment. The major reason for the reported shortcoming is that in conventional nadir view remote sensing images roof portions alone are visible and damages along the façades cannot be assessed (Gerke and Kerle, 2011a). Therefore, such analysis depends strongly on the use of proxies such as presence of blow-out debris or shadow changes (Kerle and Hoffman, 2013b). Airborne oblique images, which provide information of both roofs and facades, have been recommended as a potential data source for a more comprehensive building damage assessment (Dong and Shan, 2013; Kerle and Hoffman, 2013b). However, in a study of damage caused by the 2010 Haiti earthquake using oblique Pictometry data, the reported classification accuracy was only around 70% when the categories no-moderate damage (D1-D3), heavy damage (D4) and complete destruction (D5), using the European Macroseismic Scale (EMS 98) for damage classification (Gerke and Kerle, 2011a). The main shortcoming of this approach is that through relatively simple appearance based features, collected on a segment basis, the large variability of damage cannot be easily represented. In fact, such an approach leads to ambiguities in the damage classification approach. Each damage evidence is unique in its characteristics and requires unique features and a unique processing strategy to recognize them. For instance, the recognition of inclined elements requires 3D geometric features, whereas detection of cracks, spalling, etc., requires radiometric features, with damage such as partial collapse requiring both. Also the damage evidences have different impact when they occur on different elements of the building, especially when affecting structural (roof, façade, etc.) vs.

non-structural elements (windows, doors, etc.). Hence, an explicit definition of damage evidences, along with a categorisation of building elements, is important for accurate damage classification. Therefore, for a detailed and comprehensive damage assessment, a multi-level processing is required, where at each level a relevant set of features and a processing appropriate strategy should be adopted to recognize a specific type of damage evidence. This kind of assessment demands a detailed inventory of the building with its rich geometric and radiometric features. 3D point clouds have been recognized as an ideal source to infer the 3D geometric features (Lafarge and Mallet, 2012).

Unmanned Aerial Vehicles (UAVs) have been recognized as an ideal platform for capturing images suitable for high quality 3D point cloud generation (Colomina and Molina, 2014). UAVs provide flexibility in capturing images of the building in multiple views and from multiple positions, which helps to diminish occlusion that is a major challenge in the use of oblique images acquired by traditional survey aircraft. UAVs can also be operated at lower altitudes. Hence, they can provide images with better spatial resolution than manned aircraft.

There are a number of challenges associated with the use of image-based 3D point clouds that need to be addressed for reliable damage assessment. One of those relates to gaps (holes) in the 3D point cloud of the building elements, which can be the result of actual damage (partially broken), image matching problems (low texture, lack of image coverage), occlusion, or absence of objects or surfaces (real openings). The identification of gaps in the 3D point cloud of an intact building, as well as their specific nature, is a prerequisite for the damage assessment scenario development. Additional radiometric features from images are required for accurate detection and classification. To our knowledge, a gap classification scheme for 3D point clouds that is suitable for damage assessment scenarios has not yet been developed.

The objective of this research is to develop a methodology for delineation of individual buildings in the image-based 3D point clouds and mapping of specific kinds of damage evidences (e.g., broken segments) along the elements of the building that are related to gaps in 3D point clouds. This methodology attempts to utilize the advantages of both 3D geometry and multiple image features to resolve the ambiguities/limitations that cannot be addressed by either alone. The methodology includes three principal processes: (i) delineation of individual buildings to be assessed; (ii) 3D point clouds gap detection and (iii) gap classification, in particular identification of gaps associated with different types of damage. The challenges associated with each process, related methodology and results are independently analysed and discussed in the chapter.

2.2 Previous work

Numerous methods have been reported for building damage assessment using various remote sensing data such as optical, SAR and LiDAR (Dell'Acqua and Gamba, 2012; Dong and Shan, 2013; Plank, 2014b). The reported methods can be categorized into two groups: 1) damage assessment by analysing the significant changes between the pre and post event data, and 2) damage assessment by analysing the post event data alone (mono-temporal). The latter approach has been found to be more suitable for time critical applications such as disaster management, as it can be operational even in absence of reference data (Dong and Shan, 2013). With advances in related fields such as computer vision, pattern recognition and machine learning, numerous methods have been developed to infer the damage pattern from mono-temporal post event data. Various studies explored texture-based features as important cues for damage classification, since damaged regions tends to show unique texture pattern in contrast to undamaged manmade or natural objects, see for example, Ma and Qin (2012a), Radhika et al. (2012) and Yamazaki and Matsuoka (2007). Li et al. (2012) detected collapsed buildings using mono-temporal post-event data by fusing morphological texture features with spectral information, and reported that collapsed buildings were detected even in a complex urban environment. Some of the studies highlighted that 3D features were useful for detecting specific types of damage based on geometric reasoning (Gerke and Kerle, 2011c; Rehor and Vögtle 2008; Shen et al., 2010). So far, very few studies have investigated the potential of 3D point clouds, especially from LiDAR, for building damage assessment. Oude Elberink et al. (2011) used mono-temporal LiDAR point clouds to detect completely collapsed buildings. Shen et al. (2010) identified damaged buildings by estimating building inclination in post-disaster airborne laser data. Khoshelham et al. (2013a) classified building roofs into intact and damaged, using segment-based 3D features derived from LiDAR data. They reported that partially damaged larger roof segments were mostly misclassified as intact roofs. This implies that geometric features alone are not sufficient for reliable damage assessment. Apparently a combined use of 3D geometric features from LiDAR and radiometric features from optical images could be the ideal combination for damage assessment. Hussain et al. (2011) demonstrated that the combined use of LiDAR data with GeoEye-1 imagery produced the effective damage assessment map for the damage caused by earthquakes in dense urban areas, such as in Port-au-Prince in 2010. Trinder and Salah (2012) reported improved results when LiDAR data are used along with aerial images rather than using aerial images alone.

All the mentioned methods assess the damage state of the building as a whole and assign the single damage label for the entire building, even if only a specific portion of the building is severely damaged. Essentially element-specific detailed building damage information is wanted since different stakeholders (rescue team to insurance company) need different levels of damage information. This requires an accurate recognition of individual elements of the building. Recent studies in non-disaster domains have shown that 3D point clouds can be used to automatically recognize the structural components of the buildings, even in the presence of significant clutter and occlusion (Xiong et al., 2013). Most of the reported 3D point cloud-based urban scene classifications have so far been based on LiDAR point clouds. However, the technological advancements in the field of Computer Vision and computing made image-based stereo photogrammetric 3D cloud points comparable with LiDAR cloud points, at least in urban environments (Leberl et al., 2010). Also the advancement in UAV technology allows the capturing of high resolution data, suitable for photogrammetric processing for generating dense 3D point clouds of the scene.

2.3 Methods

A framework for building delineation from an image-based 3D point cloud and mapping of damaged elements of the delineated building that are related to the gaps in the 3D point cloud is developed. The framework comprises of three independent methods 1) building detection and delineation from image-based 3D point cloud, 2) detection of gaps in the 3D point cloud corresponding to individual elements of the building and 3) identification of gaps caused by damage. The description and related background of each method are provided in the following subsections.

2.3.1 Initial building detection and delineation

In our overall building damage assessment strategy we differentiate between fully demolished buildings and those where the roof is at least partially intact. In this study, we concentrate on the latter case. Hence, the prerequisite for our building damage assessment is delineation of individual buildings from 3D point clouds that are fully/partially intact with some roof structures. In this study, we intended to explore the suitability of methods initially designed for LiDAR point clouds (Rottensteiner et al., 2014; Sun and Salvaggio, 2013) for image-based 3D point clouds. In addition we prefer a rooftop-based building delineation approach. This is for two reasons, 1) roof segments can be identified using simple geometric constraints i.e. they will always be located at a certain height above the ground and characteristically they will be horizontal or slanted planes, and 2) generally roof segments have better point cloud quality than other building elements such as façades, since roofs are visible in both nadir and oblique views and less affected by occlusion effects when compared to façades. Our aim is to detect only the building roofs, but often tree segments are misclassified as roof segment. Hence, tree detection will generally be carried out as a pre-processing step in 3D point cloud-based building detection (Sun and Salvaggio, 2013). Some studies indicated local distributions of point normal as a useful feature to classify between tree and non-tree component (Sun and Salvaggio, 2013). However, a tree portion with dense leafs may possess required uniformity in their point normal distribution to get recognized as a planar segment. In such cases, the differentiation of tree and roof segments based on geometric features becomes ambiguous. However, the image-based 3D point cloud additionally has spectral information for each 3D point, which can be used to overcome the ambiguity in classification between vegetation and building roof segment.

Another issue to be addressed is the presence of large noise or outliers in the image-based 3D point clouds, which is mainly due to the generation of mismatches (Rupnik et al., 2014b). During plane-based segmentation a certain point offset from an optimal plane is allowed, hence erroneous 3D points might form a segment and possibly they may get connected with roof segments, thereby affecting the building delineation accuracy. In general, we assume that segments formed by erroneous points would be too compact or too narrow when compared to real roof segments. This kind of noisy segment may be connected to roof segments through single points. In contrast, real roof segments should be larger in general. Refer to threshold T_{width} below.

The building detection and delineation has been carried out based on identifying spatially connected roof segments. As an initial step the 3D points are classified into terrain and off-terrain points using the method proposed by Axelsson (2000b) which is implemented as part of the software *lastools* (Rapidlasso, 2013). The off-terrain 3D point cloud is segmented into disjointed planar segments, using the plane-based segmentation as described in Vosselman (2012). Then the height normalization of the 3D point cloud is carried out i.e. the above ground height for each off-terrain 3D point is computed as the height difference between the off-terrain 3D point and the nearest terrain point. The roof segments are then identified based on the z component of plane normal of that segment. Then a connected component analysis is applied to merge the spatially connected roof segments that satisfy predefined merging criteria to obtain a complete roof of a single building. Four merging criteria are used in order to avoid tree and noisy segments to get included as building roof segments.
Merging criteria,

i. Distance between the segments < $T_{distance}$.
ii. Number of points that create a connection between the segments (i.e. number of points in one segment that are spatially connected to the convex hull of another segment) > T_{points}.
iii. The distance between any two connecting points from step (ii) > T_{width}.
iv. Spectral index (G/(R+G+B)) > T_{index}, RGB corresponds to the red, green and blue band, respectively.

The choice of thresholds for those merging criteria plays an important role and will be discussed in the results section.

The building detection and delineation process has been carried out through a number of steps as follows,

1. The Z component of surface normal of each segment (nz) is computed through a plane fitting method (Rabbani et al., 2006). It is ensured that all normals point outside, i.e. all nz values are larger than zero.
2. The area for each segment is computed using the convex hull.
3. The segments that have nz > T_Z (pre-defined threshold) are pre-selected as roof segments and sorted by area in descending order.
4. The roof segments that are spatially connected are considered as roof segments of a single building. The connected roof segments are merged to form a complete roof segment that represents the area of the

building. The building delineation based on roof merging process is carried out through a number of steps as follows,

a) The largest roof segment is taken as the base element for the roof merging process.
b) The boundary of the base element is defined using the convex hull.
c) All segments that are spatially connected to the boundary of the base element are merged with the base element and its boundary is updated using the convex hull. The spatial connectivity constraint (conditions for segments to be agreed as connected) and the merging process (conditions for merging the connected segments) are based on the predefined merging criteria.
d) The above process (c) is repeated with the updated base segment until no new segment is added to the base segment.
e) The final base segment will be considered as a complete roof segment of a single building.
f) All 3D points (both segmented and un-segmented) that lie within the 2D boundary defined by a convex hull of the final roof segment (i.e. all elements that lie below the final roof segment) will also be registered as points of that building.
g) Finally the convex hull of the detected building is considered as the outline of the building.
h) The above steps (a) to (g) are continued until the roof segment list becomes empty.

2.3.2 Gap detection on element level

A building delineated from 3D point clouds consists of a collection of segments (e.g., building elements like roofs, façades, etc.) and each segment consists of a collection of 3D points, in addition isolated points in the vicinity are added (step f. above).

For the gap detection process, 3D points of a selected segment are voxelized based on a pre-defined voxel size. Then the voxels are classified into occupied voxels (at least one 3D point lies inside the voxel) and unoccupied voxels (no 3D point inside the voxel). In this context it is important to consider also the minimum number of image points required to compute a 3D point. This parameter is set within the image matching method. Per default we assume that this number is three. The unoccupied voxels are considered as gaps in the 3D point cloud, and can be further classified into occluded empty voxels (visible in less than three camera views) and visible empty voxels (visible in three or more camera views) by applying a visibility test analysis. In this work, we used the voxel-ray intersection method as described by Alsadik et al. (2014b) for mapping of the occluded empty voxels. The remaining voxels which are visible in a sufficient number of cameras, but are not occupied will be further considered for the gap classification process.

2.3.3 Gap classification

A gap classification process is crucial to identify the gaps caused by damage. In general, gaps can be classified into four categories: 1) gap due to occlusion, 2) gap due to failure in 3D point generation (e.g., low texture, shadow), 3) gap due to openings in architectural design and 4) gap due to damage. The gap due to occlusion will already be classified during the gap detection process through visibility test analysis, but the remaining gap categories cannot be recognized from 3D point features alone. For example, 3D points cannot be generated for texture-less surfaces which leads to a gap in the 3D point could. This kind of gaps can be classified only by analysing the surface radiometric characteristics of that gap region in an image. Also the radiometric features are required to differentiate between a gap due to an opening in architectural design and a gap created due to damage. For this analysis we assume that undamaged urban objects are more homogenous in nature and possess uniform radiometric distribution whereas damaged region will tend to show more irregular radiometric distribution patterns. For example, any deformation in the concrete surface creates a sign of spalling or debris around the deformed region which are generally rough and flaky in nature and possess uneven radiometric distribution pattern as depicted in Figure 2-1Figure 1-1. Therefore we assume that a gap with spalling or debris in its surrounding region can be classified as a gap due to damage. Spalling and debris regions in an image can be identified using the radiometric descriptors, as explained in the next subsection.

Figure 2-1 Texture pattern of damaged region
Source: Wikipedia (2015)

2.3.3.1 Radiometric descriptors as damage indicator:

Numerous radiometric features have been reported for mapping damaged regions. Among them features describing the texture have been noted as a significant indicators of unevenness radiometric distribution that corresponds to damaged region (Dong and Shan, 2013). In addition our working hypothesis is that the gradient orientation distribution of a region could indicate the actual damage state of that region. To this end we analyse the histogram of gradient orientation (HoG) feature, as well (Minetto et al., 2013; Salas et al., 2012). Based on those findings, the significance of HoG and other texture descriptors is analysed in identification of damaged regions in this work.

i) Texture features

The texture feature extraction methods can be categorized into statistical and signal processing approaches. Statistical methods define texture in terms of local grey–level statistics by analysing the spatial distribution of grey values within a specified image. An example, features based on the grey level co-occurrence matrix (GLCM). Signal processing based texture analysis is carried out by analysing the spatial-frequency characteristics of an image region, e.g., wavelets based features. Wavelet-based texture features were found to be superior to GLCM texture features in many applications including classification of remote sensing images (Ruiz et al., 2004a). Also in damage assessment scenarios wavelet features were indicated as potential features in recognition of debris pattern (Radhika et al., 2012). Among wavelets, Gabor wavelets have been noted as an efficient tool for texture analysis especially for analysing highly specific frequency and orientation characteristics of an image region which is crucial for damage detection (Arivazhagan et al., 2006). Gabor wavelets consist of set of filter banks, where each filter is tuned to capture specific frequency information at specific orientations. Using Gabor filters, the image regions can be differentiated with respect to its dominant spatial frequency and orientation. Detailed information about Gabor filter banks generation and application for pattern recognition are given by Arivazhagan et al. (2006).

ii) Histogram of gradient orientation (HoG)

HoG features are widely used in computer vision for object detection and classification by measuring the spatial variation of edge orientations within a region (Kobayashi et al., 2008). The principle behind HoG feature-based image analysis is that local object appearance within an image can be characterized by measuring the distribution of the gradient directions. The HoG features can be derived locally by dividing the image into m x n rectangular grids and by computing the histogram of gradient orientations within the rectangular grid. This histogram will then be converted into a feature vector that represents the gradient orientation pattern of an image region corresponding to the rectangular grid.

2.3.3.2 Procedure for gap classification

The gap classification process has been carried out through a number of steps as follows,

i) Choice of suitable image for gap classification:

The first step is the choice of an appropriate image for gap classification. The appropriateness of an image is decided based on the following criteria,

1. The angle between the normal orientation of the segment that contains one or more gap regions and the direction of the optical axis of the camera should be within the threshold T_{angle}. The image is automatically selected as the one where the angle between optical axis and face normal is minimal.
2. The gap region must lie within the boundary of the camera view and not occluded by other objects.

3. From the selected images based on the above two criteria, select the one where the distance between the segment and the position of the camera is minimal for better spatial resolution.

ii) Delineation of gap region on the image for analysis

The next step is to delineate the gap regions in a selected image. As a first step, the centroids of gap voxels are projected onto the image. A morphological closing operation is performed to fill the gaps between the projected points. Then a connected component analysis is used to delineate the individual gap regions. A gap due to damage is classified based on the presence of damage patterns around the gap region. Therefore the delineated gap regions are dilated using a square structural element of size m, in order to include neighbouring pixels around the gap region for further analysis.

iii) Classification of delineated gap regions

a) Gap due to damage

The gaps due to damage are classified based on the presence/ absence of damage pattern around them. The Gabor wavelets and HoG features are extracted around the gap region and independently evaluated for mapping the damaged region.

Supervised learning methods are being widely used for pattern recognition and classification in specific application, since they are more effective to develop a relational model by just providing a set of training samples rather than defining a relationship based on domain knowledge. In image-based damage assessment applications, supervised learning approaches have extensively been applied to recognize damage pattern from radiometric descriptors which is evident from the review by Dong and Shan (2013). Hence, the potential of Gabor wavelets and HoG features for damage mapping is evaluated using two learning algorithms of different paradigms, Support Vector Machine (SVM) (Schölkopf and Smola, 2001) and Random Forests (RF) (Breiman, 2001). SVM and RF are state-of-the-art and widely used machine learning algorithms of two different families. SVM and RF work in unique way. For example, random forest is a decision tree-based learning algorithm that works well for linearly separable features, and that can handle data with missing values and outliers, whereas SVM employs kernel-based learning and works well for non-linear features. A main objective of experiments conducted in the study is thus to analyse the performance of features (Gabor/HoG) and the respective learning model (SVM/RF) for the classification of region into damaged and non-damaged region. Another objective in this context relates to the question of generalisation performance: given that the classifier is trained using images from different areas, and not the actual one, how is the overall classification performance? Especially in rapid damage assessment this is an important property in order to save time for the assessment.

b) Gap due to openings in the architectural design

The gap region that has no damage pattern around them is then tested for an opening in the architectural design. All 3D points that are registered with the delineated building (c.f. step 4f under section 2.3.1.1) are projected over the selected image. The occurrence of projected 3D points over the delineated gap region in an image indicates the presence of visible surface around the gap region. Therefore, the gap region in an image with projected 3D points and no damage pattern around is classified as an opening in architectural design.

c) Gap due to image matching failure

The gap region in an image with no damage pattern around them and no projected 3D points is then analysed based on its radiometric characteristics to further classify them into gap region corresponding to textured or non-textured surface/area. Three features (contrast, entropy and peak histogram intensity) are used to analyse the surface characteristics of the gap. High contrast and entropy values indicate that a good surface texture is present, but the matching algorithm was not successful in point creation in the area. On the other hand, if the named observations show low values, it is a hint on a poorly textured or shadowed area which hampered the image matching

The automated procedure for overall gap classification process is depicted in Figure 2-2.

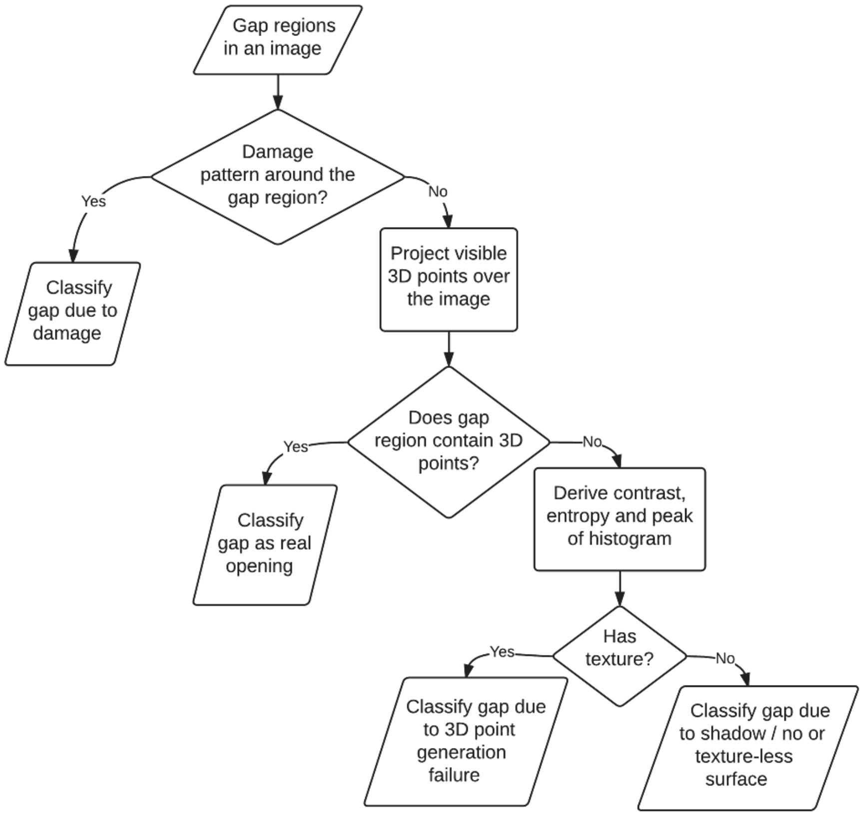


Figure 2-2 Work flow of the gap classification process

The integrated work flow of all the three processes such as building delineation, gap detection and gap classification is depicted in Figure 2-3. All the processes can be carried out automatically in sequence, once the parameters associated with each process are initialized.

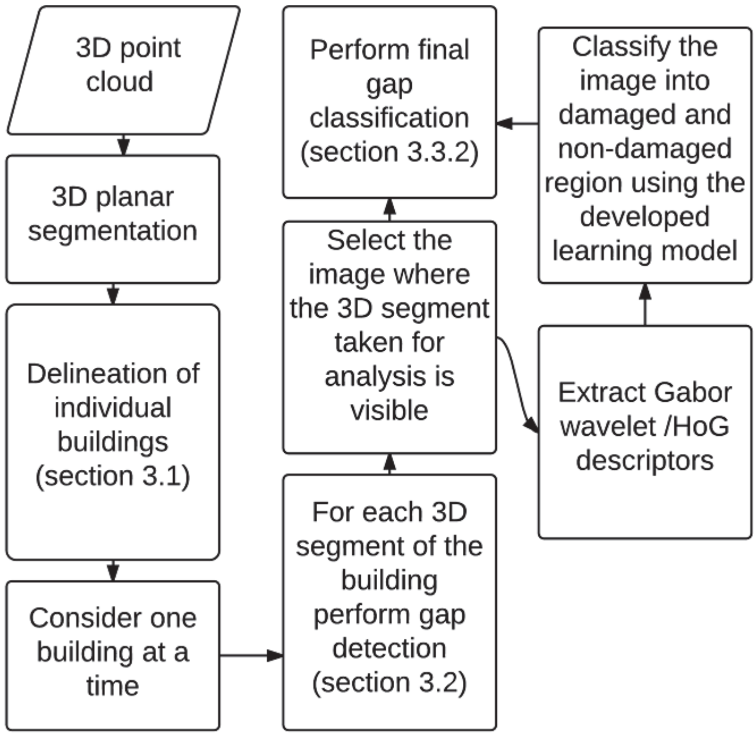


Figure 2-3 Integrated work flow of all three processes (building delineation, gap detection and gap classification)

2.4 Experimental results and discussion

2.4.1 Building detection and delineation – dataset: oblique images from manned aircraft

2.4.1.1 Data description

At the time of preparing the experiments for the building detection and delineation method we had no access to UAV data over a large earthquake-affected area. For this reason we chose a multiple view oblique airborne image data set which is described below. The data for the gap detection and classification, however, was taken from a UAV flight. The airborne oblique images captured over the city of Mirabello (Italy) after an earthquake in 2012 were taken as the primary data to evaluate the building delineation process where most of the buildings only show limited or no damaged. The images with an average ground sampling distance (GSD) of 14 cm were captured by Blom-CGR S.p.A. (*http://www.cgrspa.com*), using the Midas- Oblique system composed of 5 cameras –one nadir and

4 oblique rotated in the four cardinal directions and tilted by 45° with respect to nadir. From the 70 captured images a dense 3D point cloud of the scene with an average point density of 15 points per m^2 was generated by automatic orientation of the images, followed by dense matching using the software pix4Dmapper (http://pix4d.com). A subset of the 3D point cloud corresponding to the densely built-up region was extracted for testing of the building delineation process. The selected region consists of 198 buildings. Of those 106 are isolated buildings (not connected to neighbouring buildings), and the remaining 92 buildings can be categorized as chains of buildings (building physically connected to neighbouring buildings). The selected region also consists of numerous densely leafed tall trees, some of which overlap with buildings.

2.4.1.2 Results of building detection and delineation process

The height normalized above ground 3D points was segmented into disjointed planar segments. The segments that has Z component of normal greater than T_Z (0.6) were filtered out as a roof segments. The building delineation method described in section 0 was used to delineate the buildings from the detected roof segments of the buildings.

The building delineation process was carried out based on four merging criteria. We used $T_{distance}$ as 0.5 m, T_{index} as 0.34, T_{points} as 3 and T_{width} as 30cm for building delineation process.

The threshold T_{index} for classifying 3D points as tree and non-tree component was defined based on Otsu threshold selection method (Otsu, 1979). This spectral index performed well for our test data where most of the tree segments were identified. For visual comparison, the images projected with 3D points of the delineated buildings with and without use of spectral index as merging criteria for removing the tree segments are shown in Figure 2-4.

The thresholds T_{points} and T_{width} were defined based on trial and error basis. It was observed that the threshold T_{width} has significant impact on building delineation process, since for higher values of T_{width} e.g., T_{width} > 50 cm, a single building tended to delineate as multiple buildings as depicted in Figure 2-5.

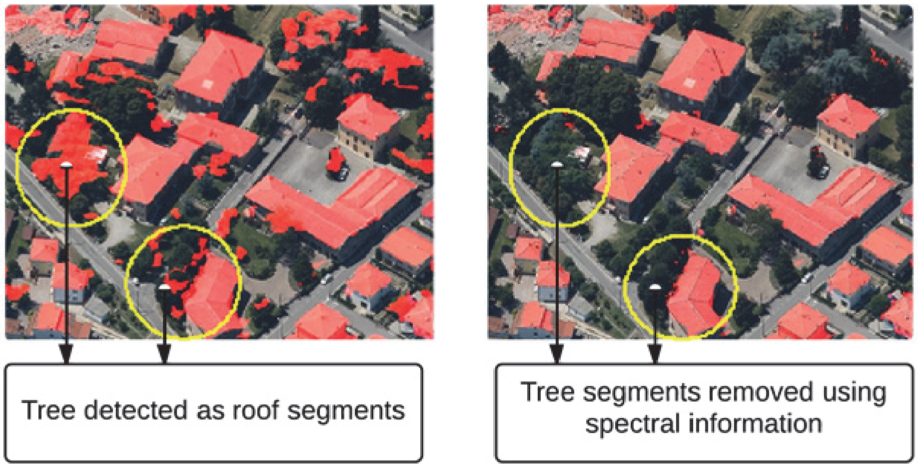
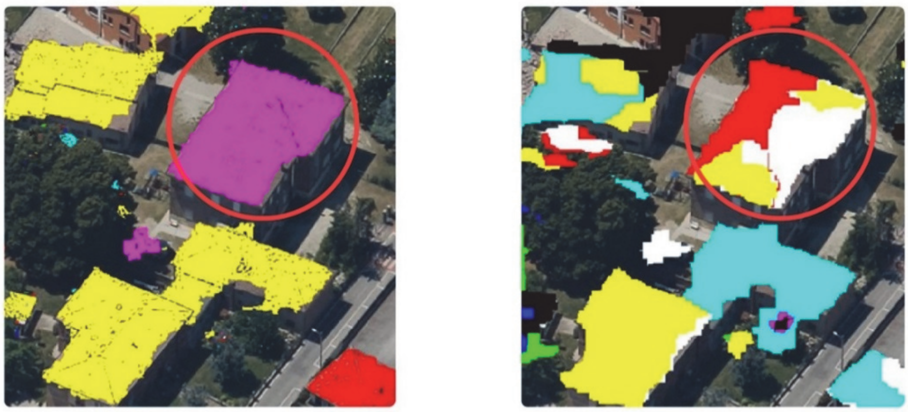


Figure 2-4 The delineated buildings projected over the images (highlighted in red) – building delineation without use of spectral index as merging criteria (left), building delineation with spectral index as one of the merging criteria (right), images © BLOM Italy



(a) Single building detected as single building

(b) Single building detected as multiple buildings

Figure 2-5 Example of the impact of the merging criteria threshold on building delineation – (a) T_{width} > 30 cm, and (b) T_{width} >50 cm, images © BLOM Italy

The 3D points of the delineated buildings were projected over the image to analyse the results visually. A portion of an image projected with 3D points of the delineated buildings is shown in Figure 2-6. From the results we differentiate between the following cases (refer to Figure 2-6, where one example building for each case is annotated):

Case 1: Isolated single building delineated as single building – this is the case, where a single building was correctly detected as single building and hence it was considered as topologically correct.

Case 2: Single building delineated as two or more buildings – this is due to the gaps between the 3D points corresponding to a single roof segment. In the image-based point cloud, the point density varies from region to region (i.e. from object to object), since the point density depends on the surface characterizes of a region. Smoothly textured surfaces (e.g., homogenous roof segment) lead to sparsely distributed 3D points. Therefore spatially separated roof segments of a single building were found to be delineated as multiple buildings. In this case, an object was correctly classified as building but it was topologically incorrect, since a single building was detected as multiple buildings.

Case 3: Spatially separated individual buildings delineated as single building –the noisy points in the 3D point cloud were found to act as a bridge between the physically separated buildings. This is the case where the object was correctly classified as building but topologically incorrect.

Case 4: Building delineated with some portion of tree – this is the case, when a portion of tree was detected as a planar segment and spatially connected to a building was recognized as a part of the building. Therefore the outline of the detected building was considered as topologically incorrect.

Case 5: Chain of buildings detected as single building –where physically connected buildings (a block of buildings) were delineated as a single building. In this case, the detected buildings were considered as topologically correct, since, with our data-driven approach, the physically connected buildings cannot be identified as separate buildings.

Case 6: Chain of buildings detected as multiple individual buildings – this happened due to two reasons, 1) significant change in the height of the connected buildings and 2) sparse 3D point cloud as described in case 2. The detected buildings were considered as topologically incorrect, since the physically connected buildings were considered as single building as per the above case.

Case 7: Non-building detected as building –the above ground non-building objects with horizontal or slanted planar segments were detected as a building. For example, in our test data, non-building objects such as large vehicles were detected as building.

***Case 8: Buildings not detected* –** some of the buildings were partially detected and some not detected because of sparse or missing 3D points due to the radiometric characteristics of the surface.

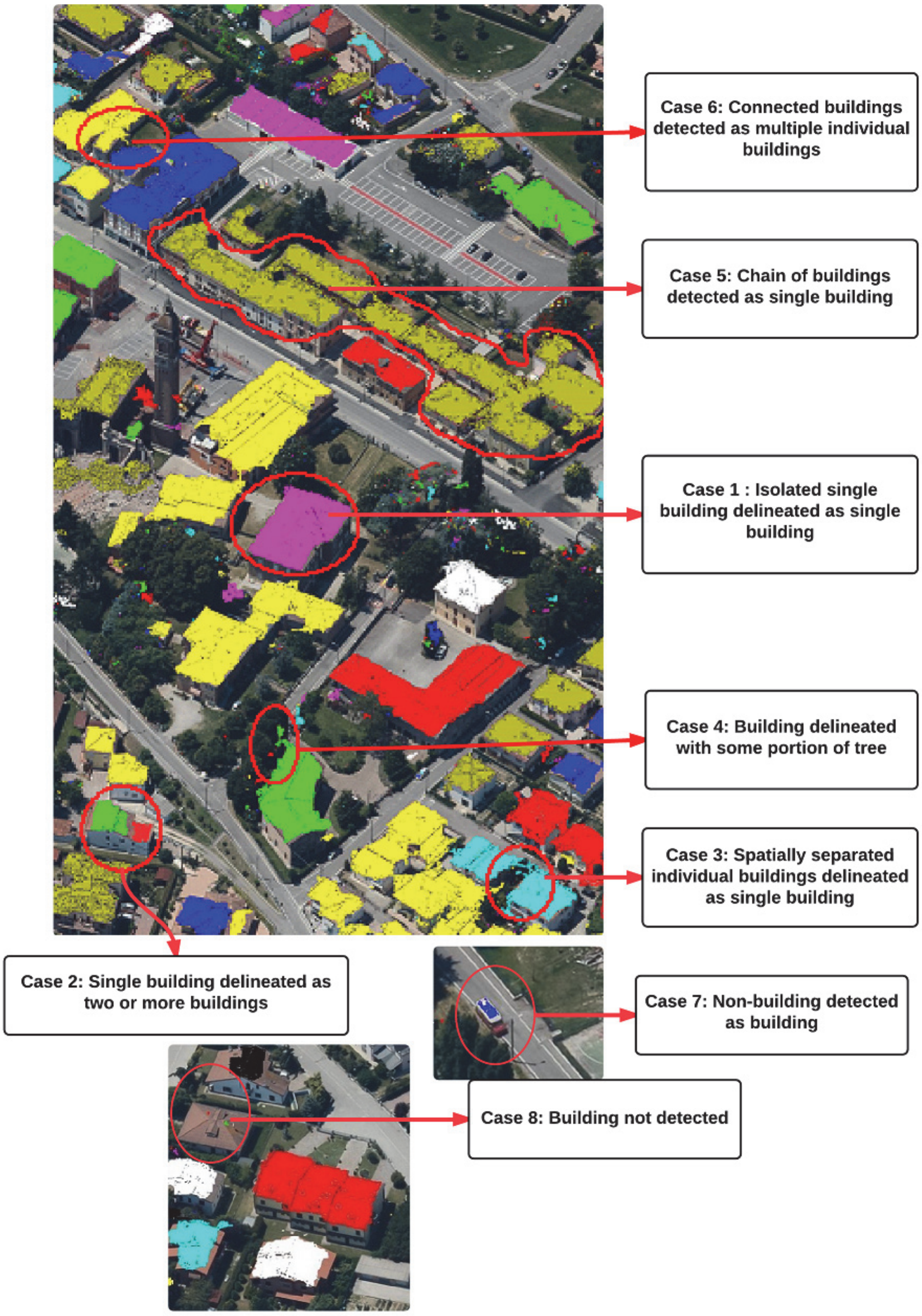


Figure 2-6 Subsets of aerial image projected with 3D points of the delineated buildings, images © BLOM Italy

Table 2-1 Number of buildings that falls under each category

Category of the delineated buildings	***Number of buildings***
Case 1 : Isolated single building delineated as single building	*83*
Case 2: Single building delineated as two or more buildings	*2*
Case 3: Spatially separated individual buildings delineated as single building	*9*
Case 4: Building delineated with some portion of tree	*8*
Case 5: Chain of buildings detected as single building	*76*
Case 6: Chain of buildings detected as multiple individual buildings	*13*
Case 7: Non-building detected as building	*16*
Case 8: Buildings not detected	*7*

The efficiency of the building delineation method was assessed based on how well the buildings were detected among other above ground objects in the scene. We adopted commonly used evaluation metrics such as precision, recall and accuracy to assess the efficiency of the building delineation process. The precision, recall and accuracy can be defined as, ***Precision*** = TP/ (TP+FP); ***Recall*** = TP/ (TP+FN); ***Accuracy =*** (TP)/ (TP+ FP+FN) where TP, FP and FN are:

True positive (TP) –Buildings detected as buildings: the buildings that comes under the category of case 1 to 6 were considered as true positives.

False positive (FP) – Non-building detected as buildings: The objects that are categorized under case 7 are the non-building objects detected as buildings which were considered as false positives.

False negative (FN) – Buildings not detected as buildings: The buildings that are categorized under case 8 were the non-detected buildings in the scene which were considered as false negatives.

Table 2-2 Results of building detection process

Evaluation metrics for building detection process	
TP = 191 *(case 1+case 2+case 3+case 4+case 5+case6),* ***FN = 7*** *(case 8),* ***FP = 16*** *(case 7)*	
Precision	*92 %*
Recall	*96%*
Accuracy	*89%*

The recall measure indicates that 96% of the buildings in the scene were detected using the roof-based building delineation approach (Table 2-2). The precision and accuracy measures (Table 2-2) were slightly affected by the detection of non-building objects as buildings. This is because here we consider only the height and surface normal orientation of the segment as features for identifying the roof segments of the buildings. So any above ground object with horizontal or slanted segments will be recognized as roof segment. For example the upper part of large vehicle which is horizontally oriented and lies above the ground was detected as a building as depicted in Figure 2-6. This implies that geometric constraints alone cannot be sufficient to avoid the false positive detections. Among the detected buildings, only 83 % of the buildings (159 out of 191 buildings) that were categorized under case 1 and case 5 were considered as correctly delineated buildings, i.e. topologically error free. The visual assessment was carried out using the aerial images projected with 3D points of the delineated buildings as depicted in Figure 2-6. The boundary of remaining 17 % of the detected buildings deviated from the actual boundary due to many reasons such as inclusion of tree segments, disconnected roof segments due to missing 3D points, etc., (refer to cases – 2,3,4 and 6 in Table 2-1). Overall the roof-based building delineation approach seems to be working well for image-based 3D point cloud where most of the building delineations were near the actual boundary of the building which is sufficient for the damage assessment scenario.

2.4.2 Gap detection and classification on element level – dataset: images from UAV

All the three processes such as building delineation, gap detection and gap classification were carried out in sequence for this data set and the results have been reported in the following sections.

2.4.2.1 Data description

A small region around the church –'*Church of Saint Paul*' in Mirabello captured by an UAV was considered for the gap detection and classification process. The images of the selected region were captured by a VTOL (vertical take-off and landing) UAV from various heights, positions and views (oblique and nadir). An average GSD of the captured images is around 1 cm. A 3D point cloud of the scene was generated from 152 images with an average point density of 650 points per m^2. The selected region contained six buildings and among them only one was damaged.

2.4.2.2 Results of gap detection process

All the six buildings in the scene were delineated using the building delineation approach. The merging criteria's thresholds (c.f. 3.1) that were defined for the building delineation from the point cloud of data set1 (average point density of 15 points per m^2) seem to work well when applied to much higher density point cloud of data set 2 (average point density of 650 points per m^2). Among the six buildings, four are isolated (case 1) and two are spatially connected (case 5) which were correctly delineated using the building delineation approach. An example for the delineation of a single building from the 3D point cloud of UAV images is given in Figure 2-7.

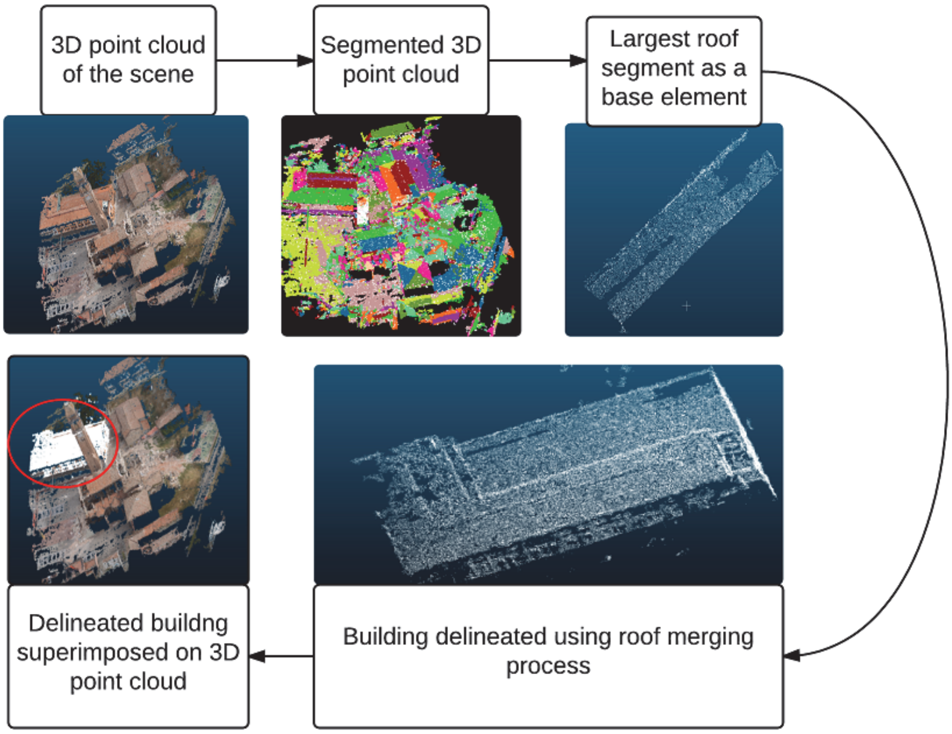


Figure 2-7 An example for delineation of a single building from the 3D point cloud, images © Aibotix Italy

Each delineated building is comprised of a number of 3D point segments. The gap detection process was carried out individually for every 3D segment of the building. A 3D segment was chosen and voxelized. The choice of voxel size is indeed a compromise between a loss of information (large voxels compared to given point density) and oversampling caused by voxels that are too small, leading to artificial gaps. In fact, the voxel size must be large enough to contain actual information from the point cloud (if available) and small enough for the given application: which gap size is of interest? In this work, we used 0.4 m voxel which was found to be sufficient for damage detection process, being aware of the fact that the original point density is much larger. The gap voxels were identified based on the gap detection procedure described in section 0 (e.g., refer to Figure 2-8). In total 7 regions were detected as gaps. The gap detection process is a straightforward approach that can accurately detect gap voxels within a segment.

2.4.2.3 Gap classification

The 3D segments that contain the gaps were selected for the gap classification process. As a first step, each 3D gap segment was mapped with an image using the image selection criteria described in section 0. This is to extract the radiometric descriptors around the gap region for the gap classification. After selecting an appropriate image for each gap segment, the corresponding gap regions in the image were delineated using the procedure described in section 2.3.3.2. An example for gap region delineation process on image is portrayed in Figure 2-8.

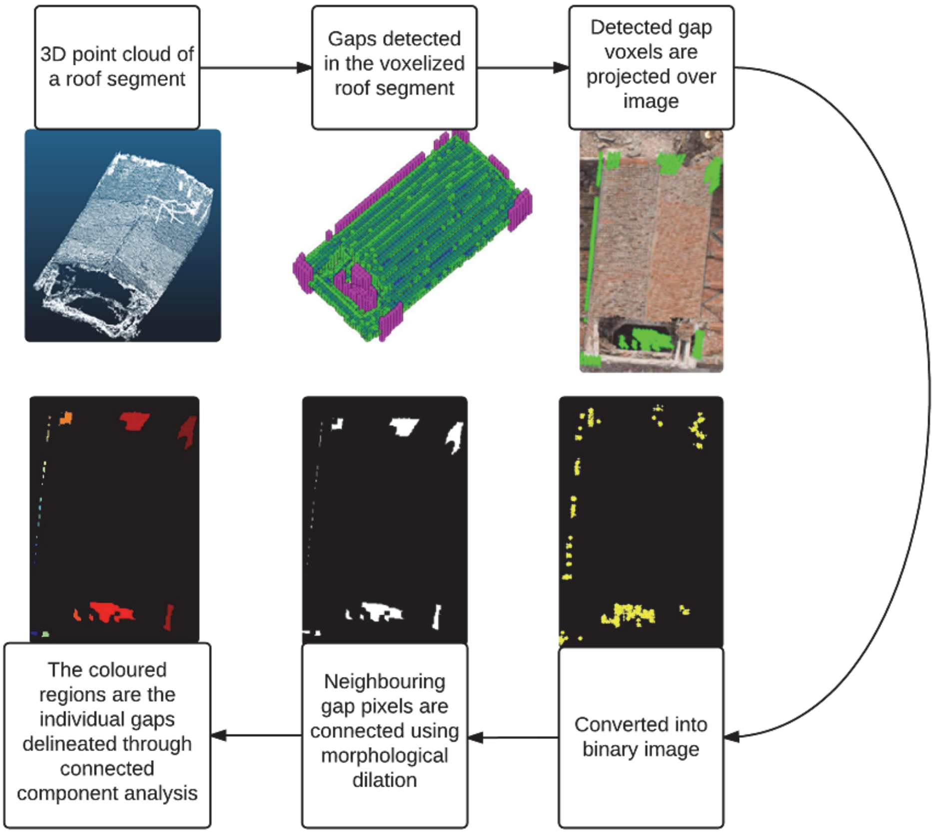


Figure 2-8 Delineated gap regions in the image

The next step after gap region delineation is the gap classification. The gap classification process requires information about the presence and absence of damage evidences around the gap region. Therefore, the detection of damaged regions in the image is a prerequisite for the gap classification process, which was carried out as follows.

2.4.2.4 Classification of entire scene into damaged and undamaged regions

Here two kinds of features, Gabor wavelets and HoG, were independently analysed for the damage detection process. A supervised learning approach was used to develop a model-based on the above mentioned features by providing the relevant training samples to perform the damage detection process. The significance of each kind of features for damage classification when trained with supervised algorithms was evaluated in two experiments:

a) Training and testing with samples belonging to the same study site.
b) Training with samples collected from various places, and testing with the samples from our study site.

The above two experiments were conducted in order to analyse the generalization capability of the classification. This is because it has been reported that often the supervised models developed based on samples from one study site produce weak predictions when tested with samples from another unseen site, due to many reasons such as the quality of the data, limitations of the selected features and limitations of the learning model itself (Foody et al., 2003).

2.4.2.5 Results of the gap classification process

i) Gabor wavelet features for damaged region detection

The Gabor feature images were obtained by applying the filter banks of Gabor wavelets over the selected image. The Gabor filter banks were obtained using the procedure as described in Haghighat et al. (2013). From the feature images it was inferred that Gabor features have the potential to differentiate between the regions based on their surface pattern, irrespective of their intensity. For example consider Figure 2-9, where the RGB image depicts the scene that contains three different types of building roofs, annotated as A, B and C, and a damaged region annotated as D. The Gabor features images that are depicted in Figure 2-9 are the feature images corresponding to different frequencies and orientations. The roof segments A, B and C, were clearly differentiated by those Gabor wavelet features. For example, for the roof A the Gabor feature 3 showed a strong signal, whereas B and C got highlighted by other Gabor features as depicted in Figure 2-9. In all the feature images the damaged region annotated as D in Figure 2-9 was found to show similar characteristics. This is because the man-made objects have a dominant orientation; hence the respective feature corresponding to that orientation shows a clear peak. Conversely, the damaged region has a gradient orientation in many directions; hence they possess similar characteristics in most of the feature images corresponding to different orientations. The visual assessment indicates that Gabor features have the potential to differentiate the objects in the scene based on their dominant frequency and orientation characteristics.

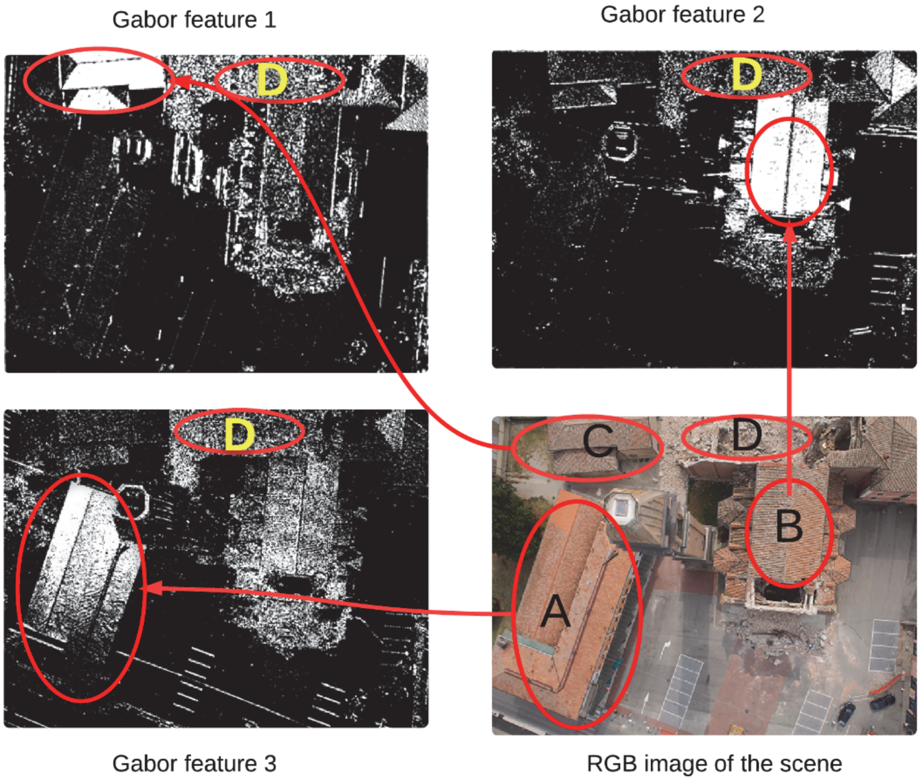


Figure 2-9 Subset of the UAV image and its Gabor feature images – the marked regions A, B and C are the roofs of different buildings and D is the region of debris

The Gabor filter banks consist of 40 filters (5 frequencies at 8 orientations), thus each pixel in the image was represented by a feature vector of size 1x40. To select the training samples for building a supervised model, the image was divided into rectangular grids with a grid size of 40x40 pixels. Each rectangular grid was assigned with a feature vector by finding the median of features corresponding to all pixels within the rectangular grid. Manually 466 grids corresponding to damaged (267 grids) and non-damaged (199 grids) regions were selected to frame training samples for constructing a supervised model. Using the training samples, supervised learning models were developed based on the two learning approaches, SVM and Random Forests. The performance of the learning models was analysed using a 10 fold cross validation process. The results are provided in Table 2-3.

Table 2-3 Results of Gabor feature-based supervised models for damaged region detection

	Confusion matrix (D= Damaged ND =Non-damaged)				***Precision (%)***	***Recall (%)***	***Accuracy (%)***
Random Forest		*Predicted Class*			*95.6*	*97.4*	*95.9*
	Actual Class		*D*	*ND*			
		D	*260*	*007*			
		ND	*012*	*187*			
SVM		*Predicted Class*			*89.3*	*97.4*	*91.8*
	Actual Class		*D*	*ND*			
		D	*260*	*007*			
		ND	*031*	*168*			

Both SVM and RF produced same recall measure which indicates that they both correctly classified 97% of the damaged region. The precision produced by SVM was relatively low (89%) when compared to RF (95%). This shows that SVM tends to produce a larger number of false positive predictions than RF for this application.

ii) HoG features for damaged region detection

For damage detection using HoG features the selected image was also divided into rectangular grids with a grid size of 40x40 pixels. The HoG pattern was found to be different for damaged and undamaged regions. For damaged regions, the histogram was spread over many directions, as depicted in Figure 2-10, whereas HoG of undamaged regions possessed gradient orientations in few directions, which can be considered the dominant gradient orientation of that region (e.g., Figure 2-10). From the visual assessment, the HoG feature was found to be capable of differentiating between the damaged and undamaged regions based on its local gradient orientation pattern. The significance of the HoG feature for damage pattern recognition was evaluated using SVM and RF learning algorithms based on the same procedure as followed for the Gabor features. The obtained results for HoG feature are presented in Table 2.4, which are quite similar to the results obtained with the Gabor features.

Table 2-4 Results of the HoG feature-based supervised models for damaged region detection

<table>
<tr><th></th><th colspan="4">Confusion matrix
(D= Damaged
ND =Non-damaged)</th><th>Precision (%)</th><th>Recall (%)</th><th>Accuracy (%)</th></tr>
<tr><td rowspan="4">Random Forest</td><td></td><td colspan="3">Predicted Class</td><td rowspan="4">94.1</td><td rowspan="4">95.5</td><td rowspan="4">94.0</td></tr>
<tr><td rowspan="3">Actual Class</td><td></td><td>D</td><td>ND</td></tr>
<tr><td>D</td><td>255</td><td>012</td></tr>
<tr><td>ND</td><td>016</td><td>183</td></tr>
<tr><td rowspan="4">SVM</td><td></td><td colspan="3">Predicted Class</td><td rowspan="4">91.9</td><td rowspan="4">92.9</td><td rowspan="4">91.2</td></tr>
<tr><td rowspan="3">Actual Class</td><td></td><td>D</td><td>ND</td></tr>
<tr><td>D</td><td>248</td><td>019</td></tr>
<tr><td>ND</td><td>022</td><td>177</td></tr>
</table>

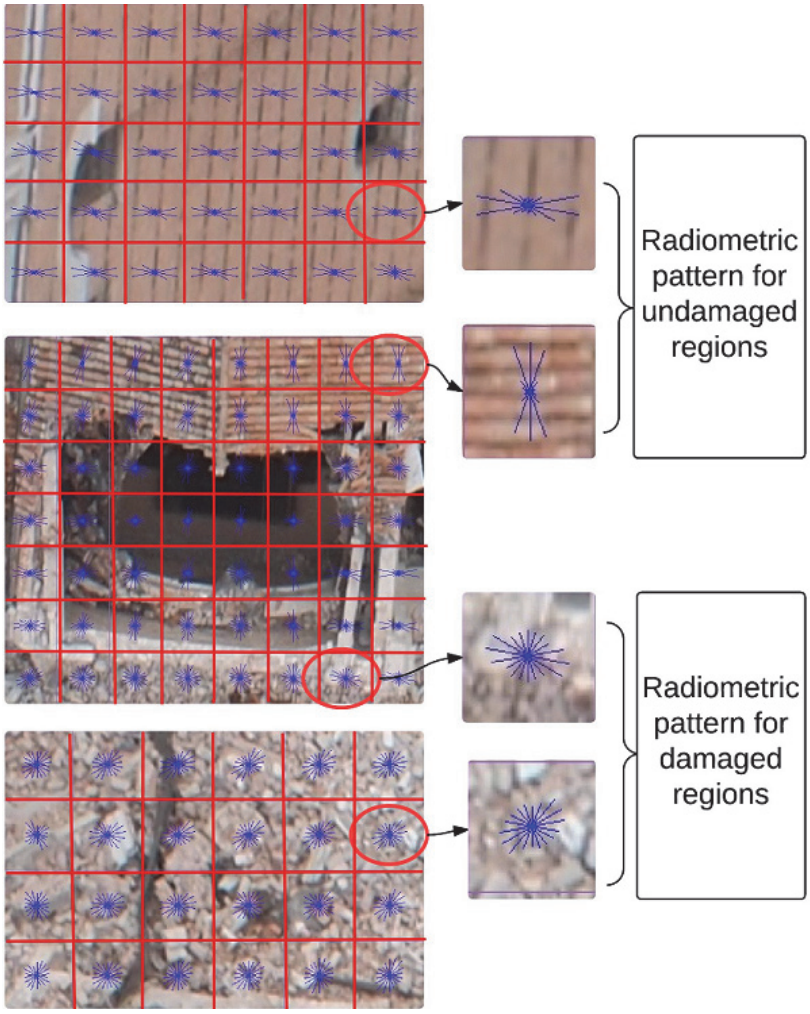


Figure 2-10 HoG pattern for damaged and undamaged regions

Both Gabor and HoG features were found to work well for our test site when the supervised model was trained and tested with samples from the same study site. To assess the generalization capability of the supervised approach, a supervised model was developed using the training samples with wide variety of building and damage types, collected from various geographic locations and tested that model with the samples from our study site. The data used for preparing the training samples were the street view images of the damaged buildings collected from various countries after the disaster events. The training set consists of 344 positive samples (damaged region) and 276 negative samples (non-damaged region). The results of the supervised models when tested with the data from the unseen site are presented in Table 2-5. The performances of these supervised models were significantly lower than the models trained and tested with samples from the same site. The maximum recall value (84%) was obtained by the model developed using HoG feature with RF when tested with data from unseen site (Table 2-5). But the same model produced relatively low precision of 71%. All the supervised models developed by unique combinations of features (Gabor and HoG) and learning algorithms (RF and SVM) produced similar results with higher recall value than precision when trained and tested with samples from different locations. This indicates that the accuracy of those models was mostly affected by the high false positive rates

Table 2-5 Results of the supervised models for damaged region detection when trained using samples from various locations and tested with samples from our study site (values in brackets indicate change to the earlier experiment where the classifier got trained in the actual area)

<table>
<tr><th>Features</th><th>Learn-ing algorit-hm</th><th>Confusion matrix (D= Damaged ND =Non-damaged)</th><th>Precisi-on (%)</th><th>Recall (%)</th><th>Accur-acy (%)</th></tr>
<tr><td rowspan="2">Gabor</td><td>RF</td><td><table><tr><td></td><td colspan="3">Predicted Class</td></tr><tr><td rowspan="3">Actual Class</td><td></td><td>D</td><td>ND</td></tr><tr><td>D</td><td>193</td><td>064</td></tr><tr><td>ND</td><td>096</td><td>103</td></tr></table></td><td>66.7 (29.0)</td><td>75.1 (22.3)</td><td>64.9 (31.0)</td></tr>
<tr><td>SVM</td><td><table><tr><td></td><td colspan="3">Predicted Class</td></tr><tr><td rowspan="3">Actual Class</td><td></td><td>D</td><td>ND</td></tr><tr><td>D</td><td>205</td><td>052</td></tr><tr><td>ND</td><td>072</td><td>127</td></tr></table></td><td>74.1 (15.2)</td><td>79.7 (17.7)</td><td>72.8 (19.0)</td></tr>
<tr><td rowspan="2">HoG</td><td>RF</td><td><table><tr><td></td><td colspan="3">Predicted Class</td></tr><tr><td rowspan="3">Actual Class</td><td></td><td>D</td><td>ND</td></tr><tr><td>D</td><td>218</td><td>039</td></tr><tr><td>ND</td><td>086</td><td>113</td></tr></table></td><td>71.7 (22.4)</td><td>84.8 (10.7)</td><td>72.58 (21.4)</td></tr>
<tr><td>SVM</td><td><table><tr><td></td><td colspan="3">Predicted Class</td></tr><tr><td rowspan="3">Actual Class</td><td></td><td>D</td><td>ND</td></tr><tr><td>D</td><td>210</td><td>047</td></tr><tr><td>ND</td><td>082</td><td>117</td></tr></table></td><td>71.9 (20.0)</td><td>81.7 (11.2)</td><td>71.7 (19.5)</td></tr>
</table>

Overall, the supervised model developed based on Gabor wavelet features with RF (Table 2-3) produced better results than other developed supervised models for our test site and this classification was used in the further gap classification process. The image classified into damaged and undamaged regions using the selected supervised model is shown in Figure 2-11.

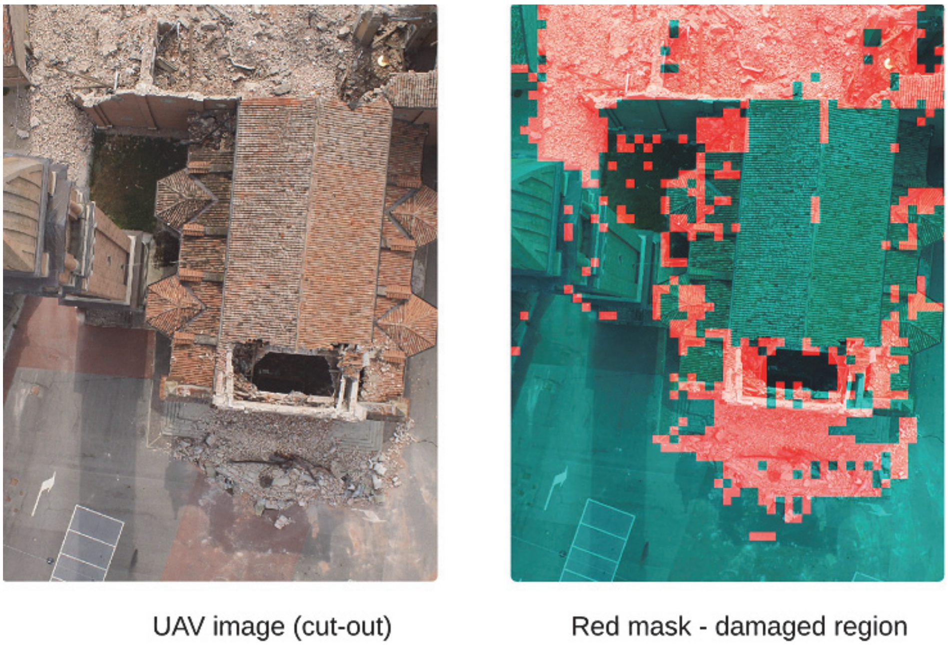


Figure 2-11 Image classified into damaged and undamaged region using the selected supervised model

2.4.2.6 Classification of delineated gap regions:

The delineated gap regions in the image were classified into gaps due to damage, gaps due to openings, and gaps due to surface characteristics issues, based on the procedure described in section 2.3.3.2. In total 7 gaps were considered. Among them only two were due to the damage, and both show evidence of damage (broken rubble pieces) around them. The damaged regions around those gaps were correctly detected as damaged regions by the selected supervised model for damage detection. The other gaps due to openings in the architectural design and image matching failure issues were also correctly classified. A sample of results for each gap category are given below.

Category 1: Gap due to damage

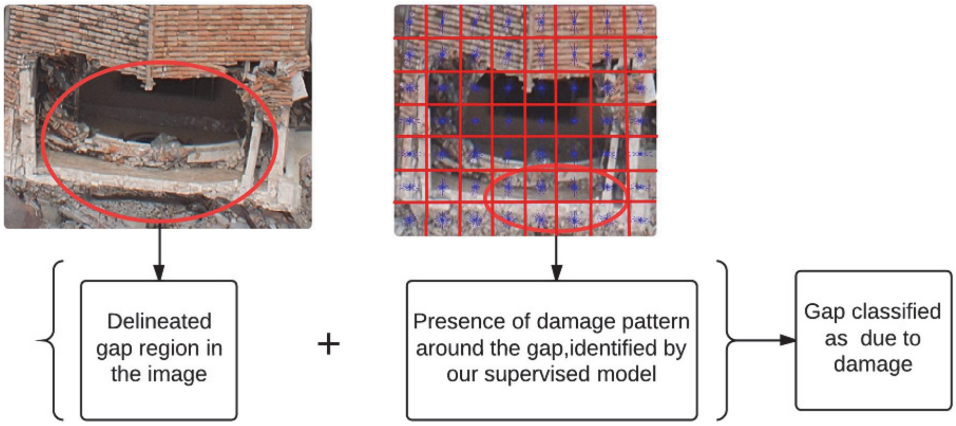


Figure 2-12 An example for gap due to damage

Category 2: Gap due to an opening in the architectural design

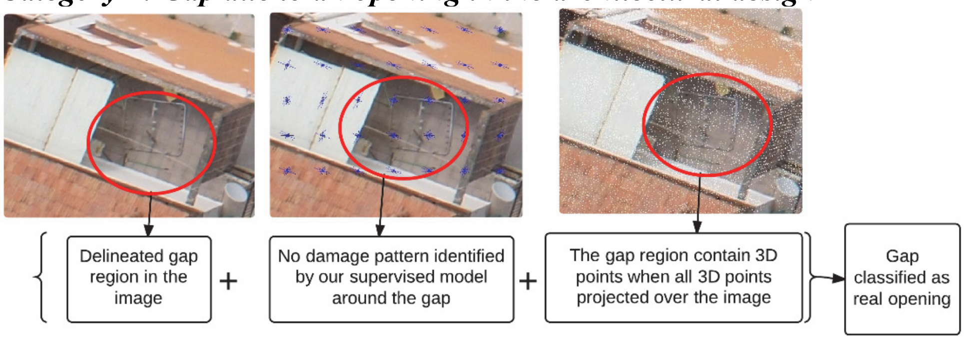


Figure 2-13 An example for a gap due to a natural opening

Category 3: Gap due to surface characteristics issues

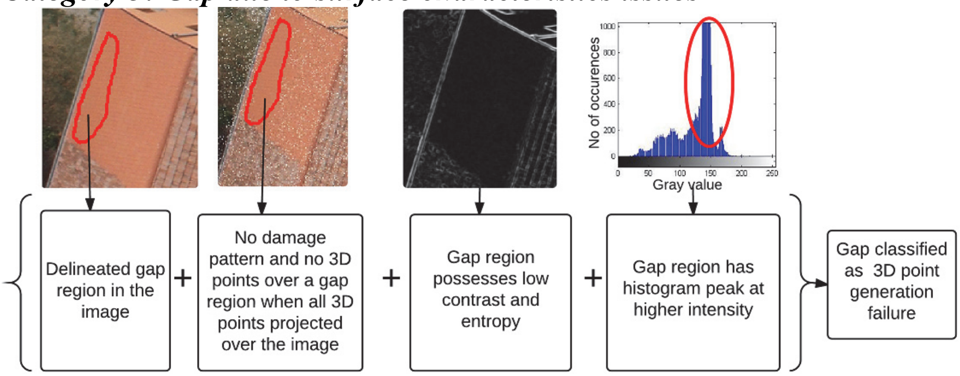


Figure 2-14 An example for gap due to surface characteristics issue

2.4.3 Gap detection and classification on element level – dataset 3: Nunspeet UAV

So far, we have no access to UAV images showing real earthquake-damaged buildings, other than what was shown in 4.2. In order to further assess the accuracy of detecting gaps in the point cloud that are due to a failure of 3D point generation or to real openings, the gap detection and classification methods were tested on another data over a non-damaged urban area in the municipality of Nunspeet in The Netherlands (Hinsbergh et al., 2013). The 3D point cloud of the selected urban region was generated with an average point density of 250 points per m^2 from the images captured by an UAV with an average GSD of 1.5 cm. We observed a number of gaps in this 3D point cloud due to two reasons; 1) presence of texture-less objects such as glass windows in the gable roof of the buildings and 2) occlusion (visible in less than 3 images) as depicted in Figure 2-15. In total 11 buildings were considered for the gap classification process, in which we found 14 gaps in the roof segments of the building after automatically excluding the gaps due to occlusion. Among 14 gaps, 5 were classified as texture-less surface and 9 were classified errors due to image matching failures, even though they all corresponding to the same object such as glass windows. This is because, the glass objects showed different kind of reflection when it was viewed from different positions as depicted in Figure 2-16. From the results, it was inferred that the presence of significant gap within the 3D segment of a point cloud will be detected by our method.

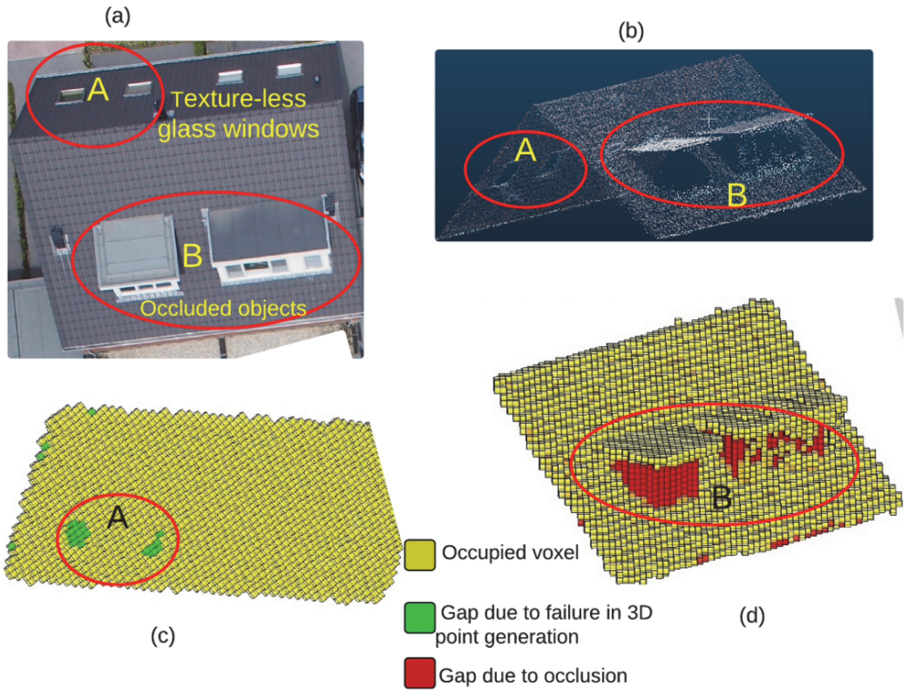


Figure 2-15 (a) an image cut out shows a building with texture-less and occluded objects, (b) Subset of 3D point cloud corresponding to that building, (c) and (d) Voxelized 3D point cloud segments with highlighted gap voxels, images © Dutch Kadaster

Figure 2-16 Image subsets showing the variation in window glass reflection for images captured from different positions, images © Dutch Kadaster

2.5 Overall discussion and conclusion

In this chapter a framework was developed to delineate buildings from an image-based 3D point cloud, and to identify the broken elements of the delineated buildings that let to gaps in the 3D point cloud.

We adopted a roof-based building delineation approach. Often, tree segments that display some degree of planarity were misclassified as roofs, and thus wrongly detected as buildings. The spectral information from images was used to differentiate between the roofs and tree components, which removed most of the wrongly classified tree segments. Our approach detected 96% (Table 2-2) of the buildings in the selected area. The remaining buildings were not detected due to the missing or sparse 3D points. This has been identified as the major limitation of image-based point clouds that show high object-to-object point density variation. Because of this low point density issue, some of the detected buildings were not accurately delineated, and a single building was wrongly delineated as multiple buildings. Also, the presence of noisy points that were generated due to mismatches, and wrong detection of tree segments as roofs that are overlapped with the roof segments, affected the delineation accuracy of the detected buildings. In total, only 83% of the detected buildings were delineated correctly. However, our objective is to perform gap detection and classification at a segment level, where we require only the segments of the buildings, and where the building delineation accuracy has a limited impact. However, for future comprehensive damage assessment at a per-building level, the realization of the geometric characteristics of the building is important. This requires at least an approximate boundary of the building. The boundary inaccuracy problem in building delineation arises due to the limitations of roof merging criteria, where only the spatial relationship between the segments is considered for the merging of roof segments of a single building. Generally, those possess similar surface characteristics (i.e. colour and texture). Therefore, defining merging criteria based on the combination of spatial connectivity and radiometric homogeneity (same colour and texture) between the segments could overcome the above limitations, and thereby improve the building delineation accuracy. One of the objectives of this research was to identify the gaps in the 3D point cloud that are caused by the damage to building elements. Those gaps were identified based on their surrounding damage patterns. Since the HoG and Gabor wavelets features performed well in previously published pattern recognition and objection classification studies, we anticipated that they could identify and classify the radiometric patterns related to damages, such as spalling and debris. We found that both HoG and Gabor wavelets has the potential to recognize the damage patterns, finding around 95% of the damaged regions when used in a supervised learning approach. However, we observed a problem of generalization with the damage detection process based on the supervised approach, as the supervised model developed from training samples based on different geographic locations produced weak results when tested with the samples from an unseen site (c.f. Table 2-5). In case of time critical applications such as damage assessment, the development of a local supervised model, suitable to the specific study site, and by collecting the training samples from the same study site, is not practical. Hence, derivation of a generalized relationship between the features and damage pattern is important. The generalization capability of the supervised model may improve when the training sample size is increased.

The regions for damage analysis were defined by dividing the image into rectangular grids of constant size. In this case, often single objects get split into several regions. Therefore, the actual orientation characteristics of the object cannot be completely captured. This may lead to an incorrect prediction about the damage state of the region. Hence, deriving features at an appropriate scale is important. Therefore, instead of defining the regions based on a gridding approach, image segmentation can be used to identify the unique regions within the image for damage analysis. We anticipate that image segmentation based on Gabor features can identify those unique regions in the image, and that performing damage analysis on those regions will improve the accuracy of the damage classification (damaged region identification).

Moreover, gaps due to damage were detected based on the surrounding damage pattern such as spalling or debris. In case such evidence is missing, damage-related gaps cannot be accurately classified. The domain semantics could be of help here. For example, according to the general nomenclature of the building, any openings in the building, such as a balcony, will be located only in a specific, predictable position. Also, real openings based on an architectural design will typically be uniform in geometry. Therefore, the gaps in inappropriate locations, and with irregular geometry, can be classified as gaps due to damage, after eliminating other possible reasons.

From the analysis it has been inferred that the incorporation of domain knowledge into our approaches is required to make them robust. However, the major challenge is that the conceptualization and formalization of these kinds of domain specific semantics into an operational format (e.g., a set of rules), which is generally referred to as ontology, e.g., Belgiu et al. (2014). However, considerable further research is required to develop such ontology-based robust approaches for building delineation and damage detection processes.

The results and analysis of this study clearly indicate the importance of image-based radiometric features in both the building delineation and damage detection processes. Thereby it implies that an image-based 3D point cloud, which provides both radiometric and 3D geometric features, is a suitable source for structural damage assessment.

Overall, the proposed framework for building delineation from image-based 3D point clouds, and mapping of damaged elements of the delineated building that are related to the gaps in the 3D point cloud, was found to be successful. The extension of this framework by developing methods for mapping other kind of damages such as inclined elements, cracks, and spalling, along every element of the building, would lead to accomplishing our desired objective of comprehensive and detailed building damage assessment.

2.6 References of Chapter 2

Alsadik, B., Gerke, M., Vosselman, G., 2014. Visibility analysis of point cloud in close range photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-5, pp. 9-16.

Arivazhagan, S., Ganesan, L., Priyal, S.P., 2006. Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters,* 27 (16), pp. 1976-1982.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing,* 33 (part B4/1), pp. 111-118.

Belgiu, M., Tomljenovic, I., Lampoltshammer, T.J., Blaschke, T., Höfle, B., 2014. Ontology-based classification of building types detected from airborne laser scanning data. *Remote Sensing,* 6 (2), pp. 1347-1366.

Breiman, L., 2001. Random forests. *Machine learning,* 45 (1), pp. 5-32.

Chen, W., Ao, T., Li, H., Yang, W., Li, X., 2011. An improved approach of information extraction for earthquake-damaged buildings using high-resolution imagery. *Journal of Earthquake and Tsunami,* 05 (04), pp. 389-399.
Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 92, pp. 79-97.
Dell'Acqua, F., Gamba, P., 2012. Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE,* 100 (10), pp. 2876-2890.
Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.
Ehrlich, D., Guo, H.D., Molch, K., Ma, J.W., Pesaresi, M., 2009. Identifying damage caused by the 2008 Wenchuan earthquake from VHR remote sensing data. *International Journal of Digital Earth,* 2 (4), pp. 309-326.
Foody, G.M., Boyd, D.S., Cutler, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment,* 85 (4), pp. 463-474.
Gerke, M., Kerle, N., 2011a. Automatic structural seismic damage assessment with airborne oblique pictometry imagery. *Photogrammetric Engineering & Remote Sensing,* 77 (9), pp. 885-898.
Gerke, M., Kerle, N., 2011b. Graph matching in 3D space for structural seismic damage assessment. *lEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 6-13 Nov, 2011, pp. 204-211.
Haghighat, M., Zonouz, S., Abdel-Mottaleb, M., 2013. Identification using encrypted biometrics. Computer Analysis of Images and Patterns, Springer, 440-448.
Hinsbergh, W.H.M.v., Rijsdijk, M.J., Witteveen, W., 2013. UAS for cadastral applications: testing suitability for boundary identification in urban areas. *GIM International,* 27, pp. 20-25.
Hussain, E., Ural, S., Kim, K., Fu, C.S., Shan, J., 2011. Building extraction and rubble mapping for city port-au-prince post-2010 earthquake with geoeye-1 imagery and lidar data. *Photogrammetric Engineering & Remote Sensing,* 77 (10), pp. 1011-1023.
Kerle, N., 2010. Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it? *International Journal of Applied Earth Observation and Geoinformation,* 12 (6), pp. 466-476.
Kerle, N., Hoffman, R.R., 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural hazards and earth system sciences (NHESS) : open access,* 13, pp. 97-113.
Khoshelham, K., Oude Elberink, S., Sudan, X., 2013. Segment-based classification of damaged building roofs in aerial laser scanning data. *IEEE Geoscience and Remote Sensing Letters,* 10 (5), pp. 1258-1262.

Kobayashi, T., Hidaka, A., Kurita, T., 2008. Selection of histograms of oriented gradients features for pedestrian detection. Neural Information Processing, Springer, Berlin Heidelberg, 598-607.

Lafarge, F., Mallet, C., 2012. Creating large-scale city models from 3d-point clouds: A robust approach with hybrid representation. *International Journal of Computer Vision,* 99 (1), pp. 69-85.

Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S., Wiechert, A., 2010. Point clouds: Lidar versus 3D vision. *Photogrammetric Engineering & Remote Sensing,* 76 (10), pp. 1123-1134.

Li, L., Zhang, B., Wu, Y., 2012. Fusing spectral and texture information for collapsed buildings detection in airborne image. *IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 22-27 Jul, 2012, pp. 186-189.

Ma, J., Qin, S., 2012. Automatic depicting algorithm of earthquake collapsed buildings with airborne high resolution image. *IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 22-27 Jul, 2012, pp. 939-942.

Minetto, R., Thome, N., Cord, M., Leite, N.J., Stolfi, J., 2013. T-HOG: An effective gradient-based descriptor for single line text regions. *Pattern Recognition,* 46 (3), pp. 1078-1090.

Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics,* 9 (1), pp. 62-66.

Oude Elberink, S., Shoko, M., Fathi, S.A., Rutzinger, M., 2011. Detection of collapsed buildings by classifying segmented airborne laser scanner data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII, , pp. 307-312.

Plank, S., 2014. Rapid damage assessment by means of multi-temporal SAR — A comprehensive review and outlook to Sentinel-1. *Remote Sensing,* 6 (6), pp. 4870-4906.

Rabbani, T., van den Heuvel, F., Vosselman, G., 2006. Segmentation of point clouds using smoothness constraint. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 36 (5), pp. 248-253.

Radhika, S., Tamura, Y., Matsui, M., 2012. Use of post-storm images for automated tornado-borne debris path identification using texture-wavelet analysis. *Journal of Wind Engineering and Industrial Aerodynamics,* 107, pp. 202-213.

Rapidlasso, 2013, Homepage lastools. http://lastools.org, (Accessed 15 Jun, 2014).

Rathje, E.M., Crawford, M., Woo, K., Neuenschwander, A., 2005. Damage patterns from satellite images of the 2003 Bam, Iran, earthquake. *Earthquake spectra,* 21 (S1), pp. 295-307.

Rehor, M., Vögtle , T., 2008. Improvement of building damage detection and classification based on laser scanning data by integrating spectral information. *International Archive of Photogrammetry and Remote Sensing (IAPRS),* 37 (B7), pp. 1599-1605.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J.D., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing,* 93, pp. 256-271.

Ruiz, L., Fdez-Sarría, A., Recio, J., 2004. Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study. *20th ISPRS Congress*, 35 (part B), pp. 1109-1114.

Rupnik, E., Nex, F., Remondino, F., 2014. Oblique multi-camera systems - orientation and dense matching issues. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 50 (3/W1), pp. 107-114.

Salas, Y.S., Bermudez, D.V., Peña, A.M.L., Gomez, D.G., Gevers, T., 2012. Improving HOG with image segmentation: Application to human detection. Advanced Concepts for Intelligent Vision Systems, Springer, 178-189.

Schölkopf, B., Smola, A.J., 2001. Learning with kernels: support vector machines, regularization, optimization, and beyond (adaptive computation and machine learning). MIT Press, Cambridge, MA.

Shen, Y., Wu, L., Wang, Z., 2010. Identification of inclined buildings from aerial LIDAR data for disaster management. *18th IEEE International Conference on Geoinformatics*, Beijing, China, pp. 1-5.

Sun, S., Salvaggio, C., 2013. Aerial 3D building detection and modeling from airborne LiDAR point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6 (3), pp. 1440-1449.

Trinder, J., Salah, M., 2012. Aerial images and lidar data fusion for disaster change detection. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-4, , pp. 227-232.

Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., Mehl, H., 2007. Satellite image analysis for disaster and crisis-management support. *IEEE Transactions on Geoscience and Remote Sensing,* 45 (6), pp. 1520-1528.

Vosselman, G., 2012. Automated planimetric quality control in high accuracy airborne laser scanning surveys. *ISPRS Journal of Photogrammetry and Remote Sensing,* 74, pp. 90-100.

Wikipedia, 2015, http://simple.wikipedia.org/wiki/2011_Christchurch_earthquake#mediaviewer/File:Piko_building_corner_Barbadoes_Kilmore.jpg, (Accessed 06 Feb, 2015).

Xiong, X., Adan, A., Akinci, B., Huber, D., 2013. Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction,* 31, pp. 325-337.

Yamazaki, F., Matsuoka, M., 2007. Remote sensing technologies in post-disaster damage assessment. *Journal of Earthquake and Tsunami,* 1 (03), pp. 193-210.

3 Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach*

[*] This chapter is based on the article:
Vetrivel, A., Gerke, M., Kerle, N. and Vosselman, G. (2016) Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach. In: Remote Sensing : open access, 8 (2016)3article no. 231, 22 p.

Abstract

Automatic post-disaster mapping of building damage using remote sensing images is an important and time critical element of disaster management. The characteristics of remote sensing images available immediately after the disaster are not certain, since they may vary in terms of capturing platform, sensor-view, image scale and scene complexity. Therefore, a generalised method for damage detection that is invariant to the mentioned image characteristics is desirable. This study aims to develop a method to perform the grid-level damage classification of remote sensing images by detecting the damages corresponding to debris, rubble piles and heavy spalling within a defined grid, regardless of the aforementioned image characteristics. The Visual-Bag-of-Words (BoW) is one of the most widely used and proven frameworks for image classification in the field of computer vision. The framework adopts a kind of feature representation strategy that has been shown to be more efficient for image classification regardless of the scale and clutter than conventional global feature representations. In this study supervised models using various radiometric descriptors (histogram of gradient orientations (HoG) and Gabor wavelets) and classifiers (SVM, Random Forests and Adaboost) were developed for damage classification based on both BoW and conventional global feature representations, and tested with four datasets. Those vary according to the aforementioned image characteristics. The BoW framework outperformed conventional global feature representation approaches in all scenarios (i.e. for all combinations of feature descriptors, classifiers and datasets), and produced an average accuracy of approximately 90%. Particularly encouraging was an accuracy improvement by 14% (from 77% to 91%) produced by BoW over global representation for the most complex dataset, which was used to test the generalization capability.

3.1 Introduction

Rapid damage assessment after a disaster event such as an earthquake is critical for efficient response and recovery actions. Direct manual field inspection is labour intensive, time consuming and cannot assess the damages in inaccessible areas. Remote sensing technology is the most predominant and early source to provide data for performing such assessments, either manually or using automated image analysis procedures (Dell'Acqua and Gamba, 2012; Dong and Shan, 2013). Various kinds of remote sensing data such as optical, synthetic aperture radar (SAR) and LiDAR are being used for the damage assessment process (Dong and Shan, 2013). However, optical data are often preferred as they are relatively easy to interpret (Dong and Shan, 2013). Moreover, optical remote sensing provides very high resolution images ranging from decimetre to centimetre scale through various platforms such as satellites, manned aircrafts and unmanned aerial vehicles (UAVs) (Adams et al., 2013; Gerke and Kerle, 2011a; Li et al., 2010a). This allows performing comprehensive damage assessment through identifying different levels of damage evidences, ranging from complete collapse to cracks on the building roof or façades, by choosing images at appropriate scales. Particularly oblique airborne images are recognized as the most suitable source, as they facilitate the damage assessment on both roofs and lateral elements (Fernandez Galarreta et al., 2015a; Gerke, 2011). For example, even extensive building damage such as inter-story collapse or pancake collapse can be identified reliably only with oblique view images, while conventional nadir views at best provide damage proxies such as external debris (Dong and Shan, 2013; Fernandez Galarreta et al., 2015a; Kerle and Hoffman, 2013a). Although current remote sensing yields images at a vast range of views and scales, automatic recognition of even heavy damages to buildings is still challenging (Dong and Shan, 2013). This is due to various reasons, such as the complexity of the scene, uncertain characteristics of damage patterns and the varying scale problem in oblique view images.

Generally, the regions corresponding to heavy damage are determined through the identification of damage patterns corresponding to rubbles piles, debris and spalling in an image region (refer to Figure 3-1Figure 2-1) (Kerle and Hoffman, 2013a). Those damage evidences have a specific meaning and play a major role in damage classification. For example, the presence of significant amounts of debris/ rubble piles around the building is the strong indication of (partial) building collapse. Spalling is an indicator of minor damage or partially broken structural elements. The recognition process of those damage patterns can be performed by analysing features extracted either in pixel or region level (Dong and Shan, 2013; Kaya et al., 2010; Miura et al., 2013). However, the pixel level analysis is not meaningful for very high spatial resolution images, particularly in the context of damage assessment, as the evidences are identified based on the characteristics of their radiometric distribution pattern, which can be captured more precisely at a region level. However, in region-level classification the critical part is to define the region that is appropriate to identify the specific damage patterns. Generally, image regions are obtained either through a gridding approach or though image segmentation (Ma et al., 2014). The most simple, efficient and straightforward strategy is the gridding approach, where the image is split into uniform rectangular cells. However, the regions derived from gridding are often cluttered, as they may comprise different kinds of objects. For example, a single cell may contain trees, building elements, cars, road sections, debris, etc. Moreover, oblique images are more complex compared to nadir images, since they also capture façades that frequently comprise various elements, such as windows, balconies, staircases, etc. They generally also look more cluttered than nadir images containing largely roofs, and reveal façade information only at image border, depending on the lens opening angle. It is quite challenging to identify the damage patterns in such kind of cluttered region. This can be alleviated by using a segmentation approach, which segments the damaged portions and other objects in the scene as separate regions. However, the selection of appropriate features and a segmentation algorithm that is suitable for a given damaged and cluttered environment is a challenging problem, one that requires substantial semantic analysis. Apart from clutter, the regions obtained from oblique images vary in scale. Nevertheless, the identification of damage patterns regardless of image scale is an important prerequisite in damage assessment. For example, damages at a building level such as inter-story collapse can be captured better at coarser scales (e.g., 100 x100 pixel region in an image of decimetre scale), while minor damages such as spalling at a building element level require finer scales (e.g., 100x100 pixel region in an image of centimetre scale). Therefore, a robust method is required to recognize the damage pattern in a defined region irrespective of the scale and clutter. This is an analogue of the human visual pattern recognition system which is extremely proficient in identifying the damage patterns regardless of the scale and complexity of the scene.

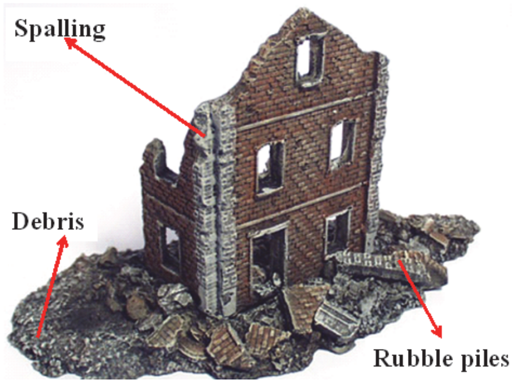


Figure 3-1 An example for debris, rubble piles and spalling –Source: http://www.combatgroupdynamix.com/Diorama/WargameSeries/Accessories/Buildings/WargameAccessoriesTallBuilding.htm

In the field of computer vision, various methods have been reported for pattern recognition tasks in various applications, such as object categorization, face recognition and natural scene classification (Jin and Ruan, 2009; Yang et al., 2009; Zhang et al., 2007). These methods are mostly based on supervised learning approaches, which work well for conventional image classification applications. However, the overall performance of the learning approach completely depends on the discriminative power of the image descriptors (features) considered for the classification (Huang et al., 2014). Generally, images are described through either global (e.g., textures) or local features, like point descriptors such as Scale Invariant Feature Transform (SIFT) (Oliva and Torralba, 2006; Zhang et al., 2007). However, most global features are very sensitive to scale and clutter (Carneiro and Jepson, 2009). In contrast, the local descriptors are robust to clutter but cannot capture the global characteristics of the image (Lou et al., 2014; Zuo and Zhang, 2011). An alternate feature representation strategy, such as Visual-Bag-of-Words (BoW), captures the global characteristics of the image through encoding a set of local features, which makes them robust to scale and clutter (Ferraz et al., 2014a; Lu and Wang, 2015). For example, in texture-based classification, the global texture pattern of the image is captured by the frequencies of the co-occurrence of the local texture patterns. This kind of feature representation outperforms the conventional global feature representation approaches in image classification (Zhuang et al., 2013). Apart from general image classification, the Bag-of-Words framework has been demonstrated as a potential approach in many image-based domain specific applications including image retrieval (Wu et al., 2009), human action and facial expression recognition (Li et al., 2010b; Wang and Mori, 2009), image quality assessment (Ye and Doermann, 2012) and medical image annotation (Bouslimi et al., 2013). Conceptually, thus, the BoW approach seems to be appropriate for identifying the damaged regions in airborne oblique images, which generally look cluttered and vary in scale. Pattern recognition methods including BoW are based on a supervised learning approach that attempts to learn the underlying relationship between the image-derived features and the pattern of a specific category, in this case the damage pattern. Therefore, apart from a feature representation strategy, the choice of features that best discriminate the damaged and non-damaged regions is also a key element. Numerous studies reported that textures are the most influential feature for damage pattern recognition, as the damaged regions tend to show uneven and peculiar texture patterns in contrast to non-damaged regions (Ma and Qin, 2012b; Radhika et al., 2012; Yamazaki and Matsuoka, 2007). Many damage classification studies used statistical textures such as grey level co-occurrence matrix (GLCM)-based features for the damage pattern recognition (Miura et al., 2013; Reinartz et al., 2013; Sui et al., 2014a). However, other texture measures such as wavelets have been recognized as superior to GLCM in many pattern recognition problems, including land cover classification (Stavrakoudis et al., 2011). Particularly for region-level pattern classification problems, descriptors such as Histogram of Gradient Orientation (HoG), Gabor wavelets, SIFT and Speeded Up Robust Features (SURF) have led to good results (Conde et al., 2013; Khan et al., 2011; Lin et al., 2011). All these features describe the pattern of the given region in a unique way, based on the magnitude of gradient along various orientations and scales. Vetrivel et al. (2015a) demonstrated the potential of HoG and Gabor features to classify the damaged regions in very high resolution UAV images. However, they found limitations with the conventional global representation of HoG and Gabor features, especially with respect to generalization. So far, however, to our knowledge no work exists which combines the named features in a BoW fashion for damage mapping.

The objective of this research work is thus to develop a robust method based on the BoW approach that is suitable especially (but not only) for oblique images to identify the damage patterns related to rubble piles, debris and spalling, regardless of the scale and the clutter of the defined region in an image. Following the above argumentation, a grid-based region definition is pursued. The robustness of the developed method based on this BoW approach is analysed by comparing the performance of various learning algorithms and image descriptors (Gabor and HoG) under both the conventional and the BoW approach. Also, the generalization capability of the developed method is analysed, by testing it on a variety of images corresponding to various scales, camera views, capturing platforms and levels of scene complexity.

3.2 Methods

For the identification of damaged regions in an image, as a preparation step we provide reference data. That is, the given image is split into MxN regions which are termed image patches. The image patches are manually labelled as damaged if any kind of damage pattern related to debris, spalling and rubble piles is observed in them. The automatic detection of those damage patterns within the patches is carried out using two different feature representation approaches: global and BoW representation. The feature descriptors and learning algorithms considered for both the global and BoW-based damage classification process are described in the respective sub-sections.

3.2.1 Damage classification based on global representation of features

This process includes two steps: 1) extraction of image descriptors that provide the global description of the given image patch, and 2) classification of the given image patch as damaged or non-damaged, based on the extracted feature descriptors using a supervised learning algorithm.

3.2.1.1 Extraction of feature descriptors

The HoG and Gabor wavelets-based feature descriptors are considered for the global feature representation-based damage classification process.

a) Histogram of Gradient Orientation (HoG)

The standard approach is used to extract the HoG features (e.g., Dalal and Triggs (2005)), where the given image patch is split into a number of overlapping blocks, and histograms of gradient orientation derived for each block are concatenated to form a feature vector. This gives the global representation of the image patch.

Procedure:

1. Derive gradient magnitude and its orientation for each pixel in the image patch.
2. Split the gradient image into AxB cells.

3. Again split the gradient image into a number of overlapping blocks, where each block contains CxD cells with 50% of overlapping cells between the blocks.
4. Define the bin size for the histogram of gradient orientation, where each bin corresponds to a specific orientation (the bin size remains fixed for all experiments later).
5. For each cell, compute the histogram of gradient orientation by adding the magnitude of the gradient to its corresponding orientation bin. Therefore, the size of the feature description of each cell is equal to the number of bins.
6. Concatenate the histograms of gradient orientation of all cells within each block to get the block level description. Normalize the histograms magnitude within the block to compensate for the local illumination variations (Déniz et al., 2011).
7. Concatenate all block level descriptors to form the global descriptor of the patch.

b) Gabor wavelets descriptors

The Gabor wavelets descriptors are obtained by convolving the image with a set of Gabor wavelet filters. These filters are derived by appropriate rotation and scaling of the mother Gabor wavelet function, where each filter is tuned to capture the pixel information at a specific orientation and frequency. The detailed procedure for Gabor wavelets filter generation can be found in Arivazhagan et al. (2006). After obtaining the Gabor filter responses for each pixel in the image patch, the region-level Gabor wavelet descriptor is represented by the histogram of magnitude of filter responses for all combinations of orientations and frequencies (cf. Yi and Su (2014)). This histogram is computed for three consecutive pyramid levels of image patches, in order to capture the variation across scales, in addition to the variation across frequencies and orientations. The procedure used for extracting the global Gabor feature descriptors for an image patch is described below.

Procedure:

1. Generate IxJ number of 2D Gabor wavelet filters, where I and J are the number of frequencies and number of orientations used to generate the Gabor wavelet filters, respectively.
2. Convolve the image patch with the generated filter banks, which results in IxJ number of feature images.
3. Normalize each feature image using l^2 normalization.

4. Compute the histogram of Gabor filter responses, where each histogram bin corresponds to a specific frequency and orientation. Therefore, the number of histogram bins is equal to IxJ, which is the size of the final feature vector.
5. Also, extract the Gabor wavelet features for the other two pyramid levels of the image patch, by subsampling it to ½ and ¼ of the image patch size.
6. Feature vectors derived at different scales are concatenated to form the final feature vector. Therefore, this final feature vector will comprise features extracted at multiple scales, multiple frequencies and multiple orientations.

3.2.1.2 Damage classification using the derived global feature descriptors

Supervised learning approaches are adopted to classify the given image patch as damaged or non-damaged, based on the global feature descriptors. Three state-of-the-art and widely used supervised learning algorithms, Support Vector Machines (SVM) (Schölkopf and Smola, 2002), Random Forests (RF) (Breiman, 2001) and Adaboost (Rätsch et al., 2001), are considered for the damage classification process. These learning algorithms belong to the families of different learning paradigms, which learn the underlying relationship between the input features and the output label in a unique way. Three different learning paradigms are considered in order to analyse whether the considered feature descriptors are independent of the supervised algorithm, i.e. how the classification task is solved independently of the applied learning strategy. Also, each learning algorithm has a number of tuneable parameters, referred to as hyper-parameters, which have a significant impact on the performance of the learning model (Bergstra and Bengio, 2012). Therefore, the hyper-parameters are tuned for the best model by searching the parameter space using the grid space search approach (Bergstra et al., 2011). This approach constructs a number of learning models for different settings of the hyper-parameters, using the training set. The performance of each model is assessed using a cross validation procedure. The best performing model is selected as the final model with tuned hyper-parameters, and then evaluated using the testing set.

3.2.2 Damage detection using Visual-Bag-of-Words

The standard BoW framework is adopted for the damage classification process. The BoW framework comprises different components, such as feature point detection, feature descriptors, visual word dictionary, and a classifier. The algorithms used for each component and the overall procedure are described below.

Overall, the BoW-based damage classification process is carried out in two stages: 1) construction of visual word dictionary, and 2) representation of the image in terms of BoW (histogram of visual words), and training the classifier based on them.

Stage 1:

a) ***Feature point detection***

The basic idea behind this step is that an image can be described by a small number of significant pixels (salient points). For example, pixels corresponding to edges and corners contain the most significant information compared to pixels of homogenous regions. Salient point descriptors that are invariant to scale and orientation are most appropriate to build an image classification model that is robust to scale and rotation. Numerous such salient point detection methods are available, with SIFT and SURF commonly being used in the BoW context (Lou et al., 2014). In this study, SURF was used since it is faster than SIFT and its descriptor is suitable to be used as the feature in the BoW framework, as discussed in the following sub-section. A description of the SURF points detection process can be found in Bay et al. (2006).

b) Feature extraction

The purpose of this step is to extract the local feature descriptor for each salient point in the given image patch. The feature descriptors HoG and Gabor wavelets that are used in the global representation-based damage classification are also considered here for the local description of salient points in the BoW-based damage classification. This allows to compare the potential of BoW and global feature representation irrespective of the features. In the BoW approach the SURF descriptor is additionally used to describe the salient points. This is because SURF is a well-proven point descriptor (local descriptor), and widely used in BoW-based image classification processes (Tsai, 2012). Furthermore, SURF descriptors are based on wavelet responses, which also describe the image region in terms of textures, similar to HOG and Gabor feature descriptors. Therefore, the three feature descriptors HoG, Gabor wavelets and SURF are used independently to describe each salient point in the given image patch for the BoW-based damage classification. The local pattern description for each salient point is derived by considering a local neighbourhood of PxQ pixels around the salient point. The same procedure as described in section 2.1.1 is followed to extract the Gabor and HoG features. The standard procedure is used to extract the SURF feature descriptor (cf. Bay et al. (2006)).

c) ***Visual words dictionary construction***
The feature descriptors of salient points from all image patches (regardless of their class) are concatenated into a single feature vector. Numerous feature encoding methods have been reported for visual word dictionary construction (Peng et al., 2013). We adopted the most commonly used iterative k-means clustering algorithm (Tsai, 2012). The obtained feature vector is clustered into k clusters using the iterative k-means clustering (Vattani, 2011). Each cluster centre is considered as the visual word, and the cluster centres are collectively referred to as visual word dictionary.

Stage 2:

a) ***Image description based on visual words***
To represent the given image patch in terms of BoW (histogram of visual words), the salient points in the image patch are detected and feature descriptors are obtained for each point. The detected points in the image are assigned to their closest visual word in the dictionary. Subsequently, the frequency of occurrence of the visual words in the image is represented as a single histogram, which is referred to as the BoW representation of the image, which will be fed into the classifier in the next step.

b) ***Classification of visual words using machine learning algorithms***
Again, the three learning algorithms SVM, RF and Adaboost are used as classifier for classifying the damage and non-damaged image patches based on BoW. The procedure as described in section 2.1.2 is followed to develop the supervised learning models based on the BoW features.

The overall workflow of BoW-based damage classification process is depicted in Figure 3-2.

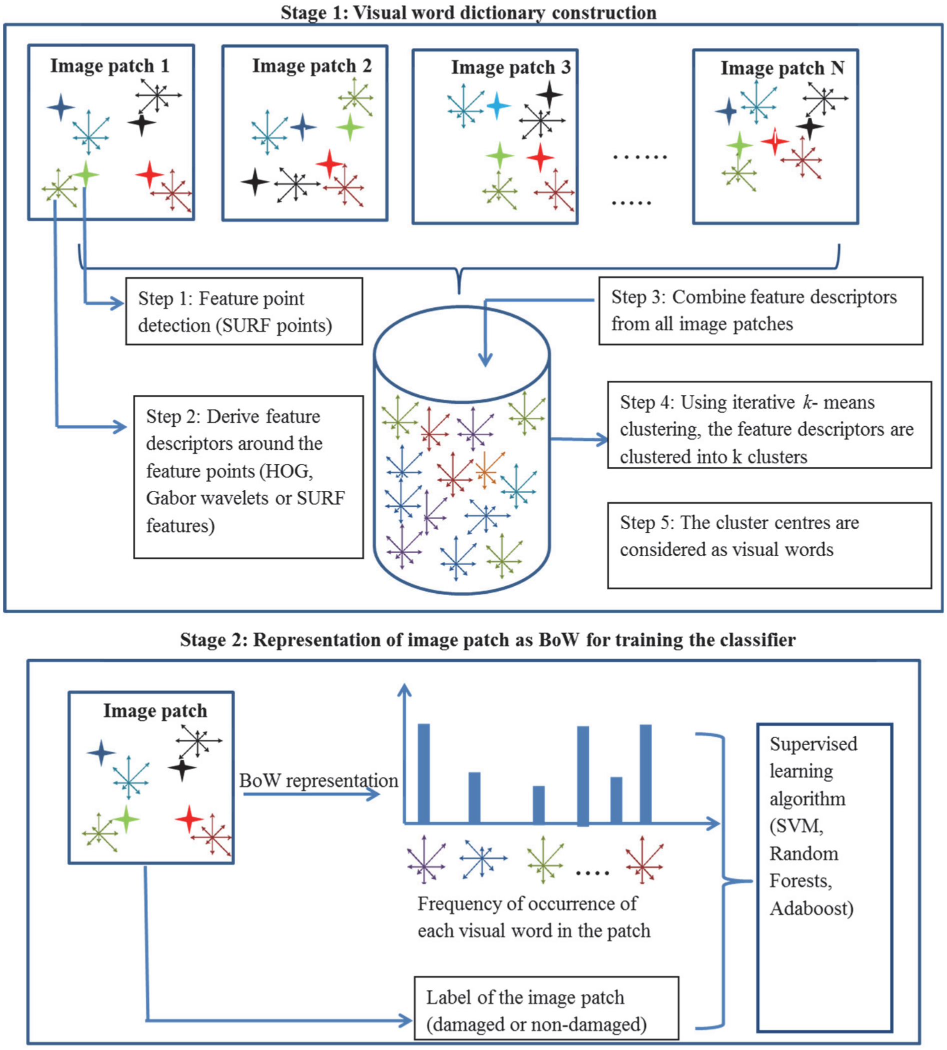


Figure 3-2 Overall process of the BoW-based damage classification

3.3 Experiments and Results

The damage classification method was evaluated using four different data sets, with each differing in its image characteristics such as scale, camera view, capturing platform and scene complexity. Each data set was independently analysed for the damage classification process based on the three feature descriptors HoG, Gabor wavelets and SURF. The performances of HoG and Gabor wavelets for damage classification were analysed by representing them in both a conventional and BoW framework. Also the potential of the SURF descriptor was analysed for damage classification by representing it in a BoW framework and comparing it with BoW based Gabor and HoG.

Three supervised learning algorithms, SVM, RF and Adaboost, were used for analysing the performance of the feature descriptors. Therefore, each data set was tested with different combinations of feature descriptors and supervised learning algorithms, as depicted in Figure 3-3.

The conducted experiments for the damage classification process include a number of algorithms, and each algorithm was associated with a number of parameters. The values used for the parameters of the algorithm are shown in Table 3-1. The hyper-parameters considered for tuning the learning algorithms (cf. 2.1.2) are described in Table 3-2.

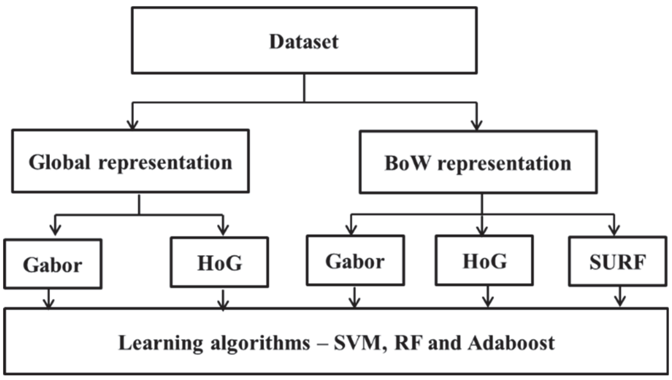


Figure 3-3 Combinations of feature descriptors and learning algorithms tested for each dataset

Table 3-1 Definition of parameters associated with each algorithm/method used in the experiment

Algorithm/ method	Parameter values	Description	Reference
Image patch generation	M=100; N=100	To generate 100x100 image patches	Section 2
HoG procedure	A=25; B=25	Cell size AxB - 25x25pixels	Section 2.1.1
	C= 4; D=4	Block size CxD– 4x4 cells	
	bin size =9	Bin size of histogram of the gradient orientations	
Gabor wavelet descriptor	I=5; J=8	I, J are the number of frequencies and orientations respectively to generate the Gabor wavelet filters	Section 2.1.1
Feature extraction	P=10; Q=10	10 x10 local neighbourhood is considered for deriving descriptor for each salient point	Section 2.2
Visual word dictionary construction	*k* = 500	k value for *k*-means clustering	Section 2.2
Supervised model for damage classification	10-fold cross validation	Cross-validation to identify the optimal hyper-parameters for a learning model based on the grid search approach	Section 2.1.2
	The dataset is split into 70% and 30% for training and testing, respectively	Training set is used to train the model and also for cross-validation for tuning the hyper-parameters. Testing set is used for evaluating the trained model.	

Table 3-2 Definition of grid search space for tuning the hyper-parameters of the classifiers

Supervised classifier	Hyper-parameter	Grid search space	Description
SVM	C	0.001 to 100, step size – multiples of 10	Regularization parameter which has a significant effect on the generalization performance of the classifier.
	Kernel	Linear, radial basis function (RBF) and histogram intersection	The function used to compute the kernel matrix for classification.
	gamma	0.0001 to 1.0, step size – multiples of 10	Regularization parameter used in RBF kernel (Gaussian kernel function) which has significant impact in the performance of the kernel.
RF	N_estimators	3 to 20, step size 2	Number of trees in the forest.
	Max_depth	1 to 5, step size 1	Maximum depth of the tree.
	Min_samples_split	1 to 4, step size 1	Minimum number of samples required to split a node.
	Min_samples_leaf	1 to 3, step size 1	Minimum number of samples required in newly created leaf after the split.
Adaboost	N_estimators	100 to 1000, step size 100	The maximum number of estimators that can be used to build the ensemble learning model.
	Learning rate	0.01 to 0.1, step size 0.01	Regularization parameter that scales the contribution of each weak estimator.

3.3.1 Dataset 1: UAV images

UAV images captured over two different areas were considered: 1) a small region around a church ('Church of Saint Paul') in Mirabello, Italy, damaged by a 2012 earthquake; 2) a small region around a partly demolished industrial facility in Gronau, Germany. Both regions possess similar characteristics, and they contain only a few buildings that are largely disconnected. One building in each region was partially collapsed and produced a significant quantity of debris and rubble piles (cf. top left image in Figure 3-4– UAV image-subset of the Mirabello church). The UAV images were captured at different heights, positions and views (nadir and oblique) with a spatial resolution of 1-2 cm. The images of both regions corresponding to various orientations and heights were split into

100x100 pixel rectangular image patches for framing of the training and testing datasets for the damage classification process. The patches are labelled as damage if at least 25% of their area represents damage evidences (debris/rubbles or spalling). It is difficult to describe the characteristics of damage and undamaged samples. Hence, for each dataset, the samples of damaged and undamaged image patches are portrayed to provide better insight (refer to Figure 3-4 to Figure 3-6). Since the image resolution is very high, the defined rectangular patches cover only a small region (approximately 1 m^2) and, therefore, most of them contain only either damage evidences or single homogenous object, i.e. the defined regions are mostly uncluttered; refer to the image training samples in Figure 3-4. In total 966 samples (482 damaged, 484 non-damaged) each of size 100x100 pixels were considered. The dataset was constructed by selecting the specific samples across different regions within the scene that highly vary in their characteristics to avoid a large number of repetitive samples. The damage classification was performed for this dataset based on different combinations of feature descriptors and learning algorithm as described above, and the results are reported in Table 3-3.

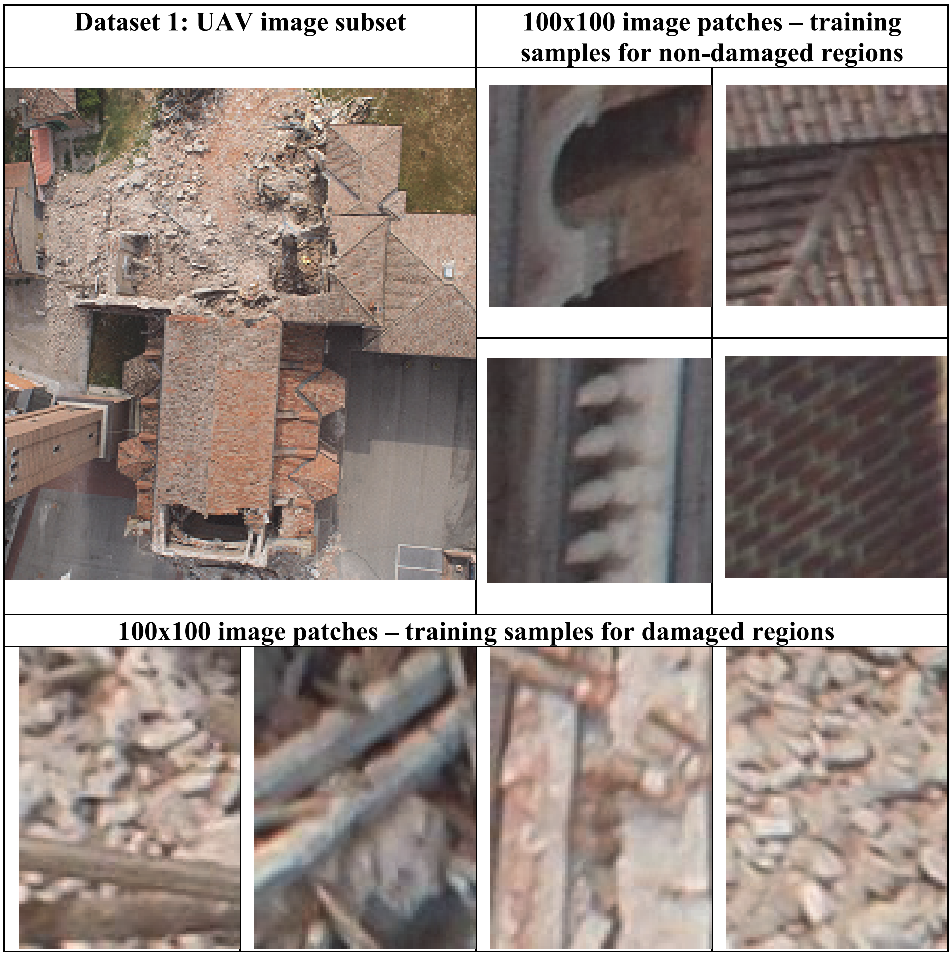


Figure 3-4 Samples of image patches in dataset 1- UAV images

Table 3-3 Performance of feature descriptors when associated with different learning algorithms for dataset 1 comprising patches from the UAV images (training samples = #676, testing samples = # 290) - bold numbers indicate best performance per indicator

Data -set1	**SVM**			**RF**			**Adaboost**		
	Precisn	**Reca -ll**	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**
Gabor	0.91	0.87	0.90	**0.99**	0.92	0.95	0.81	0.76	0.79
HoG	0.87	0.86	0.86	0.94	0.93	0.93	0.71	0.67	0.69
BoW-Gabor	0.96	0.93	0.95	**0.99**	**0.98**	**0.98**	**0.96**	0.71	0.83
BoW-HOG	**0.98**	**0.97**	**0.98**	0.97	0.95	0.95	0.95	**0.87**	**0.90**
BoW-SURF	0.97	0.92	0.94	0.90	0.88	0.90	0.80	0.81	0.81

3.3.2 Dataset 2: Oblique view manned aircraft images

The airborne oblique images (Pictometry) with a Ground Sampling Distance (GSD) between 10 cm (foreground) and 16cm (background) captured over Port-au- Prince after the 2010 Haiti earthquake were considered. The images cover almost entire city, and contain numerous buildings ranging from simple to complex. Most of the buildings are densely clustered in such a way that it is difficult to differentiate each building even visually from the images. Numerous buildings are partially covered with densely leafed tall trees, adding to the clutter of the scene. A significant number of buildings is damaged, ranging from complete/partial collapse to heavy spalling on the intact elements of the building (cf. Figure 3-5). The images are split into 100x100 pixel images/patches to frame the training and testing datasets for the damage classification process. The defined image patches are highly cluttered as they cover a large area (at least 10 m^2) and comprise different kinds of objects, such as trees, building elements, cars, road sections and debris (cf. Figure 3-5). The dataset was constructed by selecting the specific samples across different regions within the city that highly vary in their characteristics. Again, the selection of samples was driven by the idea to cover different damage characteristics rather than piling up redundant information. In total 1256 samples (698 damaged, 558 non-damaged) were selected and tested for the damage classification based on the developed approach. The patches cover an area of approximately 13,000 m². The results are reported in Table 3-4.

Figure 3-5 Samples of image patches in dataset 2, images © Pictometry

Table 3-4 Performance of feature descriptors when associated with different learning algorithms for dataset 2 comprising patches from Pictometry images (training samples = #879, testing samples = # 377).

Dataset2	SVM			RF			Adaboost		
	Precision	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**
Gabor	0.81	0.76	0.79	0.82	0.76	0.79	0.78	0.61	0.72
HoG	0.78	0.61	0.72	0.67	0.61	0.66	0.62	0.58	0.63
BoW-Gabor	0.89	0.86	0.88	**0.88**	**0.88**	**0.88**	**0.80**	**0.79**	0.79
BoW-HOG	**0.93**	**0.89**	**0.91**	0.85	0.83	0.84	**0.80**	0.69	0.75
BoW-SURF	0.91	**0.89**	0.90	0.84	0.82	0.83	**0.80**	0.78	**0.80**

3.3.3 Dataset 3: Street view images

Street view close-range images of damaged buildings captured by hand-held cameras after earthquakes in different geographic locations were used. These images were collected from two sources: 1) Governmental organization: the German Federal Agency for Technical Relief, THW; (http://www.thw.de), and 2) the internet (various websites). The collected images vary in scale; however, the actual scale is unknown. Therefore, the 100x100 pixel patches generated from those images may cover small areas (e.g., an element of the building) or large areas (e.g., entire or major portion of the building). The collected images contain buildings with various kinds of damages, such as complete collapse, partial collapse, inter-story collapse and heavy spalling. In total 887 samples (384 damaged, 503 non-damaged) were considered to construct and evaluate the supervised model for the damage classification. Samples image patches used for the training and testing of the supervised model are depicted in Figure 3-6. The results are reported in Table 3.5.

.

Dataset 3: Scale varying street view images

100x100 image regions – training samples for damaged regions	**100x100 image regions – training samples for non-damaged regions**

Figure 3-6 Samples of image patches in dataset 3- street view images

Table 3-5 Performance of feature descriptors when associated with different learning algorithms for dataset 3 comprising patches from street view images (training samples = #620, testing samples = # 267).

Dataset3	**SVM**			**RF**			**Adaboost**		
	Precision	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**
Gabor	0.89	0.85	0.87	0.88	0.80	0.85	0.91	0.85	0.89
HoG	0.95	0.74	0.86	0.94	0.81	0.89	0.84	**0.84**	0.85
BoW-Gabor	0.99	**0.91**	0.95	0.92	0.77	0.86	0.92	0.72	0.84
BoW-HOG	**1.0**	0.93	**0.96**	**0.98**	**0.94**	**0.96**	**0.98**	0.82	**0.90**
BoW-SURF	0.99	0.89	0.94	**0.98**	0.82	0.91	**0.98**	0.77	0.89

3.3.4 Dataset 4: Datasets 1, 2 and 3 are combined

The samples from datasets 1, 2 and 3, which vary in scale, camera view, and platform and scene complexity, were combined into a single dataset in order to assess the generalization capability of the damage classification methods. In total 3109 samples (1564 damaged, 1545 non-damaged; subsequently termed COM_{3109}) were used to develop and test the supervised models for damage classification. The results are reported in Table 3-6. For visual analysis, an UAV image of dataset 1 and a Pictometry image of dataset 2 were classified for the damage detection using the best performing model (BoW-Gabor with SVM). The classified images are depicted in Figure 3-7. The classification is quite accurate, showing only very few false positives and false negatives, which are also highlighted in the classified images (cf. Figure 3-7). The false positives and negatives are the examples where our assumption fails: i.e. a surface with unusual radiometric characteristics is assumed to be damaged, while manmade objects are assumed to have a regular shape and uniform radiometric characteristics. For example, in Figure 3-7b the leaf-off tree was misdetected as damage, since it shows strong irregular texture pattern. Similarly, the damaged regions are not detected as they show smooth texture.

(a) UAV image of dataset 1 (left); detected damaged regions are highlighted in red, and the false positives are highlighted using yellow circles (right)

(b) Subset of Pictometry image of dataset 2 (left); detected damaged regions are highlighted in red, and the false positives and false negatives are highlighted using yellow and green circles, respectively (right). images © Pictometry

(c) Street view image of dataset3 (left); detected damaged regions are highlighted in red, and the false positives and false negatives are highlighted using yellow and green circles, respectively (right).

Figure 3-7 Damage classification of images based on best performing supervised model

Table 3-6 Performance of feature descriptors when associated with different learning algorithms for dataset 4 (COM_{3109}) comprising patches from UAV, Pictometry and street-view images (training samples = #2176, testing samples = # 933).

Dataset4	**SVM**			**RF**			**Adaboost**		
	Precision	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**
Gabor	0.79	0.75	0.77	0.76	0.64	0.72	0.62	0.58	0.62
HoG	0.81	0.62	0.73	0.79	0.64	0.71	0.71	0.57	0.61
BoW-Gabor	**0.95**	**0.88**	**0.91**	**0.93**	**0.79**	**0.86**	0.64	0.68	0.67
BoW-HOG	0.89	0.87	0.88	0.83	0.76	0.80	**0.80**	0.64	0.74
BoW-SURF	0.90	0.84	0.87	0.83	0.77	0.80	0.79	**0.75**	**0.77**

3.4 Observations and analysis

For convenient analysis of the results, the datasets 1 and 3, which were not cluttered and less affected by shadows and trees, are referred to as non-complex datasets, while datasets 2 and 4, where the image patches were mostly cluttered and severely affected by shadows and trees, are referred to as complex datasets. Also, for convenience, the datasets are named based on the image characteristics and number of samples as described in Table 3-7.

Table 3-7 Naming of datasets based on the image characteristics and number of samples

Dataset	Name	Description	Scene complexity
Dataset 1	UAV_{966}	966 image patches generated from UAV images.	Non-complex
Dataset 2	PIC_{1256}	1256 image patches generated from Pictometry images.	Complex
Dataset 3	SVI_{887}	887 image patches generated from street view images.	Non-complex
Dataset 4	COM_{3109}	Comprehensive dataset, where datasets 1, 2 & 3 are combined, containing 3109 image patches.	Complex

3.4.1 Global representation of HoG and Gabor wavelet for damage classification

The results show that the global representations of HoG and Gabor wavelet feature descriptors have great potential to identify the damaged regions in the image, if the defined image patches are non-complex. For example, the supervised models constructed for UAV_{966} (non-complex) based on the global representation of Gabor wavelet and HoG features resulted in accuracies of 95% and 93%, respectively (Table 3-3). However, the same feature descriptors Gabor and HoG produced accuracies of 82% and 72%, respectively for PIC_{1256} (Table 3-4), where the defined image patches were mostly complex. Moreover, the same features Gabor and HoG produced highly inferior results for COM_{3109}, which was more complex than the other datasets (Table 3-6). This clearly indicates that the robustness of the global representation of HoG and Gabor features declines with an increase in image patch complexity. This is because in the global representation the radiometric characteristics of the complex region (e.g., clutter, shadows and trees) resemble the radiometric characteristics of damaged regions, which are generally more non-uniform than radiometric patterns of non-damaged regions (cf. Figure 3-8). Consider an image patch that contains different objects with different dominant orientations. The global description of this image patch based on gradient orientation is the aggregation of all gradient orientation information within this patch. In such a case the image patch would seem to possess gradient orientations in many directions, which resemble the radiometric characteristics of damaged regions. For example, consider Figure 3-8 a as an image patch that contains four different objects (annotated as A, B, C and D) with different gradient orientation patterns. The gradient pattern derived locally for each object was overlaid on the corresponding object with a black background. These local patterns show that each object possesses dominant orientations which were more uniform. However, the global gradient pattern derived for the whole image patch was non-uniform and resembles the characteristics of damaged regions (cf. Figure 3-8 a). Thus, it is difficult to classify an image patch based on global features in case it is cluttered. Also, trees and shadows possess irregular shapes and non-uniform gradient orientations, which also resemble the radiometric characteristics of the damaged regions. Hence, global feature descriptors-based damage classification did not efficiently classify the image patches that were strongly affected by trees and shadow.

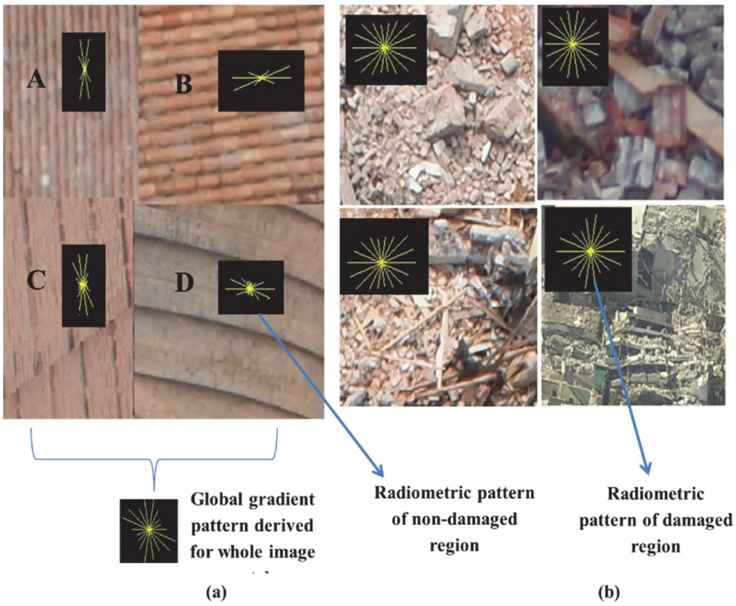


Figure 3-8 (a) Local and global gradient pattern of an image patch that contains four objects with different dominant orientations; (b) gradient pattern of damaged regions

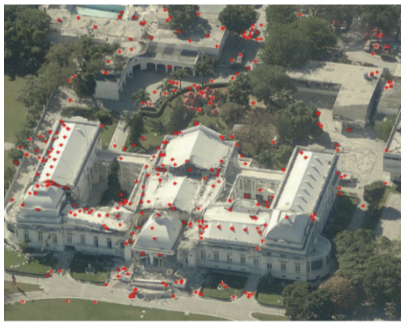
3.4.2 BoW-based feature representation for damage classification

The Gabor and HoG features produced superior results for all datasets when represented in a BoW framework than represented in conventional global scale for damage classification. Although the BoW approach produced superior results to the conventional approach, there was no significant difference in the performance between them when the considered image regions were not complex, e.g., UAV_{966} (cf. Table 3-3). However, in complex image regions there was a significant performance difference between the BoW and conventional feature representation. This is evident from the results for PIC_{1256}, where the BoW-based Gabor and HoG produced maximum accuracies of 88% and 91%, respectively, which are 9% and 19% higher than the accuracies obtained by those features when represented at a global scale (cf. Table 3-4). This shows that BoW-based Gabor and HoG features are more robust to clutter, trees and shadows than when they are represented at a global scale. The following characteristics made the BoW approach more robust compared to the global representation:

1. Unlike the global representation, the BoW approach does not aggregate the radiometric patterns within the image patch. Instead, it describes the image patch based on the number of salient points, where each point is described by the local radiometric pattern derived from its neighbourhood. Therefore, in case of no damage, the image patch will be represented by points with a uniform radiometric pattern (gradient orientation), even if the image patch contains objects with different dominant orientations. On the other hand, if the image patch contains damage it will be represented by the points with non-uniform gradient orientations. The final damage classification is performed by analysing the pattern of the occurrences of local radiometric patterns within the image patch. This eliminates the ambiguity caused by mixed radiometric pattern typical for the global representation, making the BoW comparatively more robust.
2. The BoW approach considers only the salient points as representatives to describe the image patch. The salient point selection method based on SURF mostly did not consider the pixels of shadows and trees as salient points: Figure 3-9 a & b show the strongest 300 SURF points in the image, where most of the detected points are not corresponding to trees and shadows. Thus, the BoW approach largely eliminates the shadows and trees in the damage classification process, which was one of the major problems in the global descriptors-based damage classification. Moreover, the pixels corresponding to the damaged regions were often detected as salient points, as they show a stronger gradient than other objects (cf. Figure 3-9 a). This ensures that the number of points corresponding to the damaged portion will always be significantly proportional to the number of points corresponding to the non-damaged objects, even if only a small portion of the image patch

is damaged (cf. Figure 3-9 a & b). This specific characteristic of the SURF points made the BoW-based damage classification approach invariant to scale, clutter and scene complexity.

(a)



(b)

Figure 3-9 Detected SURF points are plotted on the image: (a) Strongest 300 SURF points among 4032x3024 pixels; (b) Strongest 300 SURF points among 977x835 pixels, images © Pictometry

3.4.3 Impact of choice of learning algorithm

The results show that the choice of learning algorithm has a significant impact on damage classification performance, since the feature descriptors performed differently for different datasets when associated with different learning algorithms. The accuracies produced by SVM, RF and Adaboost for all datasets when they were associated with different feature descriptors are depicted in Figure 3-10. The plot shows that 1) SVM and RF mostly outperformed Adaboost; 2) Using the global feature descriptors the performances of RF and SVM varied with the datasets: RF produced superior results compared to SVM for UAV_{966} and SVI_{887}, whereas it produced inferior results than SVM for PIC_{1256} and COM_{3109}. This shows that the performance of the learning algorithm is highly dependent on the characteristics of the datasets, with SVM performing well for complex datasets and RF performing well for non-complex dataset. However, using the BoW approach the SVM mostly outperformed RF in the classification for all datasets, irrespective of the feature descriptors (cf. Figure 3-10). One overall conclusion from this is that the SVM based supervised models were more reliable and mostly showed better generalization performance than RF, particularly for the complex datasets.

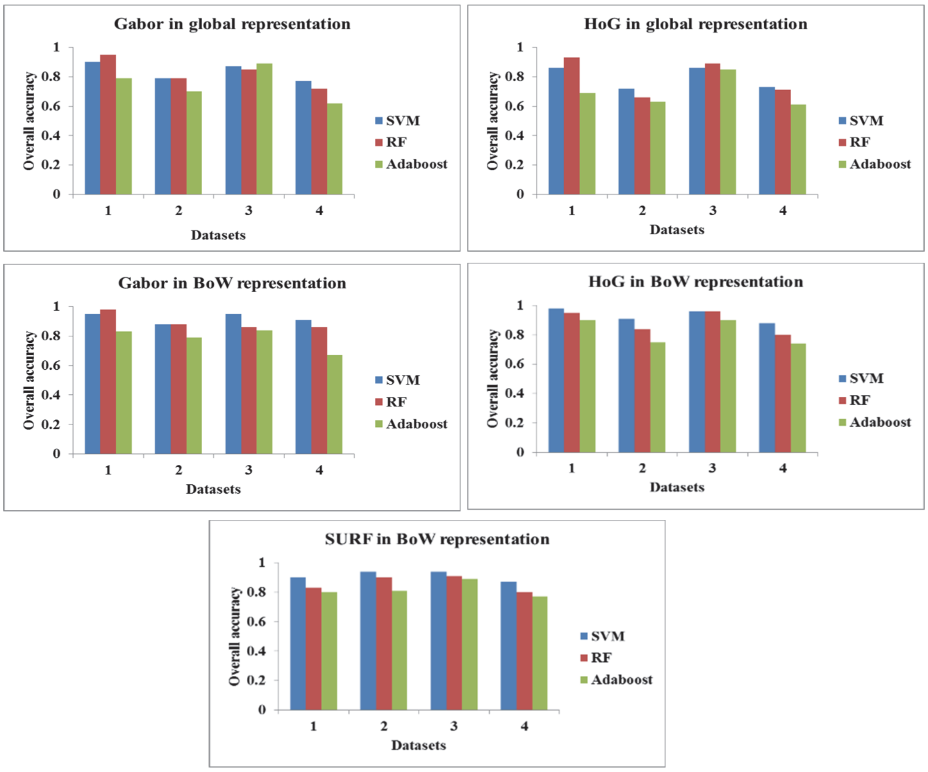


Figure 3-10 The accuracy produced by the feature descriptors for each dataset when associated with different classifiers

3.5 Discussion

The primary objective of this research work was to develop a damage classification method that classifies a given image patch as damaged or non-damaged, irrespective of scale, image characteristics and scene complexity. The damage classification method was developed by considering various feature descriptors (HoG, Gabor and SURF), different feature descriptor representations (Global and BoW), different learning algorithms (SVM, RF and Adaboost) and image datasets with different levels of scale and scene complexity. It was shown that the feature representation has a significant impact on the performance of the damage classification compared to other components such as features descriptor and learning algorithm. For all datasets, the BoW-based damage classification models performed well for all combinations of feature descriptors and learning algorithms, compared to the models developed based on global representation. Particularly, concerning COM_{3109} (the comprehensive dataset), the accuracy obtained with the best-performing feature descriptor (Gabor) and learning algorithm (SVM) with the global representation improved by 14% (from 77% to 91%), when tested with the BoW representation (cf. Table 3-6). The choice of learning algorithm was found to be the second significant factor in the performance of the damage classification model: the SVM produced significantly better results than RF and Adaboost for all feature descriptors in the BoW representation (cf. Table 3-6). The considered feature descriptors performed equally well and, hence, the choice of feature descriptor was found to have least impact on the performance of the model. The Gabor features led to a 3% and 4% improvement in accuracy compared to HoG and SURF, respectively, when the image patches were classified with SVM in the BoW framework. This small improvement also may be due to the additional information that Gabor features possess compared to HoG. For example, in Gabor, the gradient orientations information is extracted based on five different frequency scales, whereas in HoG the gradient orientation information is extracted at only one frequency scale (cf. 2.1.1). However, these improvements are significantly more modest when compared to the 14% of improvement in accuracy between BoW and global representation (cf. Table 3-6). This highlights the importance of feature representation, regardless of the potential of features. Overall, SVM associated with Gabor feature descriptors in the BoW framework was found to produce the most robust and generalized damage classification model. Even visually, the damage classification was found to be more accurate when the images of different scales, camera views and capturing platforms, and different levels of scene complexity, were classified by the best performing model (cf. Figure 3-7). Shadowed areas continue being a major problem in damage classification. Since the damaged regions covered by shadows show low contrast, they were not detected by our BoW-based approach (no SURF points in those areas). However, it is important to identify the damages in low-contrast regions as well; therefore, further tuning of the methods or identifying the optimal strategy that can make our approach work even in low contrast regions is required to increase the robustness of the model.

The BoW framework consists of various components such as feature descriptor, learning algorithm and the visual word dictionary construction. The algorithms used for each component are associated with a number of parameters (cf. Table 3-1). The performance of the BoW-based damage classification model might be further improved by tuning the parameters of the algorithm or modifying/ replacing the algorithm of the specific component. For example, the iterative *k*-means clustering was used to construct the visual word dictionary, whereas other feature encoding methods such as auto-encoders (e.g., Vincent et al. (2010)), which encode the features differently compared to *k*-means, may produce a better visual word dictionary and thereby can potentially improve the performance of the model as well. Concerning the feature descriptors, all three-feature descriptors were used independently to construct the damage classification models, whereas the combined use of feature descriptors may also improve the performance of our model. Similarly, concerning the learning algorithm we used a single kernel-based SVM for constructing the damage classification model, whereas the multiple-kernel (e.g., Bucak et al. (2014)) based learning may improve the performance of the model as well. We did not attempt to fine-tune the model by exploring all those possible approaches, because the principal focus of this research work was to analyze the potential of the BoW framework in damage classification.

The developed method can identify the damages related to debris/rubble piles that are strong indicators of building collapse or severe structural damage, which would be very useful for first responders involved in disaster response, but also other stakeholders such as governmental agencies assessing post-disaster construction needs, or insurance companies. However, for detailed building level damage assessment, these evidences alone are not sufficient to infer the complete damage state of the building, nor the total damage cost, as the latter also depends on internal (invisible) damage, and on building functions being affected, which is not always visible. However, along with other damage evidences such as cracks, inclined elements, etc., these evidences are also important in the damage classification process. From a practical point of view especially the observations we made using the combined dataset 4 (COM_{3109}) are very interesting. Although the used patches vary significantly in terms of scale and complexity, an overall accuracy of around 90% was reached (cf. Table 3-6). Transferred to an actual disaster scenario, where quick interpretation of image data is needed, this would mean that an already existing database can be used to train a model and new images can be readily classified, and a similar overall accuracy might be expected. Hence, at least for a first damage assessment, the tedious manual referencing might not be necessary.

3.6 Conclusion and outlook

A damage classification based on BoW was developed to classify a given image patch as damaged or non-damaged, irrespective of scale, image characteristics and scene complexity. Various combinations of image features (Gabor wavelets, HoG and SURF) and supervised classifiers (SVM, RF and Adaboost) are tested in both, BoW framework and conventional global feature representation approach using four different datasets. The BoW framework outperformed conventional global feature representation approaches in all scenarios (i.e. for all combinations of feature descriptors, classifiers and datasets), and produced an average accuracy of approximately 90%. Although the developed model can well identify the damaged regions in the images, it cannot classify the detected damaged regions into specific types, such as debris, rubble piles, spalling and inter-story collapse. We need contextual information and 3D geometric features such as shape, location, characteristics of the neighbouring elements and local height variation of the damaged region, to identify the actual category of damage. For example, the damage patterns on large intact planar elements could be classified as spalling, whereas the damage pattern on the ground with large local height variations and no large 3D segments could be classified as debris. Therefore, the potential extension of this work will be the development of methods for classification of the detected damaged regions into actual damage categories.

As stated earlier, the feature descriptor component in BoW framework has a significant impact on the performance of the model. Here, the texture features are chosen to examine our BoW framework as their potential in damage detection has been demonstrated well by previous studies as highlighted in the introduction. However, recent studies report that supervised feature learning methods such as convolutional neural networks (CNN) could learn the feature and its representation directly from the image pixel values chosen for a specific application (Szegedy et al., 2015). Hence, these features are referred to as data-adaptive features and they are found to be superior to well-proven handcrafted features such as Gabor and HoG for many computer vison applications including image classification (Karpathy et al., 2014; Zuo et al., 2014). Therefore, we intend to explore the potential of CNNs for damage classification in the future.

3.7 References of Chapter 3

Adams, S.M., Levitan, M.L., Friedland, C.J., 2013. High resolution imagery collection utilizing unmanned aerial vehicles (UAVs) for post-disaster studies. Advances in Hurricane Engineering: Learning from Our Past, 777-793.

Arivazhagan, S., Ganesan, L., Priyal, S.P., 2006. Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters,* 27 (16), pp. 1976-1982.
Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features, *European conference on computer vision*. Springer, pp. 404-417.
Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research,* 13 (Feb), pp. 281-305.
Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization, *Advances in Neural Information Processing Systems*, pp. 2546-2554.
Bouslimi, R., Messaoudi, A., Akaichi, J., 2013. Using a bag of words for automatic medical image annotation with a latent semantic. *arXiv preprint arXiv:1306.0178*, pp.
Breiman, L., 2001. Random forests. *Machine learning,* 45 (1), pp. 5-32.
Bucak, S.S., Jin, R., Jain, A.K., 2014. Multiple kernel learning for visual object recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 36 (7), pp. 1354-1369.
Carneiro, G., Jepson, A.D., 2009. The quantitative characterization of the distinctiveness and robustness of local image descriptors. *Image and Vision Computing,* 27 (8), pp. 1143-1156.
Conde, C., Moctezuma, D., De Diego, I.M., Cabello, E., 2013. HoGG: Gabor and HoG-based human detection for surveillance in non-controlled environments. *Neurocomputing,* 100, pp. 19-30.
Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, pp. 886-893.
Dell'Acqua, F., Gamba, P., 2012. Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE,* 100 (10), pp. 2876-2890.
Déniz, O., Bueno, G., Salido, J., De la Torre, F., 2011. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters,* 32 (12), pp. 1598-1603.
Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.
Fernandez Galarreta, J., Kerle, N., Gerke, M., 2015. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. *Natural Hazards and Earth System Sciences,* 15 (6), pp. 1087-1101.
Ferraz, C.T., Pereira, O., Rosa, M.V., Gonzaga, A., 2014. Object recognition based on bag of features and a new local pattern descriptor. *International Journal of Pattern Recognition and Artificial Intelligence,* 28 (08), pp. 32 P.
Gerke, M., 2011. Supervised classification of multiple view images in object space for seismic damage assessment. Photogrammetric Image Analysis, Springer, 221-232.

Gerke, M., Kerle, N., 2011. Automatic structural seismic damage assessment with airborne oblique pictometry imagery. *Photogrammetric Engineering & Remote Sensing,* 77 (9), pp. 885-898.

Huang, D., Zhu, C., Wang, Y., Chen, L., 2014. HSOG: a novel local image descriptor based on histograms of the second-order gradients. *IEEE Transactions on Image Processing,* 23 (11), pp. 4680-4695.

Jin, Y., Ruan, Q., 2009. Face recognition using Gabor-based improved supervised locality preserving projections. *Computing and Informatics,* 28 (1), pp. 81-95.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks, *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725-1732.

Kaya, G.T., Ersoy, O.K., Kamasak, M.E., 2010. Spectral and spatial classification of earthquake images by support vector selection and adaptation. *International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 194-197.

Kerle, N., Hoffman, R.R., 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural hazards and earth system sciences (NHESS),* 13 (1), pp. 97-113.

Khan, N.Y., McCane, B., Wyvill, G., 2011. SIFT and SURF performance evaluation against various image deformations on benchmark dataset, *International Conference on Digital Image Computing Techniques and Applications (DICTA), 2011* pp. 501-506.

Li, P., Xu, H., Guo, J., 2010a. Urban building damage detection from very high resolution imagery using OCSVM and spatial features. *International Journal of Remote Sensing,* 31 (13), pp. 3393-3409.

Li, Z., Imai, J.-i., Kaneko, M., 2010b. Face and expression recognition based on bag of words method considering holistic and local image features, *International Symposium on Communications and Information Technologies (ISCIT), 2010* IEEE, pp. 1-6.

Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., Cao, L., Huang, T., 2011. Large-scale image classification: fast feature extraction and svm training, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011* pp. 1689-1696.

Lou, X.-w., Huang, D.-c., Fan, L.-m., Xu, A.-j., 2014. An Image Classification Algorithm Based on Bag of Visual Words and Multi-kernel Learning. *Journal of Multimedia,* 9 (2), pp. 269-277.

Lu, Z., Wang, L., 2015. Learning descriptive visual representation for image classification and annotation. *Pattern Recognition,* 48 (2), pp. 498-508.

Ma, J., Qin, S., 2012. Automatic depicting algorithm of earthquake collapsed buildings with airborne high resolution image, *IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), 2012* pp. 939-942.

Ma, X., Shen, H., Yang, J., Zhang, L., Li, P., 2014. Polarimetric-spatial classification of SAR images based on the fusion of multiple classifiers. *IEEE*

Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7 (3), pp. 961-971.

Miura, H., Midorikawa, S., Kerle, N., 2013. Detection of building damage areas of the 2006 Central Java, Indonesia, earthquake through digital analysis of optical satellite images. *Earthquake Spectra,* 29 (2), pp. 453-473.

Oliva, A., Torralba, A., 2006. Building the gist of a scene: The role of global image features in recognition. *Progress in brain research,* 155, pp. 23-36.

Peng, X., Peng, Q., Qiao, Y., Chen, J., Afzal, M., 2013. A study on unsupervised dictionary learning and feature encoding for action classification. *arXiv preprint arXiv:1309.0309*, pp.

Radhika, S., Tamura, Y., Matsui, M., 2012. Use of post-storm images for automated tornado-borne debris path identification using texture-wavelet analysis. *Journal of Wind Engineering and Industrial Aerodynamics,* 107, pp. 202-213.

Rätsch, G., Onoda, T., Müller, K.-R., 2001. Soft margins for AdaBoost. *Machine learning,* 42 (3), pp. 287-320.

Reinartz, P., Tian, J., Nielsen, A.A., 2013. Building damage assessment after the earthquake in Haiti using two post-event satellite stereo imagery and DSMs, *Urban Remote Sensing Event (JURSE), 2013 Joint*, pp. 057-060.

Schölkopf, B., Smola, A.J., 2002. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

Stavrakoudis, D., Theocharis, J., Zalidis, G., 2011. A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66 (4), pp. 529-544.

Sui, H., Tu, J., Song, Z., Chen, G., Li, Q., 2014. A novel 3D building damage detection method using multiple overlapping UAV images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 40 (7), pp. 173.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9.

Tsai, C.-F., 2012. Bag-of-words representation in image annotation: A review. *ISRN Artificial Intelligence,* 2012, pp.

Vattani, A., 2011. K-means requires exponentially many iterations even in the plane. *Discrete & Computational Geometry,* 45 (4), pp. 596-616.

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, pp. 61-78.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research,* 11 (Dec), pp. 3371-3408.

Wang, Y., Mori, G., 2009. Human action recognition by semilatent topic models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 31 (10), pp. 1762-1774.

Wu, Z., Ke, Q., Sun, J., Shum, H.-Y., 2009. A multi-sample, multi-tree approach to bag-of-words image representation for image retrieval, *IEEE 12th International Conference on Computer Vision, 2009* pp. 1992-1999.

Yamazaki, F., Matsuoka, M., 2007. Remote sensing technologies in post-disaster damage assessment. *Journal of Earthquake and Tsunami,* 1 (03), pp. 193-210.

Yang, J., Yu, K., Gong, Y., Huang, T., 2009. Linear spatial pyramid matching using sparse coding for image classification, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.* , pp. 1794-1801.

Ye, P., Doermann, D., 2012. No-reference image quality assessment using visual codebooks. *IEEE Transactions on Image Processing,* 21 (7), pp. 3129-3138.

Yi, J., Su, F., 2014. Histogram of log-Gabor magnitude patterns for face recognition, *IEEE International Conference on Acoustics, speech and signal processing (ICASSP), 2014* pp. 519-523.

Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C., 2007. Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision,* 73 (2), pp. 213-238.

Zhuang, X., Wu, S., Natarajan, P., 2013. Compact bag-of-words visual representation for effective linear classification, *Proceedings of the 21st ACM international conference on Multimedia*. ACM, pp. 521-524.

Zuo, Y., Zhang, B., 2011. Robust hierarchical framework for image classification via sparse representation. *Tsinghua Science & Technology,* 16 (1), pp. 13-21.

Zuo, Z., Wang, G., Shuai, B., Zhao, L., Yang, Q., Jiang, X., 2014. Learning discriminative and shareable features for scene classification, *European Conference on Computer Vision*. Springer, pp. 552-568.

4 Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel- learning*

* This chapter is based on the article:
Vetrivel, A., Gerke, M., Kerle, N., Nex, F.C. and Vosselman, G. (2017) Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. In: ISPRS Journal of Photogrammetry and Remote Sensing, (2017) IN PRESS15 p.

Abstract

Oblique aerial images offer views of both building roofs and façades, and thus have been recognized as a potential source to detect severe building damages caused by destructive disaster events such as earthquakes. Therefore, they represent an important source of information for first responders or other stakeholders involved in the post-disaster response process. Several automated methods based on supervised learning have already been demonstrated for damage detection using oblique airborne images. However, they often do not generalize well when data from new unseen sites need to be processed, hampering their practical use. Reasons for this limitation include image and scene characteristics, though the most prominent one relates to the image features being used for training the classifier. Recently features based on deep learning approaches, such as convolutional neural networks (CNNs), have been shown to be more effective than conventional hand-crafted features, and have become the state-of-the-art in many domains, including remote sensing. Moreover, often oblique images are captured with high block overlap, facilitating the generation of dense 3D point clouds – an ideal source to derive geometric characteristics. We hypothesized that the use of CNN features, either independently or in combination with 3D point cloud features, would yield improved performance in damage detection. To this end we used CNN and 3D features, both independently and in combination, using images from manned and unmanned aerial platforms over several geographic locations that vary significantly in terms of image and scene characteristics. A multiple-kernel-learning framework, an effective way for integrating features from different modalities, was used for combining the two sets of features for classification. The results are encouraging: while CNN features produced an average classification accuracy of about 91%, the integration of 3D point cloud features led to an additional improvement of about 3% (i.e. an average classification accuracy of 94%). The significance of 3D point cloud features becomes more evident in the model transferability scenario (i.e., training and testing samples from different sites that vary slightly in the aforementioned characteristics), where the integration of CNN and 3D point cloud features significantly improved the model transferability accuracy up to a maximum of 7% compared with the accuracy achieved by CNN features alone. Overall, an average accuracy of 85% was achieved for the model transferability scenario across all experiments. Our main conclusion is that such an approach qualifies for practical use.

4.1 Introduction and related works

Automated detection of severe building damages is crucial in the coordination of fast response actions after any destructive disaster event such as earthquakes. Remote sensing technology has been recognized as a suitable source to provide timely data for automated detection of damaged buildings for larger areas (Dell'Acqua and Gamba, 2012; Dong and Shan, 2013). In particular, multi-view oblique images from manned aircraft and unmanned aerial vehicles (UAV) have been recognized as most suitable (Fernandez Galarreta et al., 2015a; Gerke and Kerle, 2011b; Kerle and Hoffman, 2013a). This is because these images capture both roofs and façades with very high spatial resolution, facilitating a holistic and detailed view of the building for damage assessment (Fernandez Galarreta et al., 2015a). Several studies have demonstrated automated detection of damaged buildings from the above mentioned image types, where the heavily damaged buildings are identified by recognizing externally visible damage evidences such as spalling, debris, rubble piles and broken elements, which are the strong indicators of severe structural damage (Dong and Shan, 2013; Vetrivel et al., 2015a). These damage evidences alone are not sufficient to infer the actual damage state of the building, as it requires additional information such as damages to internal building elements (e.g., columns and beams), which can rarely be directly inferred from images. Even though the information that can be derived from the images is limited, it is typically sufficient for satisfying the requirements of the stakeholders involved in search and rescue processes (Dong and Shan, 2013). Furthermore, the information can be used to plan for subsequent detailed assessments, for example, identifying hotspots that require immediate attention, and prioritizing the locations for field inspection. Towards this, numerous automated methods have been proposed for detection of aforementioned visual damage evidences from very high resolution images (Dong and Shan, 2013; Ma et al., 2016). These methods are largely based on two approaches: 1) comparison of pre- and post-event data, and 2) damage detection based on mono-temporal post-event data alone. The methods based on supervised learning strategies have been demonstrated to be effective for damage detection, particularly for the mono-temporal approach (Gerke and Kerle, 2011b; Vetrivel et al., 2015a). However, it is still challenging to adopt them for practical use. This is because the methods based on a supervised learning approach often do not generalize enough for them to be transferred to similar remote sensing data from unseen geographic locations, and Vetrivel et al. (2015a) discussed several reasons. One of the major factors is the poor generalization capability of the features and their representation used for constructing the supervised model, which is briefly described below:

1) Numerous image features have been examined for damage detection, and often the texture features such as Histogram of oriented Gradients (HoG) and Gabor features have been reported as effective (Samadzadegan and Rastiveisi, 2008; Tu et al., 2016a; Vetrivel et al., 2015a). Apart from feature selection, the choice of the feature representation strategy is also crucial, which is evident from the recent study by Vetrivel et al. (2016b), where the performance of the above mentioned texture features was found to be improved when represented using a Visual Bag of Words (BoW) framework (Ferraz et al., 2014b). Though the BoW representation improved the accuracy, problems related to generalization still exist. For example, Vetrivel et al. (2016b) examined the generalization capability of three different texture features: speeded up robust features (SURF), HoG features and Gabor features in a BoW framework for damage detection, using very high resolution images from different geographic locations (e.g., Italy, Haiti, India, etc.). They reported that the performance of the features is moderately inconsistent for datasets from different places, i.e. particular features perform better for specific datasets. The difference between the accuracies produced by these features for different datasets was reported to be 3 to 4%. The same set of features in a similar experimental setting as reported in Vetrivel et al. (2016b) was examined by Tu et al. (2016a) for another study area for damage detection. However, they reported contradictory findings: the difference in accuracy produced by different features was found to be higher (~10%), though there is no obvious explanation for this difference in results. Thus, identifying the generalized features for building a supervised classifier for damage detection is still challenging.
2) Additionally, all aforementioned features which have been reported as being efficient for damage detection are based on gradient orientation distribution patterns. These features are adopted for the damage detection process based on the assumption that structurally deformed regions often result in non-uniform radiometric distributions when compared to regions of undamaged man-made structural elements. For example, Figure 4-1a depicts the rudimentary gradient orientation pattern derived for damaged and undamaged image regions. However this assumption often fails in urban areas possessing complex texture (Vetrivel et al., 2016a). For example, consider Figure 4-1b where the

building elements possess complex textures that look similar to the radiometric pattern of damaged regions. In such areas, the reported texture features would fail, thereby hindering the automated assessment.

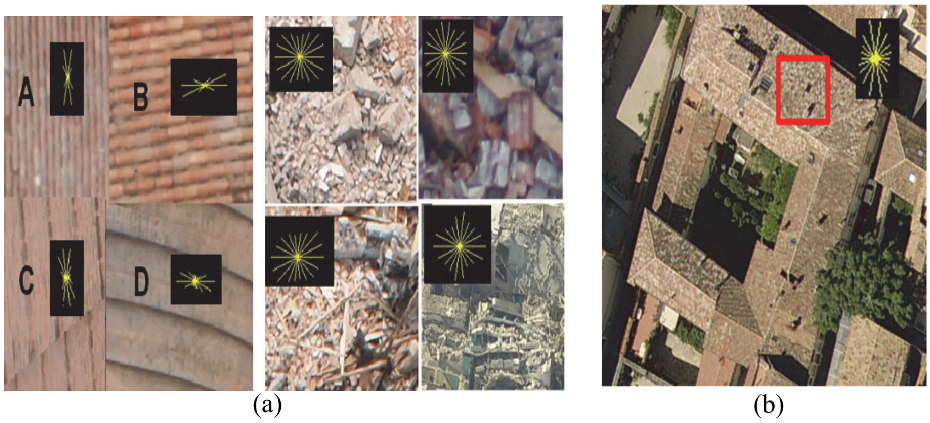


Figure 4-1 Rudimentary histogram of gradient orientation pattern depicted in yellow for (a) damaged (no annotations) and undamaged (annotated as A, B, C and D) image samples; (b) undamaged roof with complex texture highlighted in red rectangular box

Overall, the previously reported features are found to be inadequate to create a strong generalized supervised model for damage detection, and a feature descriptor robust to aforementioned limitations is highly desirable. Recently, the features from deep learning approaches such as Convolutional Neural Networks (CNNs) have been reported as being superior to conventional hand-crafted features, including the ones used in earlier state-of-the-art BoW framework for image classification in many applications including remote sensing (Hu et al., 2015; Karpathy et al., 2014; Sherrah, 2016; Szegedy et al., 2015; Zhou et al., 2015; Zuo et al., 2014). For example, several participants in the ISPRS urban scene classification challenge have achieved state-of-the-art accuracy for the ISPRS Vaihingen and Potsdam benchmark data sets using CNN features, outperforming all previously reported methods based on hand-crafted features (cf. ISPRS-Benchmark, 2016). Hence, we anticipate that CNN features would outperform the hand-crafted features in a damage detection application as well. This is examined in this study.

In the real world, man-made structural elements are complex and they often possess irregular radiometric patterns due to several reasons other than damage, including radiometric degradation of elements due to aging, or presence of dirt (cf. Figure 4-2). In such cases, our assumptions about damaged regions based on image-radiometric patterns may fail. In this scenario, the use of 3D geometric information could be of help to differentiate between the unusual radiometric pattern due to geometric deformation (damage) and other reasons. In general, 3D point clouds are an ideal source to infer geometric characteristics of structural elements. For example, Khoshelham et al. (2013b) demonstrated the potential of 3D point cloud features derived from post-event LiDAR point clouds for building damage detection. The oblique-view aerial images from manned- and unmanned aerial vehicles which have previously been identified as effective for damage detection are usually captured with high block overlap, facilitating the generation of 3D point clouds (Nex and Remondino, 2014). We assume that the integrated use of 3D features from photogrammetric point clouds and CNN features from images would yield improved results. However, it is well known that the direct integration of features, i.e. stacking of features from different sources, possibly possessing different modalities, into a single feature vector for supervised classification is inefficient (Bucak et al., 2014; Gu et al., 2015). Alternatively, integrating features from different sources using a Multiple-Kernel-Learning (MKL) approach associated with a kernel-based classifier such as SVM has been reported to be effective and it is being commonly used (Bucak et al., 2014; Gu et al., 2015). In addition to feature subsets integration, the MKL also could be used to evaluate the significance of each feature subset in the classification process (Gönen and Alpaydın, 2011). Hence, in addition to feature subsets integration, MKL is being widely used as a feature selection algorithm as well (Cao et al., 2015). In this study, MKL is used to integrate 3D point cloud-based features and CNN features for classification, and is also used as a feature selection approach by inferring the contribution of each feature subset in the classification accuracy.

(a) (b)

Figure 4-2 Image samples (a) and (b) depicting radiometric variation as a result of degradation of building elements due to aging and fouling

To the best of our knowledge, so far no studies have been reported for building damage detection using 1) CNN features, 2) mono-temporal post-event photogrammetric point cloud features, and 3) the combination of 3D point cloud and CNN features based on MKL. Thus the objective of this study is to develop an automated framework for performing building damage detection using CNN- and 3D point cloud-features based on MKL. The developed framework is evaluated using datasets containing images of manned and unmanned aerial vehicles captured from different geographic locations that are highly variable in scene and image characteristics, in order to examine the generalization capability of the developed approach.

The research problems and objectives with their background and relevant literature are introduced in Section 1. The detailed description of the methodology is provided in Section 2. The information about the experimental setup, implementation details, data description and the results is provided in Section 3. Section 4 provides the overall discussions and conclusions.

4.2 Methodology

A framework is developed for constructing a supervised learning model for damage classification using CNN features independently and in combination with 3D point cloud features. This framework includes three components: 1) a method for the generation of super-pixels to facilitate the object-level analysis; 2) methods for extracting CNN and several 3D point cloud features, and 3) a method for integration of CNN and 3D point cloud features for classification. Each component in the framework is described below.

a) Step 1: Super-pixel construction

As stated earlier, damaged regions are identified by recognizing unusual radiometric and geometric patterns. An object or segment level analysis is often preferred for identifying damaged regions, as it can capture these damage patterns more precisely than pixel level features (Vetrivel et al., 2016b). The pertinence of an object-based approach for damage classification has been already demonstrated by several studies (Dong and Shan, 2013; Li et al., 2011). Therefore, super-pixels derived by over-segmentation of images are considered as the primary entity for generating the image patches for the feature extraction and damage classification process. In this study, the state-of-the-art algorithm represented by Simple Linear Iterative Clustering (SLIC) (cf. Achanta et al. 2012) is adopted for generating the super-pixels. The information about the tunable parameters associated with SLIC is provided in the Experimental Section. The CNN architecture adopted in this study requires input in a format of rectangular image patches, hence the super-pixels are converted into such patches by fitting an outer-rectangle.

b) Step 2: Feature extraction

The methods for extracting CNN and 3D point cloud features for the image patches derived based on the super-pixels are described below:

i. **CNN features**

   Feature extraction is a fundamental step in any image classification application. Particularly feature extraction based on filtering approaches has been found to be dominating (Arivazhagan et al., 2006; Tian, 2013b). In general, object level feature extraction is carried out in two steps. As a first step, the features are extracted using a set of filters, where usually the filter weights are defined based on mathematical functions (e.g., Gabor filter banks). Subsequently, the extracted features are represented by adopting a suitable feature encoding approach such as a BoW framework, which typically serves as an input for a classifier (Xu et al., 2010). A CNN also belongs to the category of filtering approaches, where the filter weights and feature encoding strategy are directly learned from the images (Bengio et al., 2013). Hence, they are referred to as data-specific features. CNN is typically built by stacking several layers. For example, Figure 4-3 depicts the pre-trained CNN model 'imagenet-caffe-alex' (cf. MatConvNet, 2016), designed based on the popular CNN architecture proposed by Krizhevsky et al. (2012). In a CNN, the first few layers will be the convolutional layers analogous to the filter banks in the filtering approach (e.g., C1-C5 in Figure 4-3). These layers are found to be capable of learning low-level features such as blobs, edges, and oriented gradients in the first layer, and gradually learn higher level contextual information in the subsequent layers. Typically, each convolutional layer will be followed by the Rectified Linear Unit (ReLU) and a pooling layer. The ReLU performs the non-linear transformation of the feature map produced by the convolutional layer to introduce nonlinearity to the system. The pooling layer is used to reduce the size of the feature maps produced by the convolutional layer. The few layers preceding the final layer in the network will be the fully connected layers, where the output of each of these layers provides features with high-level reasoning, by summarizing the information from feature maps of previous convolutional layers. This information can be used as a final feature vector, an input to any classifier. The final layer in the network is also the fully connected (FC) layer, usually coupled with a loss function such as softmax or SVM (Tang, 2013). The dimensionality of this layer should be equal to the number of classes defined for the classification process, as its output will be the probability scores for each defined class. The learning process is carried out by tuning the weights of the neurons in various layers of the network, using the back propagation algorithm based on a large number of labelled image samples. For more details about the layers in the CNN architecture refer to Zeiler and Fergus (2014) and Krizhevsky et al. (2012).

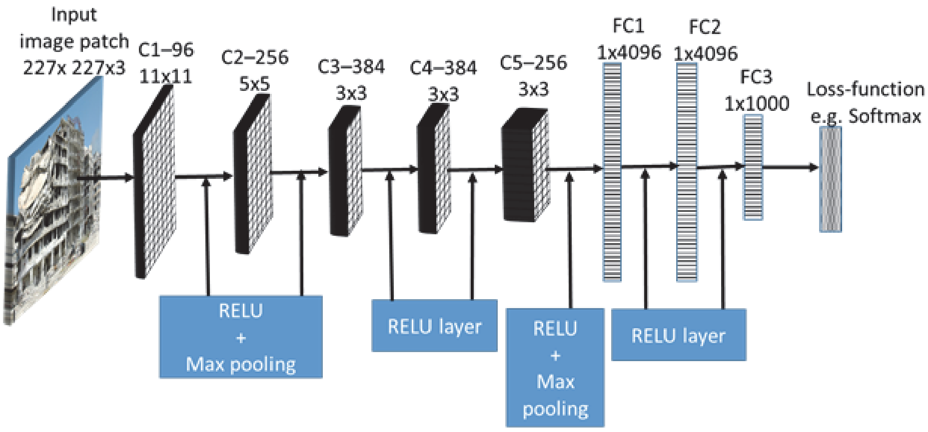


Figure 4-3. CNN architecture design of the pre-trained CNN model – 'imagenet-caffe-alex' designed based on one of the popular CNN architectures proposed by Krizhevsky et al. (2012) in which C1-C5 are convolutional layers and FC1-FC3 are fully connected layers. The values on the right-hand side and below C1-C5 indicate the number of filters and their sizes, respectively. The values below FC1-FC3 indicate their dimensions i.e., number of neurons in the fully connected layer.

In general, CNN feature extraction is performed in three scenarios:

1) ***Training from scratch:*** In this approach, the CNN architecture is designed from scratch using the layers mentioned earlier. However, designing the network, i.e. the choice of the number of layers and the number of hyper-parameters in each layer, is a challenging task. This is because there is no standard definition and it entirely depends on numerous factors, such as the characteristics of the input image (e.g., size, complexity, etc.), complexity of the application, number of classes and number of training samples available. In general, combinations of layers and their hyper-parameters that are suitable for specific applications are identified empirically based on the dataset. This approach requires a large amount of training data to avoid overfitting, hence it can be adopted only when a large amount of training sample is available. The features from this approach can be referred to as application- and data-specific features, as they are learned based on the image samples from the specific application (e.g., damage detection) and images with specific characteristics (e.g., aerial images).
2) ***Transfer learning using pre-trained models**:* Transfer learning in a CNN is achieved by learning application-specific features by adopting a pre-trained model that is already trained using large number of

samples (usually on the order of millions) from related domains, and tuning their weights using the training samples from the considered application. These kinds of pre-trained CNN models are increasingly available (MatConvNet, 2016). Moreover, the adaptability of these pre-trained models to perform transfer learning for remote sensing applications such as urban scene classification has already been demonstrated (Hu et al., 2015). While adopting the pre-trained CNN model for transfer learning it is mandatory to modify the final fully connected layer, as its dimension should be equal to the number of classes defined for the classification. Also, the architecture of the network can be altered by adding or removing layers, if required. However, in most cases the architecture of the pre-trained model is largely retained as the features learned by them are often found to be more generalized and contain information significant for any image classification application including remote sensing (Hu et al., 2015). To this end, the pre-trained model is modified in several ways and evaluated concurrently (i.e. by removing or adding some layers and concurrently evaluating using our datasets). The model which provided superior results is adopted here. For example, the pre-trained model depicted in Figure 4-3 when modified by adding two fully connected layers as depicted in Figure 4-4 provided superior results. The dimension of the last fully connected layer is set to two, as we intend to learn CNN features for performing binary classification for image regions, i.e. damaged or undamaged. The dimension of the layer FC4 preceding to the last layer has been empirically set to 256 after evaluating several dimensions (64, 128, 256 and 512) based on the dataset used in this study.

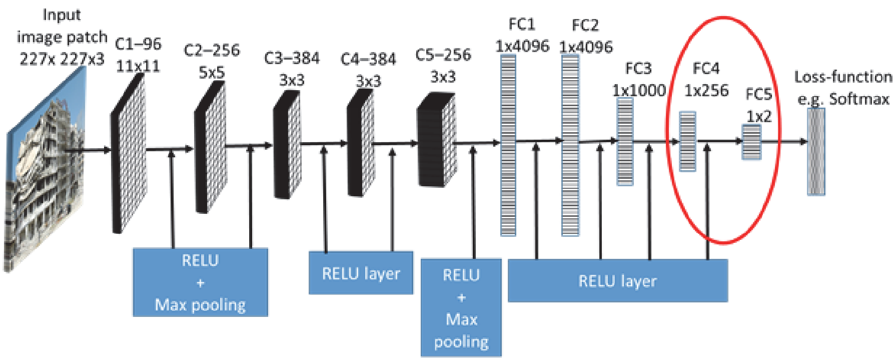


Figure 4-4. The CNN architecture for transfer learning by adding two extra fully connected layers (highlighted in red circle) to the existing pre-trained CNN model shown in Figure 4-3 for damage detection.

3) ***Pre-trained CNN model as feature extraction tool:*** In this approach the pre-trained CNN model acts as a tool for feature extraction, where the output of one of the fully connected layers in the CNN architecture is considered as the feature representation for a given image patch (super-pixels). This is based on the aforementioned assumption on generalization ability of the CNN features learned by the pre-trained CNN model using a large number of generic images with highly varying characteristics. The features extracted from the fully connected layer can be used for the classification process by fitting them with any classifier such as SVM. This approach also has been demonstrated for the classification of remote sensing images by few studies (Castelluccio et al., 2015; Cheng and Han, 2016).

In this study, the CNN features extracted using the above mentioned three approaches are examined for the damage detection process and their implementation details are provided in the experiment section.

ii. **3D point cloud features**

As stated earlier, it is hypothesized that 3D geometric features from a photogrammetric point cloud have the potential to identify geometrically deformed structural elements. Numerous 3D point cloud features have been reported for generic applications such as scene classification and object detection using 3D point clouds (Golovinskiy et al., 2009; Velizhev et al., 2012). In particular, 3D point cloud features derived based on 3D structure tensor have been reported to be effective (Gevaert et al., 2016; Hackel et al., 2016; Weinmann et al., 2015). The commonly used 3D point cloud features are listed in Table 4-1. A few of them, such as linearity, planarity and scattering, have already been demonstrated by Khoshelham et al. (2013b) to be effective for detection of damaged buildings in LiDAR point clouds. In general, these features are extracted at point level by computing the 3D structure tensor for each point based on its local neighborhood (Weinmann et al., 2015). The extraction of these features at segment level can be achieved by considering all 3D points in the segment as local neighborhood for constructing the 3D structure tensor (Weinmann et al., 2016) . Alternatively, it is also common to use the mean value of the point-level features computed for each 3D point within the segment as the segment-level representation (Khoshelham et al., 2013b). However, the above two strategies for extracting segment-level features are not suitable for the application of damage detection. This is because the damaged regions are identified by assuming that they will possess a non-uniform distribution of geometric characteristics, which cannot be captured well by the aforementioned segment-level feature extraction strategies. However, histograms are widely used as region-level descriptors and are expected to be effective for capturing the distribution pattern of local geometric variations within the segment. For example, in image-space the radiometric distribution patterns of damaged regions (image segments) have been reported to be well represented by histogram approaches such as Histogram of oriented Gradients (HoG) (Vetrivel et al., 2016b). Hence, in this study, we propose to use a similar histogram approach for representing the geometric distribution pattern based on local point-level features within the segment. The procedure for extracting various 3D point cloud features on segment-level is described below.

Table 4-1. 3D features based on 3D structure tensor derived from collection of 3D points

Features based on 3D structure tensor	***Feature definition (e_1, e_2 and e_3 are the normalized eigenvalues of the 3D structure tensor computed for 3D points derived based on the local neighborhood)***
Linearity L_λ	$(e_1-e_2)/e_1$
Planarity P_λ	$(e_2-e_3)/e_1$
Scattering S_λ	e_3/e_1
Omnivariance O_λ	$\sqrt[3]{e_1\, e_2\, e_3}$
Anisotropy A_λ	$(e_1-e_3)/e_1$
Eigenentropy E_λ	$-\sum_{i=1}^{3} e_i \ln(e_i)$
Change of curvature C_λ	$e_3/(e_1+ e_2 +e_3)$

Procedure for extracting segment level 3D point cloud features based on histogram

1) Select the over-segmented image and the corresponding 3D point cloud for damage detection. Derive the 3D point cloud features listed in Table 4-1 for each 3D point by computing the 3D structure tensor based on the local neighborhood. The optimal local neighborhood size for each 3D point is computed using the method proposed by Weinmann et al. (2015).

2) Identify the visible 3D points to the selected image using hidden point removal (HPR) algorithm (Katz et al., 2007c). Subsequently, project the features of these 3D points to the over-segmented image.
3) Derive the segment-level descriptors based on the projected point-level features using the histogram approach inspired from HoG which is described below.
   a. Select a rectangular image patch derived based on the super-pixel and split it into M × N blocks.
   b. All 3D point feature values range from 0–1 as they are normalized. Define the bin size (B_{size}) for computing the histogram.
   c. For each block compute the histogram of a specific 3D feature by adding its value to the corresponding bin.
   d. Concatenate all block-level histogram of the specific feature to derive the patch-level descriptor. The size of the final descriptor for each feature is equal to the number of blocks × size of the histogram bin.

The seven features listed in Table 4-1 and additionally the Z component of the normal vector of each 3D point within image segment are used to derive segment-level descriptors using the histogram approach as described above. Therefore, in total eight histogram-based feature subsets are used to describe a segment. For comparison, the segment-level 3D point cloud features are also derived using the aforementioned two conventional approaches: 1) deriving features by considering all 3D points within the segment as local neighborhood and 2) computing the mean of the point feature values within the segment. Therefore, overall ten 3D feature subsets are derived for each segment, as described in Table 4-2 for the damage detection process.

Table 4-2. 3D feature subsets considered for the damage detection process

Feature subsets (FS) name	***Description***	***Feature subset size***
FS1: H_L_λ	Histogram of linearity	1 x (M×N× B_{size}). For example, consider the image blocks are split into 4×4 (M×N) and histogram bin size (B_{size}) as 10, then FS size is equal to 1x160.
FS2: H_P_λ	Histogram of planarity	
FS3: H_S_λ	Histogram of scattering	
FS4: H_O_λ	Histogram of omnivariance	
FS5: H_A_λ	Histogram of anisotropy	
FS6: H_E_λ	Histogram of eigenentropy	
FS7: H_C_λ	Histogram of change of curvature	
FS8: H_N_z	Histogram of Z component of the normal vector	
FS9: S_3D	All seven features listed in Table 4-1 are derived for a segment by considering all 3D points within the segment as local neighborhood	1x7
FS10: M_3D	All seven features listed in Table 4-1 are derived for a segment by computing the mean of the point feature values within the segment	

c) Step 3: Classification

The classification process is carried out using a canonical kernel classifier, namely SVM, based on the features mentioned earlier. In this study, multiple feature subsets such as several 3D feature subsets and a CNN feature subset are considered independently and in-combination for damage classification. These feature subsets may possess different modalities. In this study, MKL is used as the tool for systematically aggregating different modalities into a single learned model. It is achieved through a convex combination of multiple kernels, each representing one feature subset:

$$k(x, x') = \sum_m \beta_m \, k_m(x, x')$$

where $k_m(x, x')$ is a basic kernel built for subjects x and x', β_m is a nonnegative weighting parameter with $\sum_m \beta_m = 1$, and m is the number of kernels where each kernel is associated with one feature subset. In this study, we adopt the widely used *Simple-MKL* proposed by Rakotomamonjy et al. (2008), where the optimal values of β_m are learned together with other SVM parameters by optimizing a single objective function. The derived values of β_m indicate the importance i.e. contribution of each feature subset in the classification process. The MKL framework adopted for damage detection is depicted in Figure 4-5.

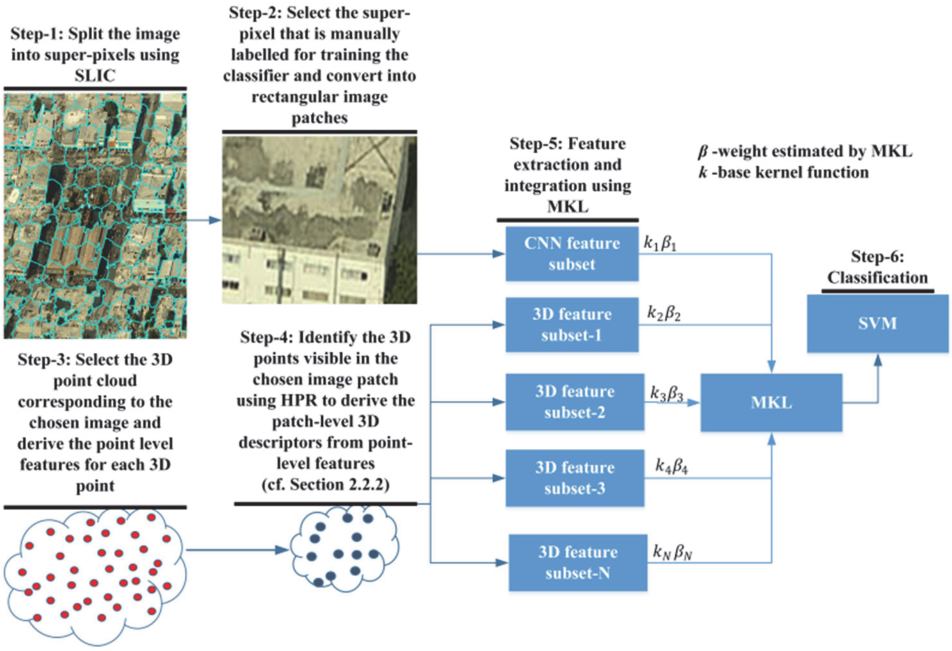


Figure 4-5. Overall steps for integrating CNN and 3D point cloud features using MKL for damage classification

4.3 Experiments

In this study we examined the potential of CNN and 3D features, both independently and synergistically for damage detection. Towards this several experiments have been conducted. Among them some experiments were designed based on the inference made in preceding experiments. Hence, the results and discussions are provided independently for each experiment.

4.3.1 Datasets

Two groups of datasets based on multi-view oblique images from two different platforms, 1) manned aircrafts and 2) UAVs, were considered in this study (cf. Table 4-3 and Table 4-4). These images were captured by different sensors for different disaster events from several geographic locations that are highly varying in scene characteristics, illumination condition and image characteristics. Particularly, the UAV images from all locations were captured with high overlap which facilitated the generation of a dense 3D point cloud. This was used for examining the combined use of image and 3D point cloud features for the damage detection process.

The information about the datasets is summarized in Table 4-3 and Table 4-4.

4.3.2 Preparation of training samples

The images from the above mentioned datasets were considered for preparing the training and testing samples for training and evaluating the supervised classifiers. The damage evidences such as debris, spalling, holes and building collapses that are clearly visible in the images were considered as damaged regions. The intact building elements, bare ground areas, cars, roads, etc., were considered as undamaged image regions for preparing negative training samples. Some portions of damaged and undamaged regions in the considered images were manually delineated using a polygon and labelled accordingly. Subsequently, the images were segmented into super-pixels using the SLIC method. The CNN model requires as input a rectangular image patch in specific size. In order to generate the super-pixels with about uniform size and shape, the parameter *'m'* in SLIC, which controls the compactness of a super-pixel, was empirically determined as 40 after examining several values (i.e. 10, 20, 30, 40 and 50). Another important parameter associated with SLIC is the number of desired super-pixels, which was set to be the dimension of the image (i.e. number of pixels) divided by desired size of the super-pixels (100x100).The super-pixels having at least 50% of their area overlapping with a polygon labeled as damaged or undamaged were chosen and converted into rectangular image patches as depicted in Figure 4-6. These are then constituted as the final training and testing samples for damage detection. In all experiments, the training and testing samples had a 70:30 ratio. The count of image patches corresponding to each dataset is provided in Table 4-3 and Table 4-4.

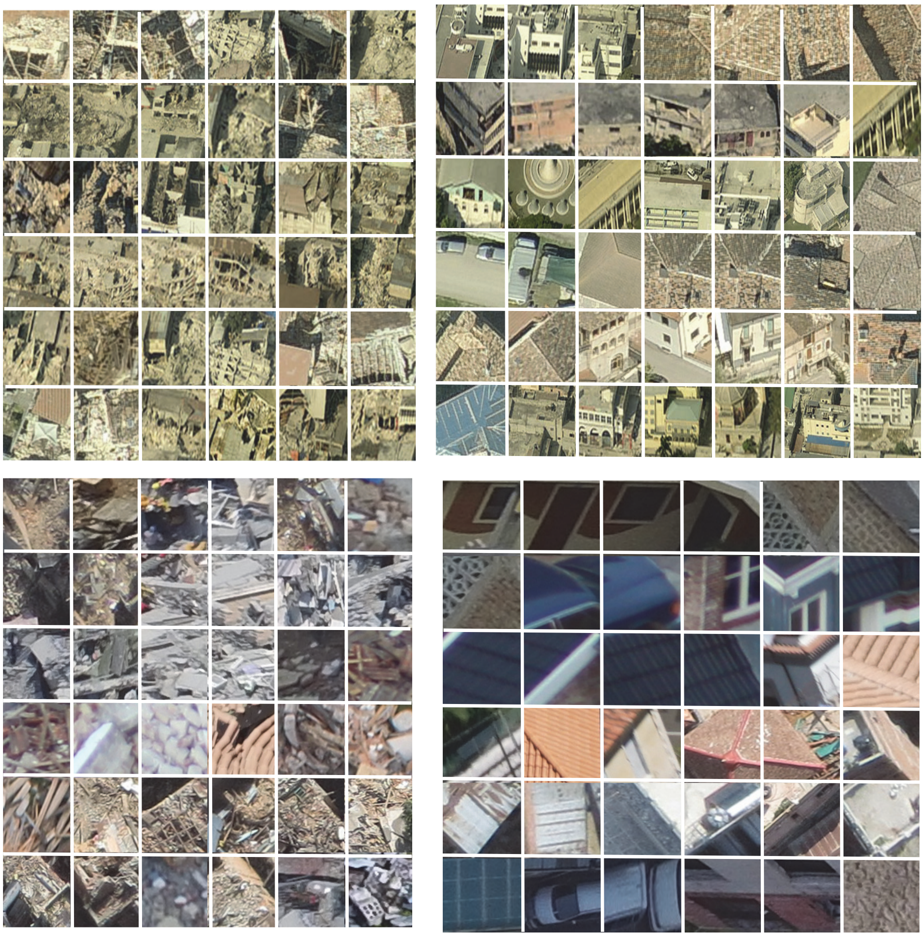


Figure 4-6. Sample image patches of damaged (left) and undamaged (right) regions generated based on super-pixels of manned aircraft images (top), and UAV images (bottom) for framing the training and testing samples for supervised classification

Table 4-3. Dataset 1 ($Aerial_{7130}$)[2] description of the training and testing samples derived from images of manned aerial platform for different geographic locations

Geographic Location	***Year of disaster event***	***Type of event***	***No of samples***			***Acronym for dataset from each location***
			Damaged	***Undamaged***	***Total***	
Port-au-Prince, Haiti	2010	Earthquake	984	942	1926	A_PAP_{1926}
Bidonville, Haiti	2010	Earthquake	795	860	1655	A_BID_{1655}
L'Aquila, Italy	2009	Earthquake	485	503	988	A_LAQ_{988}
Onna, Italy	2009	Earthquake	647	609	1256	A_ONN_{1256}
Tempera, Italy	2009	Earthquake	493	568	1061	A_TEM_{1061}
Mirabello, Italy	2012	Earthquake	97	147	244	A_MIR_{244}
Total			3501	3647	7130	$Aerial_{7130H}$
Overall description: *The samples were derived from images captured by manned aircrafts (CGR and Pictometry) with five cameras (one nadir and four oblique views), with a spatial resolution around 10 –16 cm.*						

Table 4-4. Dataset 2 (UAV_{5414}) description of the training and testing samples derived from images of UAV for different geographic locations

Location	***Year of event***	***Type of event***	***No of samples***			***Acronym for dataset from each location***
			Damaged	***Undamaged***	***Total***	
Ecuador, Peru	2016	Earthquake	307	476	783	U_ECU_{783}
Kathmandu, Nepal	2015	Earthquake	837	715	1552	U_KAT_{1552}
L'Aquila, Italy	2009	Earthquake	184	201	385	U_LAQ_{385}
Pingtung, Taiwan	2016	Earthquake	483	506	989	U_PIN_{989}
Mirabello, Italy	2012	Earthquake	568	679	1247	U_MIR_{1247}
Gronau, Germany	2013	Manually destructed industrial area	214	244	558	U_GRO_{244}
Total			2593	2821	5414	UAV_{5414}
Overall description: *The samples were derived from images captured by UAVs at different heights, views, cameras and lighting conditions, with spatial resolutions ranging from 1 –5 cm.*						

[2] Subscript in the dataset name indicates the number of samples present in the data set

4.3.3 Experiment 1: CNN features only for damage detection

The CNN features were extracted according to the three scenarios (cf. Section 2.2.1) and independently examined for damage detection. The main motivation of this experiment is to examine whether the features learned by the pre-trained CNN models based on generic images (i.e. image patches from various domains) are generalized enough to be used for the classification of damaged and undamaged regions in the remote sensing images. Alternatively features learned from the domain-specific training samples would be required. The implementation details of CNN for aforementioned three scenarios are described below.

a) CNN from scratch (CNN_S): As stated earlier, the designing strategy of CNN depends on various factors including the available number of training samples and the number of classes defined for the classification process. In this study, we have a limited number of training samples for building the binary classifier (damaged or undamaged). In this case, we cannot adopt the design of established pre-trained models as their network size is too large (i.e. too many hyper-parameters), which cannot be effectively tuned with the limited number of training samples. Hence, we designed a CNN architecture empirically, which was similar to the architecture of pre-trained models but with a relatively low number of layers and a low number of hyper-parameters in each layer. The details of the designed CNN model are given in Table 4-5. The weights in the network were initialized based on a Gaussian distribution with mean 0 and standard deviation 0.01. The network was independently trained and tested for both datasets. The training and testing processes were repeated ten times with varying training and testing sample sets. The mean and standard deviation of the performance measures overall accuracy, precision and recall obtained for ten trials are reported in Table 4-6. In this chapter, the measures reported with its range (e.g. $x \pm y$) always indicate the mean and standard deviation obtained for ten trials.

Table 4-5. Details of CNN architecture designed for learning features from scratch

CNN architecture for training from scratch		
Layer number	***Layer name***	***Properties***
1	Input layer	Input image patch size: 100x100x3
2	Convolutional	Number of filters: 9; Filter size: 11x11
3	RELU	–
4	Maxpooling	Pool size 2x2
5	Convolutional	Number of filters: 21; Filter size: 7x7
6	RELU	–
7	Maxpooling	Pool size 2x2
8	Convolutional	Number of filters: 41; Filter size: 3x3
9	RELU	–
10	Maxpooling	Pool size 2x2
11	Fully connected	Size: 1x256
12	RELU	–
13	Dropout	Dropout ratio: 0.5
14	Fully connected	Size: 1x100
15	Fully connected	Size: 1x2
16	Softmax	–

b) Transfer learning using pre-trained CNN (CNN_T): The pre-trained CNN model was modified by adding two extra fully connected layers as depicted in Figure 4-4, and was deployed for performing transfer learning i.e. fine-tuning of a pre-trained model using the domain-specific training samples. The learning rate was set to a very low (0.000005) to avoid large weight updates, preserving the useful features learned by the pre-trained model. The network was independently fine-tuned using the training samples from both datasets and tested with corresponding testing samples (cf. Table 4-6).

***c)* Pre-trained CNN model as feature extraction tool (CNN_F):** The CNN pre-trained model depicted in Figure 4-3 was adopted as a tool for CNN feature extraction without any further tuning of the weights in the network using domain-specific samples. The output of the fully connected layer FC2 with dimension 1x4096 was considered as the final feature vector for an image patch (227x227x3) input to the pre-trained model. The classification process was carried out independently for both datasets and the results are reported in Table 4-6.

Table 4-6. Results of damage detection on two different datasets using the supervised classifiers constructed based on CNN features extracted from three different scenarios

Data set / Models	*Aerial$_{7130}$*				*UAV$_{5414}$*			
	Accuracy (in %)	***Precision (in %)***	***Recall (in %)***	***CT (in s)***	***Accuracy (in %)***	***Precision (in %)***	***Recall (in %)***	***CT (in s)***
CNN_S	93.62± 0.50	95.25± 0.19	**92.13 ± 0.95**	7330	89.60± 1.19	92.44± 0.40	86.83 ± 2.53	6974
CNN_T	**94.24± 0.21**	**96.42± 0.26**	92.08 ± 0.48	8173	**92.33± 0.62**	92.42± 0.37	**92.19 ± 1.27**	7623
CNN_F	92.18± 0.16	93.38± 0.15	91.01 ± 0.38	**635**	91.15± 0.56	**92.56± 0.43**	89.54 ± 1.22	**541**

The computation time (CT) is the time taken for training and testing of the classifier on a laptop with 8 CPU cores (Intel® Core™ i7- 4710HQ), NVIDIA® GTX-860M 4G GPU and 16 GB of RAM.

The pre-trained model tuned with domain-specific samples (CNN_T) performed better than the other two cases, CNN_S and CNN_F, when tested using both datasets. CNN_S, the model trained with domain-specific samples, did not perform better than CNN_F (features learned from generic images) for the dataset UAV$_{5414}$**.** This could be because of two problems: 1) suboptimal design of the network, or 2) insufficient training samples as building a CNN model from scratch typically requires a large number of training samples. Hence, it is expected that its performance may be improved by increasing the number of training samples or further optimizing the architecture of the CNN. However, an overall observation is that there was no significant difference in the classification accuracy for CNNs used in three different scenarios. The results indicate that the CNN features learned by the pre-trained model from generic images generalize enough for adopting them directly for the damage detection application. Moreover, this approach (CNN_F) requires a significantly lower number of training samples compared to other two methods (CNN_S and CNN_T), though providing comparable accuracy with much more efficient computation. Hence, the CNN_F approach is highly recommended for the damage detection application. Samples of aerial and UAV images classified using CNN_F model are depicted in Figure 4-7 and Figure 4-8, respectively. Based on visual interpretation, the CNN_F model detected almost all damaged regions, though with a small number of false positives.

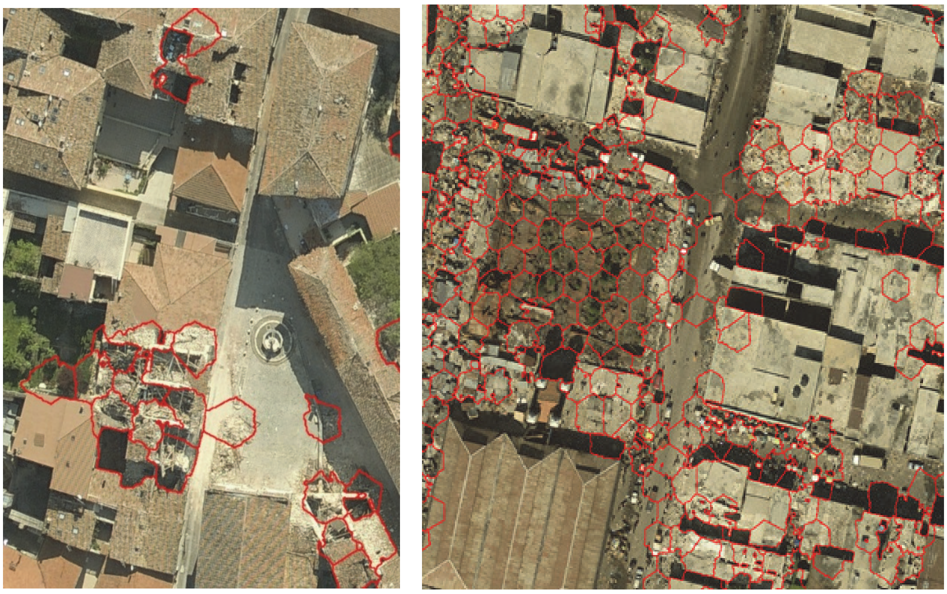


Figure 4-7. Super-pixels of aerial images that were classified as damage by the CNN_F based SVM classifier are highlighted as red polygons in the subset of an aerial image of L'Aquila (left) and Port-au-Prince, (right). © Pictometry.

Figure 4-8. Super-pixels of UAV images that were classified as damage by the CNN_F based SVM classifier are highlighted as red polygons in the subset of UAV images of Kathmandu (left-top), Mirabello (left-bottom) and L'Aquila (right).

4.3.4 Experiment *2:* Transferability of CNN features for damage detection

The transferability of the CNN features was examined by training the supervised model using the image samples from several geographic locations, and tested on the samples from unseen locations. Each of the two datasets consists of samples from six independent geographic locations. For examining the transferability, the supervised model was trained by considering samples from any three out of six geographic locations and independently tested on samples from the remaining three locations. This strategy is followed, because samples from a single site may not be sufficient for constructing the supervised model. For the same reason, the CNN_F approach alone was considered in this experiment for examining the transferability, as the number of training samples constituted from three geographic locations alone may not be sufficient for building the CNN model based on the other two scenarios (CNN_S and

CNN_T). From six locations, in total 20 unique subsets, each containing three locations, can be generated. The supervised models were trained and tested for all 20 subsets independently for each datasets. The results are reported in Table 4-7 and Table 4-8.

Table 4-7. The results of the CNN_F based SVM classifier examined for transferability using $Aerial_{7130}$ dataset. The reported accuracies are based on the classifier trained using the samples from training sites and tested on samples from unseen testing sites. The accuracies less than 80% and their corresponding testing sites are highlighted.

$Aerial_{7130}$ dataset's geographic locations: 1–A_PAP$_{1926}$; 2–A_BID$_{1655}$; 3–A_LAQ$_{988}$; 4–A_ONN$_{1256}$; 5–A_TEM$_{1061}$; 6–A_MIR$_{244}$.				
Training locations	***Testing locations***	***Overall accuracies (%) corresponding to the testing locations***		
1,2,3	4,5,6	91.32	88.37	82.56
1,2,4	3,5,6	88.69	88.09	83.12
1,2,5	3,4,6	91.87	94.42	90.65
1,2,6	3,4,5	86.86	92.97	86.56
1,3,4	**2**,5,6	**74.51**	90.21	87.40
1,3,5	**2**,4,6	81.07	95.24	87.61
1,3,6	**2**,4,5	**72.09**	92.24	89.87
1,4,5	**2**,3,6	**78.84**	93.55	92.71
1,4,6	**2**,3,5	**69.21**	86.20	87.79
1,5,6	**2**,3,4	**73.05**	92.17	95.97
2,3,4	1,5,6	88.68	88.41	88.62
2,3,5	1,4,6	82.29	94.10	81.95
2,3,6	1,4,5	86.18	88.62	91.74
2,4,5	1,3,6	86.68	92.05	82.28
2,4,6	1,3,5	83.12	89.63	87.09
2,5,6	1,3,4	90.49	91.37	91.21
3,4,5	1,**2**,6	87.98	**78.16**	88.25
3,4,6	1,**2**,5	87.41	**71.36**	90.17
3,5,6	1,**2**,4	89.50	**73.06**	93.48
4,5,6	1,**2**,3	89.92	**71.84**	92.03

Table 4-8. The results of the CNN_F based SVM classifier examined for transferability using UAV$_{5414}$ dataset. The reported accuracies are based on the classifier trained using the samples from training sites and tested on samples from unseen testing sites. The accuracies less than 80% and their corresponding testing sites are highlighted.

UAV$_{5414}$ dataset's geographic locations: 1– U_ECU$_{783}$; 2– U_KAT$_{1552}$; 3– U_LAQ$_{385}$; 4– U_PIN$_{989}$; 5– U_MIR$_{1247}$; 6– U_GRO$_{244}$.				
Training locations	***Testing locations***	***Overall accuracies (%) corresponding to the testing locations***		
1,2,3	4,**5**,6	92.69	**68.45**	85.96
1,2,4	3,5,6	84.09	83.32	86.45
1,2,5	3,4,6	83.52	91.49	86.05
1,2,6	3,**4**,**5**	91.01	**76.50**	**68.15**
1,3,4	2,5,6	85.59	83.72	87.62
1,3,5	2,4,6	85.65	90.27	87.12
1,3,6	2,**4**,**5**	80.73	**79.8**	**69.51**
1,4,5	2,**3**,6	83.11	**67.86**	83.20
1,4,6	**2**,3,5	**74.28**	94.63	85.87
1,5,6	**2**,3,**4**	**79.39**	91.01	**78.44**
2,3,4	1,5,6	92.46	91.35	90.44
2,3,5	1,4,6	86.09	90.97	85.96
2,3,6	1,**4**,5	89.88	**76.24**	80.87
2,4,5	1,**3**,6	85.55	**65.06**	82.01
2,4,6	1,3,5	90.30	85.37	93.21
2,5,6	1,**3**,4	89.56	**79.74**	84.45
3,4,5	1,**2**,6	86.47	**78.46**	83.67
3,4,6	1,**2**,5	93.34	**66.56**	94.58
3,5,6	1,**2**,4	90.75	**67.06**	85.64
4,5,6	1,**2**,**3**	87.79	**62.61**	**75.99**

The results show that the supervised model trained using CNN features was highly transferable, achieving an average transferable accuracy of 85%, and a maximum of approximately 95% for both datasets. However, the results were not too generalized as in some cases they produced a transferable accuracy of only around 65%. It was found that transferable accuracy drops to very low values only when a specific geographic location was considered as the testing site. For example, concerning the dataset Aerial$_{7130}$, model transferability accuracy drops whenever the geographic location ***A_BID$_{1655}$*** (2) was considered as the testing site. When examined, it was found that the scene characteristics of ***A_BID$_{1655}$*** (cf. Figure 4-9) differed substantially from the rest of the considered geographic locations (cf. Figure 4-7). Hence, the accuracy might drop because of the poor representativeness of the training samples used for constructing the supervised model. A similar observation was made from the results of ***UAV_{5414}*** dataset, where the accuracy drops when a specific combination of sites was considered in the training and testing process. For example, the accuracy of testing site 2 drops whenever the sites 1 and 3 did not co-occur in the training set. Overall, the results show that the supervised model developed based on CNN features generalized enough to transfer it to classify remote sensing images from unseen sites that vary in illumination conditions, spatial resolution and scene characteristics. However, it was observed that the tuning of a model from site-specific samples remains necessary for improved accuracy, particularly when the considered test site varies substantially in scene characteristics compared to the sites used for training the supervised model.

Figure 4-9 Subset of aerial image of Bidonville, Haiti, used for generating the dataset A_BID_{1655}, where the scene characteristics visually seem different from images of other locations depicted in Figure 4-7

4.3.5 Experiment 3: 3D point cloud features for damage detection alone

As stated earlier, it was anticipated that the combined use of CNN and 3D point cloud features could improve the results by compensating the weakness of one with the strength of another. Towards this, several 3D feature subsets were defined (cf. Table 4-2). Before delving into the final damage detection process using both CNN and 3D features, it is important to infer the independent significance of 3D feature subsets in damage detection and eliminate the insignificant features. This is the objective of this experiment. The ***UAV_{5414}*** dataset was considered for this experiment, as it has a 3D point cloud corresponding to all images in the dataset. Thereby, the 3D feature subsets as mentioned in Table 4-2 were derived for all image samples using the method described in Section 0. The derived features were used to fit a SVM classifier, where the integration of different 3D feature subsets and the estimation of their individual contribution in the classification process were performed using MKL. Two kernels, linear and Gaussian, were used independently in this MKL framework in order to infer the performance of 3D point cloud features when associated with different kernels. The results are reported in Table 4-9.

Table 4-9. Results of the performance of 3D feature subsets in the damage detection for the UAV_{5414} dataset, and the contribution of individual feature subset in the classification process estimated by MKL. The least contributing features are highlighted.

3D feature subsets (cf. Table 4-2)	***Weight β estimated by MKL for each feature subset***	
	Linear kernel	***Gaussian kernel***
FS1: H_L_λ	0.11±0.066	0.12±0.015
FS2: H_P_λ	0.14±0.045	0.15±0.006
FS3: H_S_λ	0.16±0.015	0.17±0.003
FS4: H_O_λ	0.13±0.016	0.11±0.004
FS5: H_A_λ	0.09±0.016	0.08±0.004
FS6: H_E_λ	0.09±0.027	0.07±0.002
FS7: H_C_λ	0.09±0.009	0.04±0.027
FS8: H_N_z	0.23±0.046	0.25±0.001
FS9: S_3D	**0.011±0.01**	**0.01±0.012**
FS10: M_3D	**0.022±0.013**	**0.01±0.003**
Overall accuracy %	75.94±1.41	81.64±1.26
Precision%	78.29±1.29	83.33±1.29
Recall%	73.62±3.3	79.95±3.04

The results in Table 4-9 show that 3D point cloud features possess some useful information for identifying the damaged regions. Also, the choice of kernel for representing 3D point cloud features was observed to be having a significant impact on the classification accuracy (cf. in Table 4-9 the accuracy difference between Gaussian kernel and linear kernel was around 5%). As anticipated, the histogram-based representation of 3D point cloud features was more important than the other two representations FS9 and FS10. In particular the histogram feature representing the distribution of Z component of the normal vector (FS8) was found to be the most significant 3D feature, contributing 23% to the classification accuracy. The other histogram 3D point cloud features were found to be moderately contributing (around 10-16%), while the other non-histogram based 3D point cloud features FS9 and FS10 were insignificant, contributing only around 1%. Hence, the latter two features were not considered for the classification process in the subsequent experiments.

4.3.6 Experiment 4: Integration of CNN and 3D point cloud features using MKL for damage detection

In this experiment the effectiveness of the combined use of CNN and 3D point cloud features for damage detection was analyzed. The histogram-based 3D point cloud features FS 1-8, which were found to be significant, and CNN features based on the CNN_F approach, were the features considered for this experiment. The integration and estimation of individual contribution of these features for the damage detection process was achieved using MKL with a SVM classifier. In the MKL framework, a Gaussian kernel was used for all 3D point cloud features based on the results from Table 4-9, while a linear kernel was used for CNN features, as its dimensionality was already high (1x4096). It is well-known that applying higher order kernels on high dimensional features may lead to overfitting (Han and Jiang, 2014). The ***UAV_{5414}*** dataset containing both images and corresponding 3D point clouds was considered for this experiment, and the results are reported in Table 4-10. It was also analyzed whether the integration of 3D point cloud features along with CNN features for model construction improves the model transferability. This was realized using the same procedure used in Experiment 2 by adopting the MKL framework, and the corresponding results are shown in Table 4-11.

Table 4-11Table 4-10. Results of the performance of integrated use of CNN and 3D feature subsets in the damage detection for the UAV_{5414} dataset, and the contribution of individual feature subset in the classification process estimated by MKL. Significantly contributing features are highlighted.

3D feature subsets (cf. Table 4-2)	***Kernel associated with each feature subset***	***Weight β estimated by MKL for each feature subset***
FS1: H_L_λ	Gaussian	0.039481
FS2: H_P_λ	Gaussian	0.043047
FS3: H_S_λ	Gaussian	0.042455
FS4: H_O_λ	Gaussian	0.026954
FS5: H_A_λ	Gaussian	0.025276
FS6: H_E_λ	Gaussian	0.011519
FS7: H_C_λ	Gaussian	0.038332
FS8: H_N_z	**Gaussian**	**0.118256**
CNN_feature	**Linear**	**0.65468**
Overall accuracy %		94.22±0.85
Precision%		95.66±0.74
Recall%		92.78±1.72

Table 4-11. The results of the transferability of supervised models developed based on the combined use of CNN and 3D feature subsets for the UAV_{5414} dataset. The difference in accuracy of the model developed by CNN+3D point cloud features and CNN alone is given within brackets for reference. The top 5 increase and top 5 decrease in accuracies are highlighted in green and yellow, respectively.

Geographic locations: 1–A_PAP$_{1926}$; 2–A_BID$_{1655}$; 3–A_LAQ$_{988}$; 4–A_ONN$_{1256}$; 5–A_TEM$_{1061}$; 6–A_MIR$_{244}$				
Training locations	***Testing locations***	***Overall accuracies (%) corresponding to the testing locations***		
1,2,3	4,5,6	94.02 (+1.33)	74.95 (+6.5)	87.01 (+1.05)
1,2,4	3,5,6	83.63 (-0.46)	86.95 (+3.63)	86.75 (+0.30)
1,2,5	3,4,6	84.34 (+0.82)	91.65 (+0.16)	86.17 (+0.12)
1,2,6	3,4,5	88.81 (-2.20)	83.47 (+6.97)	71.86 (+3.71)
1,3,4	2,5,6	86.92 (+1.33)	85.67 (+1.95)	88.81 (+1.19)
1,3,5	2,4,6	86.82 (+1.17)	90.30 (+0.03)	88.90 (+1.78)
1,3,6	2,4,5	83.08 (+2.35)	81.58 (+1.78)	73.00 (+3.49)
1,4,5	2,3,6	82.12 (-0.99)	74.98 (+ 7.12)	84.18 (+0.98)
1,4,6	2,3,5	78.00 (+3.72)	93.22 (-1.41)	86.41 (+0.54)
1,5,6	2,3,4	82.59 (+3.20)	91.72 (+0.71)	81.14 (+2.70)
2,3,4	1,5,6	93.73 (+1.27)	91.26 (-0.09)	92.37 (+1.93)
2,3,5	1,4,6	88.04 (+1.95)	92.23 (+1.26)	87.73 (+1.77)
2,3,6	1,4,5	91.66 (+1.78)	81.15 (+4.91)	83.95 (+3.08)
2,4,5	1,3,6	88.33 (+2.78)	69.37 (+ 4.31)	83.70 (+1.69)
2,4,6	1,3,5	91.07 (+0.77)	85.66 (+0.29)	93.25 (+0.04)
2,5,6	1,3,4	90.27 (+0.71)	82.57 (+2.83)	85.86 (+1.41)
3,4,5	1,2,6	88.53 (+2.06)	80.41 (+1.95)	86.47 (+2.80)
3,4,6	1,2,5	91.39 (-1.95)	70.52 (+3.96)	93.57 (-1.01)
3,5,6	1,2,4	91.04 (+0.29)	71.73 (+4.67)	86.35 (+0.71)
4,5,6	1,2,3	89.02 (+1.23)	67.54 (+4.93)	78.11 (+2.12)

The 3D point cloud features which were found to possess some potential in identifying the damaged regions led to an accuracy improvement of 3.1% (i.e. 94.2% to 91.1%) when they were integrated with CNN features (cf. Table 4-6 & 4-10). Moreover, the 3D point cloud features were found to play a significant role in model transferability. For example, in many cases, the use of 3D point cloud features along with CNN features improved the model transferability accuracy by a maximum of 7%, compared to the accuracy previously achieved by CNN features alone. When examining the reasons for accuracy improvement, we observed that the 3D point cloud features help to distinguish between the radiometric pattern discrepancy due to geometric deformation (damage) and other reasons mentioned earlier, thereby controlling the error rates. For example, 8.9% of the errors made by CNN features alone were reduced to 5.8% when including 3D point cloud features i.e. the amount of error was reduced by 35% (cf. Table 4-6 and Table 4-10). This is visually evident

from Figure 4-10, which depicts the classified image where the undamaged and geometrically intact super-pixels (highlighted with blue circles) were classified as damaged (false positive) by the CNN-features based classifier. However, they were classified with a very low confidence rate (prediction score depicting the distance between the sample point and the decision boundary in feature space), indicating a high degree of classification uncertainty. In such cases, the influence of 3D point cloud features was found to be significant, as they help moving these low confidence predictions (samples closer to the margin) into the high confidence region (far away from the margin), thereby improving the accuracy (cf. Figure 4-10). Also, the inclusion of 3D point cloud features along with CNN features lead to improved accuracy by detecting a few damaged regions that were not previously detected by CNN features-based classifier (cf. region highlighted using green circle in Figure 4-10). On the contrary, in some cases, the integration of 3D point cloud features along with CNN features was found to decrease the accuracy of the model transferability achieved previously by CNN features alone. Upon closer examination it was found that façade regions or under-segmented image super-pixels were often wrongly detected as damaged by the 3D geometric features, as they generally look cluttered, consequently possessing varying geometric distribution characteristics. This problem could be eliminated by improving the segmentation quality. Also the photogrammetry point clouds are often noisy, leading to poor geometric representation for some regions in particular façade regions (cf. Vetrivel et al. 2016a), which could also be one of the reasons for the decrease in accuracy.
In addition, we analyzed the significance of each feature when both CNN and 3D point cloud features were used in combination for damage detection. The CNN feature was estimated as the most significant feature by MKL, contributing around 65% among the nine features considered for the classification (cf. Table 4-10). The proportion of contribution of the 3D point cloud features was similar to the inference made in a preceding experiment, where the histogram of Z component of the normal vector was ranked as the most significant feature among other 3D features, contributing around 11%. The remaining seven 3D point cloud features were assigned weights in the range of 1-4% by MKL, where their total contribution was around 24%.

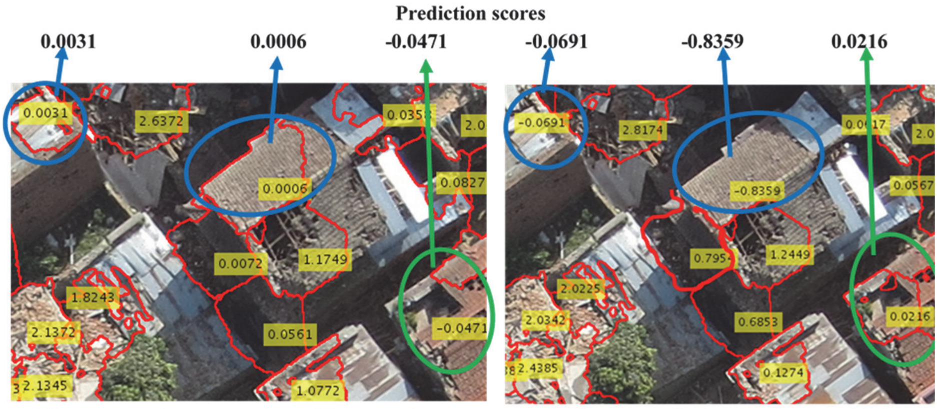


Figure 4-10. The super-pixels of an image which are wrongly predicted as damaged, and missed detections by CNN features based-classifier, are highlighted in blue and green circles, respectively, and annotated with prediction scores ,i.e. distance of the sample to the classification boundary in feature space (left); the improved predictions by CNN+3D features-based classifier (right). The super-pixels with positive and negative prediction scores indicate damaged and undamaged regions, respectively. The prediction scores are in the range [-6.9, 8.2] for the samples in the training sets.

4.4 Overall discussions, conclusions and future works

A framework was developed for the independent and integrated use of CNN and 3D point cloud features for automated identification of damaged regions using images from manned and unmanned aerial platforms. The framework was tested in different scenarios using images from several geographic locations that are highly varying in image and scene characteristics. The major inferences are discussed and summarized below.

CNN features: They were examined for damage detection in three different scenarios: 1) feature learning by training the model from scratch (CNN_S), 2) tuning features of a pre-trained CNN model using domain-specific samples (CNN_T), and 3) directly adopting the features learned by a pre-trained based on images from related domains (CNN_F). There was no significant difference in the performance of the features learned by three different scenarios, as they all produced an accuracy of 90±3%. However, the CNN_F approach is highly recommended as it requires a lower number of training samples and is light-weight in computation. For this study we selected the pre-trained model developed based on classical images, as there was no pre-trained model specific to the remote sensing domain readily available. However, it was observed that the pre-trained model tuned with domain-specific samples produced slightly better accuracy than the other two approaches. Based on this observation, we expected that creating a pre-trained model specific to the remote sensing domain (i.e. training CNN models with remote sensing images from various remote sensing applications) could learn the useful domain-specific features, further adopting this model for feature extraction might improve the accuracy.

3D point cloud features: The Z of plane normal and several structure tensor-based 3D features described in Table 4-1 were examined for the damage detection process using the $\boldsymbol{UAV_{5414}}$ dataset. The major objective was to analyze the significance of these features before using them together with CNN features for the classification process. We anticipated that the histogram-based representation of 3D point cloud features inspired from HoG would be more effective for capturing the segment-level geometric characteristics than other conventional representations as described in Section 0. To examine this, 10 feature subsets were framed as described in Table 4-2, in which eight features subsets were based on the histogram approach and the remaining two were based on other two approaches. The significance of each 3D feature subset was estimated using the MKL approach. As expected the histogram-based 3D point cloud features were found to be the most significant features, contributing 97% towards the classification accuracy of 81% (cf. Table 4-9). Also, the choice of kernel in the MKL framework for representing 3D feature was found to have a significant impact on the classification accuracy: the Gaussian kernel produced superior results over the linear kernel.

Integrated use of CNN and 3D features: The CNN and 3D feature subsets were integrated and classified using SVM by adopting the MKL framework. The integration of 3D point cloud features with CNN produced only a marginal improvement in terms of accuracy, from 91% to 94% (cf. Table 4-6 and Table 10). There was no significant difference in the predictions made by CNN when used alone or together with 3D features, because the MKL assigned a weight of 65% to CNN features, signaling its superiority and enabling it to strongly dominate the influence of 3D point cloud features in the classification process. Hence, only the highly uncertain samples (often the false positives) in the feature space of CNN could have been influenced by the 3D point cloud features when integrated with CNN, leading to a marginal accuracy improvement.

Model transferability: The transferability of the supervised models based on CNN features was examined by training the model based on images from specific geographic locations, and tested on images from unseen sites that were slightly varying in scene characteristics, illumination conditions, spatial resolution and radiometric characteristics. Often the CNN based supervised models were found to generalize well and produce an average accuracy of around 85% for the experiments conducted for analyzing the model transferability. In some cases, the model transferability accuracy dropped when the training and testing samples were from highly varying scene characteristics, highlighting the importance of site-specific samples for achieving improved results. The inclusion of 3D point cloud features along with CNN for model construction often was found to be significantly increasing the transferability of the model. In a few cases, the inclusion of the 3D point cloud features degraded the accuracy of the model due to the earlier mentioned reasons such as under-segmentation and presence of high noise in the point cloud. The under-segmentation issues could be largely resolved by using 3D point cloud features along with image features for segmentation (Vetrivel et al., 2015b). We also expected that instead of using hand-crafted 3D features, learning useful 3D point cloud features directly from the 3D point cloud using CNN might be effective. However, it is not possible to adopt CNN directly for 3D point clouds, as the used CNNs are specially designed for data in matrix format, such as images. However, in other domains, where the data are not in matrix format, such as text mining and software security, the CNN approach has been widely used for feature extraction by converting their data originally in non-matrix representations to matrix representation (Kim, 2014; Narayanan et al., 2016a). For example, the recently proposed concept word2vec for representing text in matrix format has been used for adopting CNN for text mining applications (Fu et al., 2016). It has been reported that the CNN features learned specifically for text or graph data are reported to be superior over hand-crafted features (Kim, 2014; Narayanan et al., 2016a). Similarly, designing an approach for representing 3D point clouds in matrix format to use CNN for learning the application- and data-specific 3D features might outperform the conventional hand-crafted features as observed in several studies from related domains. We intend to explore this in the future.

Overall, the proposed framework for training supervised model based on CNN features either combined with 3D point cloud features using MKL or alone was found to be significant for detecting damaged regions using very high resolution images. Particularly, the accuracy obtained on model transferability was encouraging, where the model trained with a sufficient number of samples can be transferred to an actual disaster scenario without any retraining for a quick assessment of damaged regions. This would be highly beneficial to the first responders for speedy response activities, avoiding the tedious manual referencing.

4.5 References of chapter 4

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

Arivazhagan, S., Ganesan, L., Priyal, S.P., 2006. Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters,* 27 (16), pp. 1976-1982.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35 (8), pp. 1798-1828.

Bucak, S.S., Jin, R., Jain, A.K., 2014. Multiple kernel learning for visual object recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 36 (7), pp. 1354-1369.

Cao, Z., Principe, J.C., Ouyang, B., 2015. Group feature selection in image classification with multiple kernel learning, *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-5.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land use classification in remote sensing images by convolutional neural networks, *arXiv:1508.00092 [cs.CV]*.

Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing,* 117, pp. 11-28.

Dell'Acqua, F., Gamba, P., 2012. Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE,* 100 (10), pp. 2876-2890.

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.

Fernandez Galarreta, J., Kerle, N., Gerke, M., 2015. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. *Natural Hazards and Earth System Sciences,* 15 (6), pp. 1087-1101.

Ferraz, C.T., Pereira, O., Rosa, M.V., Gonzaga, A., 2014. Object recognition based on bag of features and a new local pattern descriptor. *International Journal of Pattern Recognition and Artificial Intelligence,* 28 (08), pp. 1-32.

Fu, H., Niu, Z., Zhang, C., Ma, J., Chen, J., 2016. Visual cortex inspired CNN model for feature construction in text analysis. *Frontiers in Computational Neuroscience,* 10 (64), pp. 1-10.

Gerke, M., Kerle, N., 2011. Automatic structural seismic damage assessment with airborne oblique Pictometry© imagery. *Photogrammetric Engineering & Remote Sensing,* 77 (9), pp. 885-898.

Gevaert, C.M., Persello, C., Sliuzas, R., Vosselman, G., 2016. Classification of informal settlements through the integration of 2D and 3D features extracted from UAV data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* III-3, pp. 317-324.

Golovinskiy, A., Kim, V.G., Funkhouser, T., 2009. Shape-based recognition of 3D point clouds in urban environments, *IEEE 12th International Conference on Computer Vision*, pp. 2154-2161.

Gönen, M., Alpaydın, E., 2011. Multiple kernel learning algorithms. *Journal of Machine Learning Research,* 12 (Jul), pp. 2211-2268.

Gu, Y., Wang, Q., Jia, X., Benediktsson, J.A., 2015. A novel MKL model of integrating LIDAR data and MSI for urban area classification. *IEEE Transactions on Geoscience and Remote Sensing,* 53 (10), pp. 5312-5326.

Hackel, T., Wegner, J.D., Schindler, K., 2016. Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic,* 3, pp. 177-184.

Han, H., Jiang, X., 2014. Overcome support vector machine diagnosis overfitting. *Cancer Informatics,* 13 (Suppl 1), pp. 145-158.

Hu, F., Xia, G.-S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing,* 7 (11), pp. 14680-14707.

ISPRS_Benchmark, 2016, ISPRS semantic labeling contest. http://www2.isprs.org/vaihingen-2d-semantic-labeling-contest.html (Accessed 04/11/2016).

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725-1732.

Katz, S., Tal, A., Basri, R., 2007. Direct visibility of point sets. *ACM Transactions on Graphics (TOG)*, 26 (3), pp.

Kerle, N., Hoffman, R.R., 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural hazards and earth system sciences (NHESS),* 13 (1), pp. 97-113.

Khoshelham, K., Oude Elberink, S., Xu, S., 2013. Segment-based classification of damaged building roofs in aerial laser scanning data. *IEEE Geoscience and Remote Sensing Letters,* 10 (5), pp. 1258-1262.

Kim, Y., 2014. Convolutional neural networks for sentence classification, *arXiv:1408.5882v2 [cs.CL]*

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, pp. 1097-1105.

Li, X., Yang, W., Ao, T., Li, H., Chen, W., 2011. An improved approach of information extraction for earthquake-damaged buildings using high-resolution imagery. *Journal of Earthquake and Tsunami,* 05 (04), pp. 389-399.

Ma, Y., Chen, F., Liu, J., He, Y., Duan, J., Li, X., 2016. An automatic procedure for early disaster change mapping based on optical remote sensing. *Remote Sensing,* 8 (4), pp. 272.
MatConvNet, 2016, CNN pretrained models. http://www.vlfeat.org/matconvnet/pretrained/, (Accessed 12/10/2016).
Narayanan, A., Chandramohan, M., Chen, L., Liu, Y., Saminathan, S., 2016. Adaptive and scalable android malware detection through online learning, *arXiv:1606.08928 [cs.LG]*.
Nex, F., Remondino, F., 2014. UAV for 3D mapping applications: a review. *Applied Geomatics,* 6 (1), pp. 1-15.
Rakotomamonjy, A., Bach, F.R., Canu, S., Grandvalet, Y., 2008. SimpleMKL. *Journal of Machine Learning Research,* 9 (Nov), pp. 2491-2521.
Samadzadegan, F., Rastiveisi, H., 2008. Automatic detection and classification of damaged buildings, using high resolution satellite imagery and vector data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 37, pp. 415-420.
Sherrah, J., 2016. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery, *arXiv:1606.02585 [cs.CV]*.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9.
Tang, Y., 2013. Deep learning using linear support vector machines, *arXiv preprint arXiv:1306.0239*.
Tian, D.P., 2013. A review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering,* 8 (4), pp. 385-396.
Tu, J., Sui, H., Feng, W., Song, Z., 2016. Automatic building damage detection method using high-resolution remote sensing images and 3D GIS model. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* III-8, pp. 43-50.
Velizhev, A., Shapovalov, R., Schindler, K., 2012. Implicit shape models for object detection in 3D point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* I-3, pp. 179-184.
Vetrivel, A., Duarte, D., Nex, F., Gerke, M., Kerle, N., Vosselman, G., 2016a. Potential of multi-temporal oblique airborne imagery for structural damage assessment. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* III-3, pp. 355-362.
Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2015a. 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, pp. 61-78.
Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2015b. Segmentation of UAV-based images incorporating 3D point cloud information. *The International*

Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-3/W2, pp. 261-268.
Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2016b. Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach. *Remote Sensing,* 8 (3), pp. 231.
Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing,* 105, pp. 286-304.
Weinmann, M., Mallet, C., Bredif, M., 2016. Detection, segmentation and localization of individual trees from mms point cloud data, *In: GEOBIA 2016 : Solutions and Synergies.*, University of Twente Faculty of Geo-Information and Earth Observation (ITC) p. 8.
Xu, S., Fang, T., Li, D., Wang, S., 2010. Object classification of aerial images with bag-of-visual words. *IEEE Geoscience and Remote Sensing Letters,* 7 (2), pp. 366-370.
Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, pp. 818-833.
Zhou, P., Cheng, G., Liu, Z., Bu, S., Hu, X., 2015. Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. *Multidimensional Systems and Signal Processing*, pp. 1-20.
Zuo, Z., Wang, G., Shuai, B., Zhao, L., Yang, Q., Jiang, X., 2014. Learning discriminative and shareable features for scene classification, *European Conference on Computer Vision*. Springer, pp. 552-568.

5 Potential of multi-temporal oblique airborne imagery for structural damage assessment*

* This chapter is based on the article:
Vetrivel, A., Duarte, D., Nex, F.C., Gerke, M., Kerle, N. and Vosselman, G. (2016) Potential of multi-temporal oblique airborne imagery for structural damage assessment. In: Proceedings of the XXIII ISPRS Congress: From human history to the future with spatial information, 12-19 July 2016, Prague, Czech Republic. Peer reviewed Annals, Volume III-3, 2016 / edited by L. Halounova, ... [et al.]. ISPRS, 2016. ISSN: 2194-9050. pp. 355-362.

Abstract

Quick post-disaster actions demand automated, rapid and detailed building damage assessment. Among the available technologies, post-event oblique airborne images have already shown their potential for this task. However, existing methods usually compensate the lack of pre-event information with aprioristic assumptions of building shapes and textures that can lead to uncertainties and misdetections. However, oblique images have been already captured over many cities of the world, and the exploitation of pre- and post-event data as inputs to damage assessment is readily feasible in urban areas. In this chapter, we investigate the potential of multi-temporal oblique imagery for detailed damage assessment by developing a methodology for detecting severe structural damages related to geometrical deformation by combining the complementary information provided by photogrammetric point clouds and oblique images. The developed methodology detected 87% of damaged elements. Most of the failed detections are due to varying noise levels within the point cloud which hindered the recognition of some structural elements.

5.1 Introduction & related works

Structural damage assessment is an imperative process to be carried out immediately after the disaster event for effective planning and execution of response and recovery actions. Assessing building damages over large areas affected by hazard events with ground observations is not efficient. Alternatively, remote sensing-based approaches have been recognized as useful means for assessing synoptic building damage. Detailed information of an affected area can be provided in a short time using a variety of sensors such as optical, SAR and LiDAR (Khoshelham et al., 2013a; Miura et al., 2013; Uprety and Yamazaki, 2012). In particular airborne oblique images have been recognized as a valuable data source to assess building damages because, compared to traditional nadir views, they allow the complete inspection of the external outlines of the building, such as roofs and façades (Murtiyoso et al., 2014). Nowadays, airborne images are captured with high overlap, and the generated point clouds can be exploited in the damage assessment process as well (Sui et al., 2014b). Geometrical deformations such as partial/complete collapse, pancake collapse, inclination, broken and dislocation of elements can be derived by 3D geometric information, while damages such as cracks and spalling can be inferred from the images directly. Several papers have highlighted the potential of synergistic use of 3D point cloud and images for building damage assessment (Gerke and Kerle, 2011a; Vetrivel et al., 2015a). However, only few studies have looked at the use of digital oblique aerial imagery for structural damage assessment, and were focused on (mono-temporal) post-event information (Gerke and Kerle, 2011a; Vetrivel et al., 2015a). The major limitation of this approach is that damage is inferred based on a set of ontological assumptions: i.e. a surface with unusual radiometric or geometric characteristics is assumed to be damaged, while manmade objects are assumed to have a regular shape and uniform radiometric characteristics. These assumptions have limitation in complex environments, leading to a high rate of false alarms, which reduces their reliability and operational utility. In Vetrivel et al. (2015a), damages presenting regular and uniform shapes (false negative), or intact regions characterized by cluttered and non-uniform radiometric distributions (false positive), were incorrectly classified due to these assumptions.

The above uncertainties can be alleviated if pre-event data are available for reference. Many studies have demonstrated the potential of multi-temporal data for damage assessment, though with most focusing on nadir-view images (Dong and Shan, 2013; Murtiyoso et al., 2014). To our knowledge, no methods have been reported yet for identifying building damages using multi-temporal oblique images and/or 3D point clouds.

In this chapter the first implementation of an automated algorithm for building damage assessment from multi-temporal oblique images is presented. Although geometrically more stable cameras are used nowadays in oblique airborne systems, many data sets are captured with less sophisticated camera systems, and image overlap is often restricted to 2-fold. Hence, for such configurations one has to cope with dense image matching point clouds of minor quality (relatively large random error margin, gaps). The proposed methods take advantage of both 2D and 3D information and efficiently cope with these problems.

Thus this chapter focuses on developing methods to identify severe building damages related to geometrical deformation, using multi-temporal oblique images and 3D point clouds. To this end three different change detection methods are proposed to identify building elements that are geometrically deformed between the two epochs. Subsequently, a change classification method is proposed to identify the geometric deformation of an element marked as damaged. The detailed description of the methodologies and the results achieved on the test area of L'Aquila (Italy) will be presented in detail in subsequent sections.

5.2 Data description

The data used are corresponding to the city of L'Aquila, Italy in which an earthquake occurred on 6th April 2009. The data consist of two airborne oblique acquisitions (August 2008 and May 2009) covering the city with both oblique (4 cameras) and nadir (1 camera) imagery, captured by small format DSLR cameras. Images were acquired at a flying height of approximately 1000 m allowing for an average ground sampling distance of 16 cm on oblique views. The flight was conducted considering a forward overlap between 60-70% and side overlap between 35-45%, allowing to derive a 3D point cloud. The registration was achieved computing tie-points from all the imagery, forcing both epochs to share a local coordinate system. Dense image matching was then performed separately on both epochs.

5.3 Methodology for Point cloud based damage detection and classification

Any severe structural damage, such as partial or complete collapse, pancake collapse, dislocated or inclined elements, leads to the absence of the given elements in their actual 3D boundary in the pre-event data. These damaged elements are referred to missing elements. The missing structural elements can be identified by comparing the accurately co-registered multi-temporal (pre and post-event) 3D point clouds. However, the absence of pre-event segments in post-event 3D point clouds may be due to many reasons, such as occlusion/a building part being exposed to a lower number of camera views, or poorly textured surfaces, leading to missing 3D points. On the other hand, an element can be missing because it was damaged or cleared deliberately. Therefore, it is important to infer the reasons for the absence of a pre-event element in the second epoch data after detecting them.

The missing elements due to damage are detected using the pre- and post-event images and 3D point clouds derived from them by three pipeline processes (cf. Figure 5-1). As an initial step, the individual buildings in the area are delineated from the pre-event 3D point cloud. Subsequently, the delineated buildings in the pre-event data are compared to the post event 3D point cloud to detect changes. Finally, the changes are classified to isolate the changes caused by damage. The detailed methodology to carry out the aforementioned processes is described below.

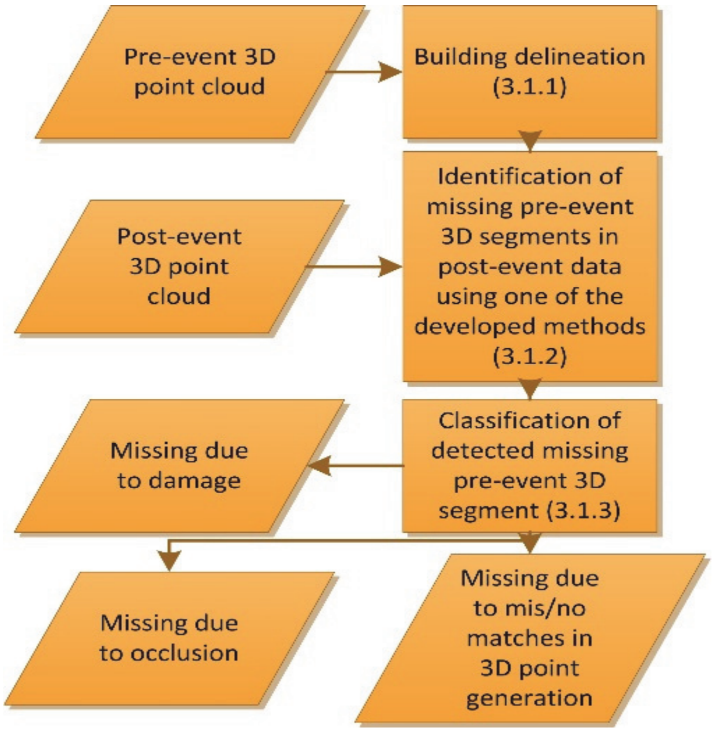


Figure 5-1. Overall workflow

5.3.1 Building delineation from 3D point cloud:

In remote sensing data, the roof segments are the most visible and least cluttered building elements. Moreover, they can be recognised based on simple geometric constraints. For instance, the roof segments are mostly elevated horizontal or slanted planar surfaces with respect to ground. Hence, in this study the roof-based building delineation approach based on 3D point clouds is adopted. Moreover, this method has previously shown reliable results (Vetrivel et al., 2015a). The procedure followed is described below.

The 3D point cloud is segmented using planar segmentation as described by Vosselman (2012). The segments with Z component of the plane normal > T_Z (0.4) and above height T_H (3 m) are labelled as roof segments. The spatially connected roof segments are identified by defining the alpha shapes with an alpha radius of T_R (0.3). Finally, the 3D points that are covered by single alpha shapes are delineated as roofs of a single building. Also, all the 3D points that lie within the 2D boundary of the alpha shape are registered as the 3D points of the building, i.e. all 3D points that lie below the roof elements are also registered as the 3D points of the building.

5.3.2 Change detection to identify the missing building elements in post event:

The pre-event building elements that are missing in the post-event point cloud are identified using the following three approaches. To provide a clear understanding of the proposed methods, they are illustrated with examples from the data used and results obtained from our experimental study. Also, the pros and cons, and appropriate scenario that make a particular method suitable, are discussed.

5.3.2.1 Voxel-based approach (VBA):

A unit of 3D space that is occupied by a specific building element in the pre-event but not the post-event data is straight-forwardly identified and classified as a change. The 3D space defined by the delineated building in the pre-event point cloud is divided into voxels. The edge length of voxels is defined based on image ground resolution (0.5 m ~ three times the image ground resolution). The 3D points of the pre- and post-event point clouds that lie within the defined 3D boundary are added to the corresponding voxels. The voxels that contain pre-event but no post-event 3D points are classified as changed voxels. However, it is challenging to differentiate between the changed and unchanged voxels as there is a high probability that 3D points of the pre- and post-event epochs of the same building element may fall into different (adjacent) voxels, due to the varying noise level between the two point clouds. To overcome this problem the spatial buffers along the horizontal and vertical directions with buffer thresholds of T_{HB} (0.5 m) and T_{VB} (1.0 m), respectively, are created for each voxel occupied with the pre-event 3D points. These voxels are classified as unchanged voxels if post-event 3D points lie within the buffer area; otherwise they are classified as changed voxels. Finally, the pre-event 3D points of the changed voxels are detected as 3D points of missing building elements and considered for the further change classification process. The overall process of this approach is shown in Figure 5-2.

Pros: In many cases photogrammetric point clouds are very noisy, hindering the recognition of individual building elements. In such cases, this approach is more suitable as it does not need any prior information or assumptions about the building, unlike the segment-based approaches discussed below.

Cons: The presence of artefact 3D points, common in photogrammetric point clouds in unfavourable image flight configurations, strongly affect the performance of this approach. Moreover, in specific scenarios, it cannot detect or accurately delineate the missing portion of the element, since the voxels are classified in a binary fashion, i.e. whether or not they contain a certain element. Therefore, even if only a minor portion of a voxel is occupied by an element, it will be classified as occupied voxel. Hence this results in a failure to detect missing elements that characterises the remaining majority portion of the voxel. Since it does not consider the geometry of the elements it may fail to detect the damages in specific scenarios. For example, consider a horizontal roof segment is missing and the below vertical elements are visible. In such a case, this approach may fail as the 3D points of the vertical elements fall into the voxels of the missing elements when the voxel and its buffer sizes are significantly large. Moreover, this approach is computationally intensive.

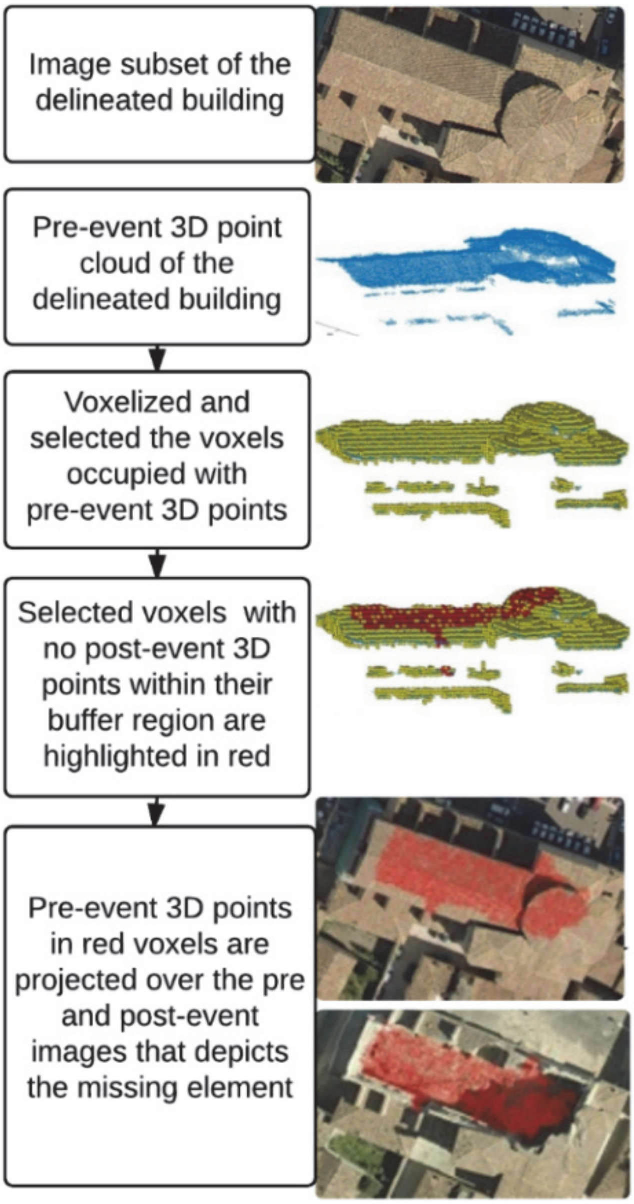


Figure 5-2. Workflow of voxel-based approach

5.3.2.2 Segment-based approach (SBA):

In general, most buildings are made of a composition of planar segments. Hence, we anticipate that comparison of pre- and post-event data based on planar segments will help to precisely identify changes on element level.

This would also lead to more object-oriented analysis compared the voxel-based approach. However, direct comparison of 3D segments that are obtained from independent segmentation of pre and post-event is not always practically possible. This is because, though the 3D point clouds of the two epochs are segmented using the same segmentation algorithm, in practice it is not always feasible to obtain the same segments even for corresponding undamaged areas. This is predominantly due to varying noise level between point clouds. Therefore, we propose a method where the pre-event point cloud alone is segmented and the corresponding segments in the post-event are derived based on the pre-event segments. This is done by fitting a plane to pre-event 3D segment and, subsequently, the post-event 3D points that lie within the plane-offset of T_D (1.0 m) to be derived as post-event segment. Then the missing (damaged) portions of pre-event segments are identified by comparing them with the corresponding derived post-event segments. Only the segments with an area greater than T_A (5 m^2) are considered for the change detection process. The overall procedure of this approach is shown in Figure 5-3.

Pros: Compared to the voxel-based approach it is 1) less sensitive to artefacts, 2) more robust in detection and accurate delineation of missing elements, and 3) computationally less demanding.

Cons: This approach fully depends on the performance of the planar segmentation algorithm. The planar segmentation fails to produce accurate segmentation for portions of point cloud corresponding to very high noise, and also for regions of non-planar elements.

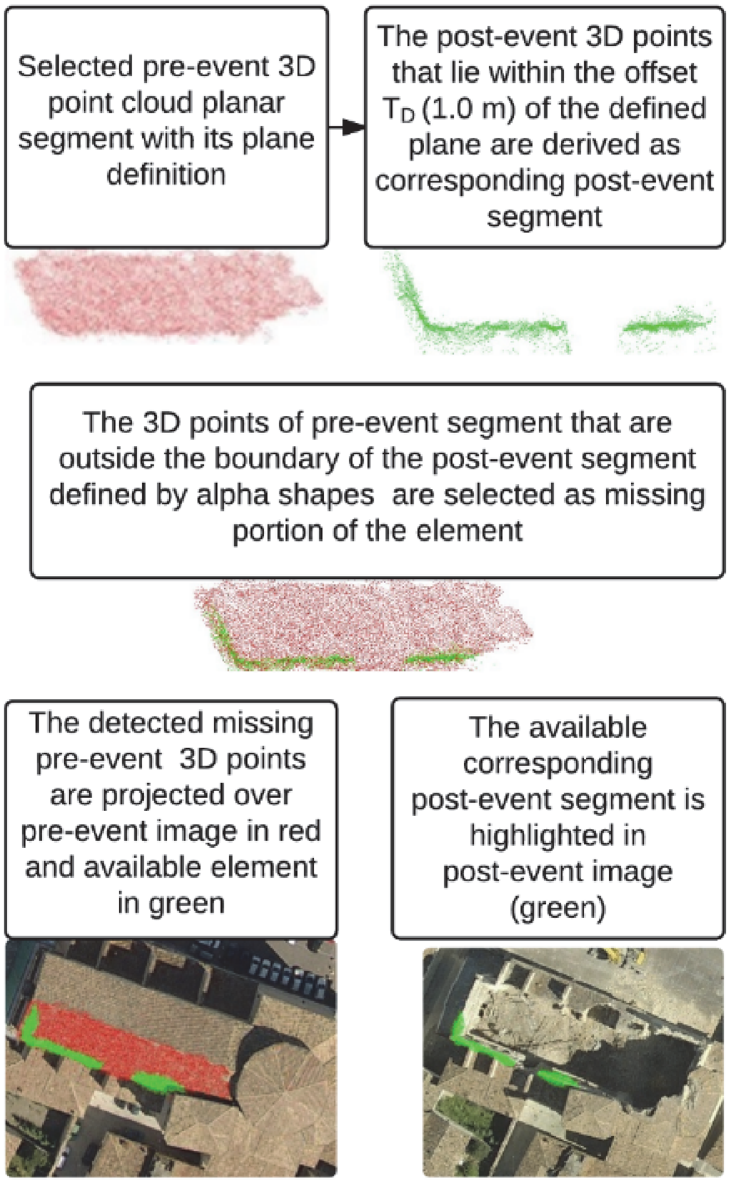


Figure 5-3. Overall workflow of segment-based approach

5.3.2.3 Composite segment-based approach (CSBA):

The above segment-based approach is defined particularly for plane-based segmentation. Here, we develop an alternative segment-based approach that is independent of the segmentation methods. Pertaining to this, a composite segment-based approach is developed where the pre and post-event point clouds are merged and segmented together using a plane-based segmentation algorithm (although any segmentation method can be used). Therefore, the corresponding segments in the two dataset share the same segment label, which facilitates direct comparison between them. Thereby the completely and partially missing elements can be identified in simple and faster way than in the above proposed methods. The proposed change detection strategy is described below.

- The pre-event segments that have no corresponding segment in the post-event point cloud represent the completely missing segments. This can be obtained in a single step by a simple set difference: i.e. suppose sets *A* and *B* are the segment label lists of pre and post-event epoch, respectively, then the completely missing segments are obtained by *A*-*B* (set difference).
- To identify the partially missing segments select the corresponding segments in the pre- and post-event data and define the boundary for the post-event 3D segment using the alpha shapes. The pre-event 3D points that lie outside the defined boundary are considered as the missing portion of the pre-event segment in the post-event data.

Pros: It is faster and simple than the other two proposed methods.

Cons: The choice of segmentation algorithm is critical as it should detect building elements of different geometry (planar and no-planar) and robust to varying noise levels between the point clouds. It is also highly vulnerable to co-registration errors of multi-temporal point clouds.

5.3.3 Change classification (inferring reason for change):

In general, the disappearance of a building element due to damage will lead to two kinds of scenarios: 1) the absence of certain elements will create an opening, leading to a visibility of the element below it. Therefore, there will be a new surface (3D points) in the post-event data, corresponding to this disappeared element (cf. Figure 5-8Figure 5-4), or 2) the disappearance of partial elements may create a hole in the structural element which appear darker due to poor radiometric reflection (cf. Figure 5-4). This is termed as structural holes and it will be a gap in the 3D point cloud. However, gaps in 3D point clouds may also be caused by mis/no matches in 3D point generation (e.g., poorly textured surfaces) and occlusions. Therefore, it is important to distinguish between these gaps to identify the ones caused by damage. As the image radiometry plays a major role here, we used pre- and post-event images in addition to point clouds for the change classification process. The procedure for inferring the reasons for the absence of pre-event element in post-event is described below.

Figure 5-4 Example of element collapses leading to an opening with the surface below it visible (left) and not visible, i.e. structural holes (right) are highlighted in red circles.

Case 1: Element missing due to occlusion/partial visibility:

The pre-event 3D points missing in the post-event data due to occlusion can be identified by analysing the visibility of those points in the post-event camera views. This is done by adding the missing pre-event 3D points to the post-event point cloud. Subsequently, the visibility of 3D points in each post-event camera view is estimated using the Hidden Point Removal operator (HPR) (Katz et al., 2007a). Finally, the newly added pre-event 3D points that are visible in less than ‘N (3)’ camera views (post-event cameras) are removed by labelling them as occluded points, and the remaining pre-event 3D points are classified as visible points and considered for further change classification process.

Case 2: Element missing due to damage and mis/no matches in 3D point generation: The change classification is preferred to be performed at segment level, as manmade objects are largely composed of planar/regular elements, thus it helps to reduce false decisions. The segment-based approaches will directly provide the 3D points of missing element in terms of segments. However, for the voxel-based approach, it will provide the collection of independent pre-event 3D points that are absent in the post-event situation. Therefore, the 3D points obtained from the latter are grouped into dis-jointed 3D segments based on their spatial connectivity, using the alpha shapes for further processing.

A rule-based approach is adopted and a set of rules is defined to classify the missing elements into the aforementioned scenarios. This classification approach is independent from the methods used to detect missing elements and is described below.

Element missing due to damage and the surface below it is visible: The rule for this class is defined below and illustrated in Figure 5-5.

Rules: The post-event 3D points should be present within the 2D boundary (change boundary) of the missing pre-event 3D segment. Also the post-event points and pre-event 3D segment should be visible in the same post-event camera. Also the area covered by the post-event 3D points should be of similar size compared to the area of the missing pre-event segment (at least cover 30 %).

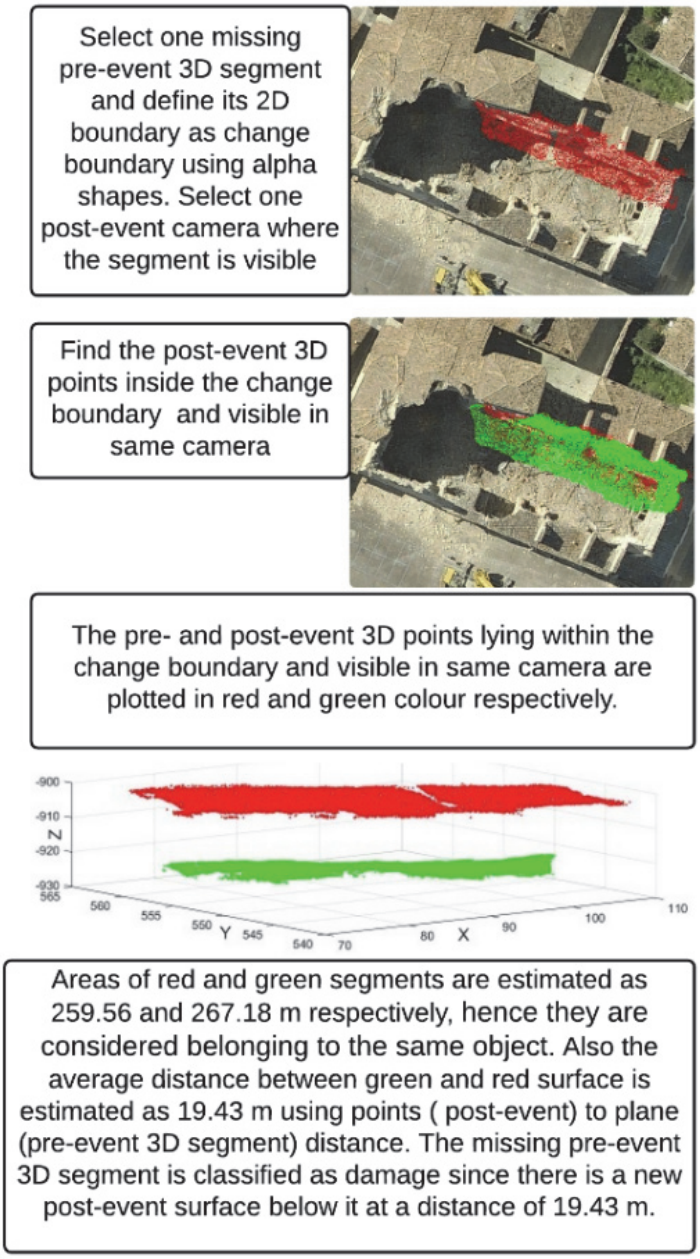


Figure 5-5. Example for missing 3D segment classified as damaged and the surface below it is visible in post-event.

Pre-event element missing in post-event due to damage (structural hole) or mis/no matches in 3D point generation: In contrast to the above scenario, if the change-boundary does not contain a significant number of post-event 3D points then the radiometric characteristics of the change boundaries in the pre- and post-event data are compared by delineating them in the corresponding images. If they are not similar and if the post-event image surface appears significantly darker compared to the pre-event one, then it is classified as damage (structural holes), otherwise it is classified as mis/no matches in 3D point generation. The grey-value (image pixel value) based histogram is used as a feature to compare the radiometric characteristics as it is well proven and widely used region-level image descriptors (Wenjing et al., 2006).

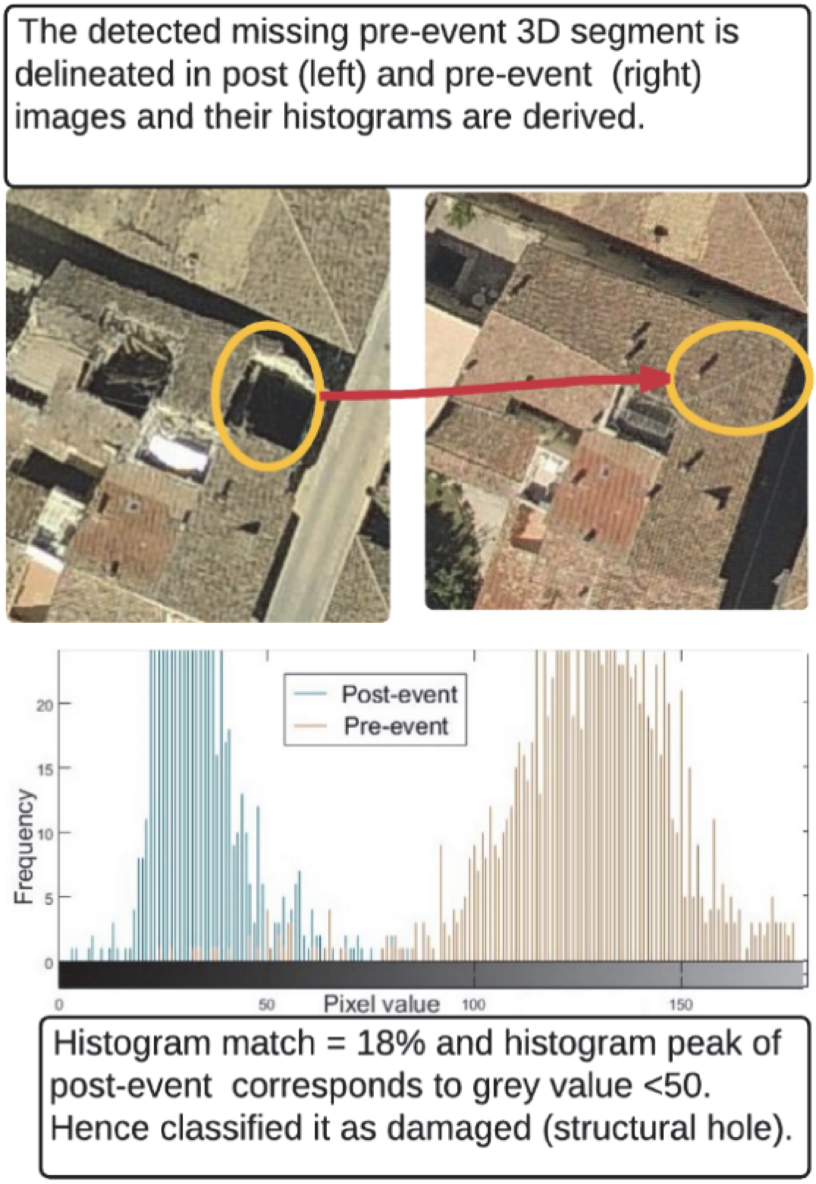


Figure 5-6. Example for missing 3D segment classified as (structural hole) caused by damage.

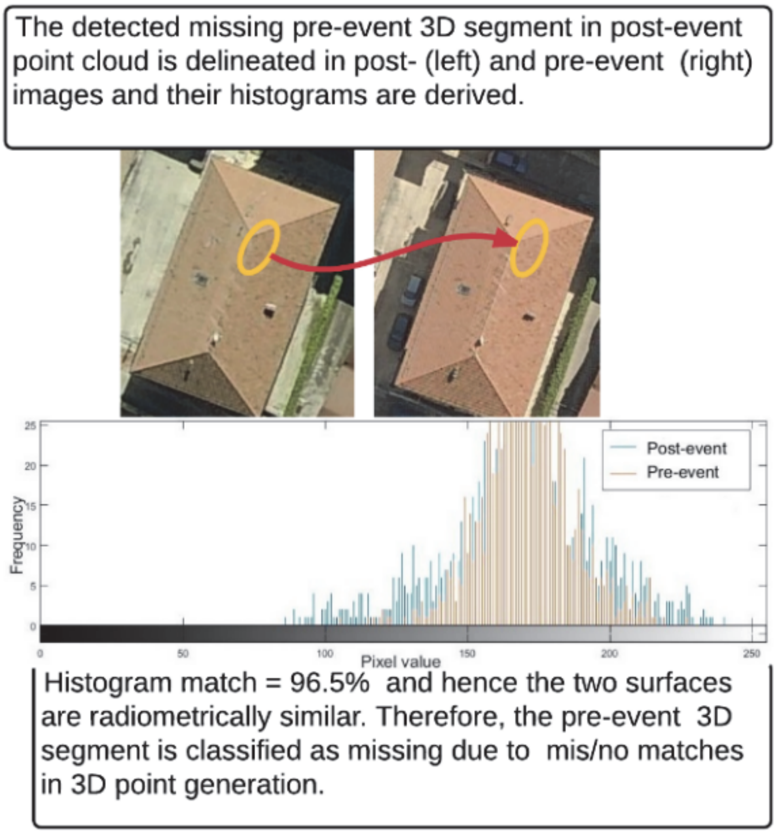


Figure 5-7. Example for missing 3D segment classified as mis/no matches in 3D point generation.

Rules*:** If the correlation between the histograms of the pre and post-image regions is less than T_{HD} (0.75) they are considered to be radiometrically dissimilar. If the histogram peak of the post-event image patch corresponds to grey values $<T_G$ (50 lower grey value) then the pre-event 3D segment is classified as ***element missing due to damage (structural holes) in the post-event. If above constraints are not satisfied then the missing pre-event element is classified as ***element missing due to mis/no matches in 3D point generation.*** The illustrations of these two classes are depicted in Figure 5-6 and Figure 5-7.

5.4 Results

5.4.1 Data used:

Three subsets of building blocks from different parts of the city were considered for the damage detection process, containing a total of 48 buildings. Of those 23 building elements are identified as either completely or partially missing in the post event data due to damage. The pre-event 3D point clouds of the considered sub-blocks are shown in Figure 5-8.

Figure 5-8. Pre-event 3D point clouds of the sub-blocks considered for damage assessment.

5.4.2 Results of building delineation:

All 48 buildings are detected using the method described in section 0 and they are categorized into three cases: 1) Buildings that were delineated with close approximation to the actual boundary (# 36); 2) buildings with some portions not delineated (# 7); 3) single buildings detected as multiple buildings, particularly the buildings with multi-level roofs. (#5). A sample result for building delineation is shown in Figure 5-9.

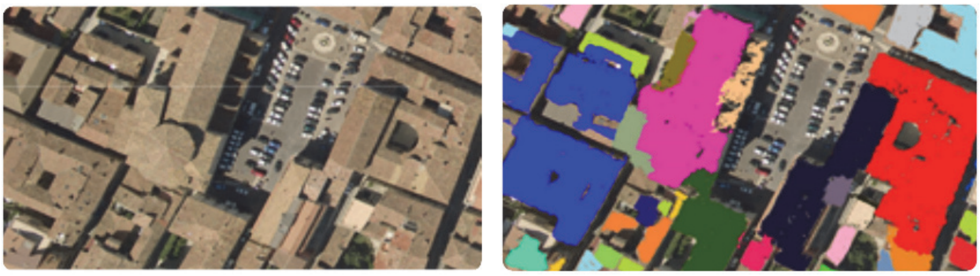


Figure 5-9. Image subset of airborne image (left) and delineated buildings based on 3D point cloud are projected over the image (right).

5.4.3 Results of the change detection methods to identify the missing building elements in post event:

In this section, the overall results and the major inferences associated with each change detection method are summarized. The overall results are provided in Table 5-1. An example result of missing elements that were detected by the composite segment-based approach for a small block is depicted in Figure 5-10.

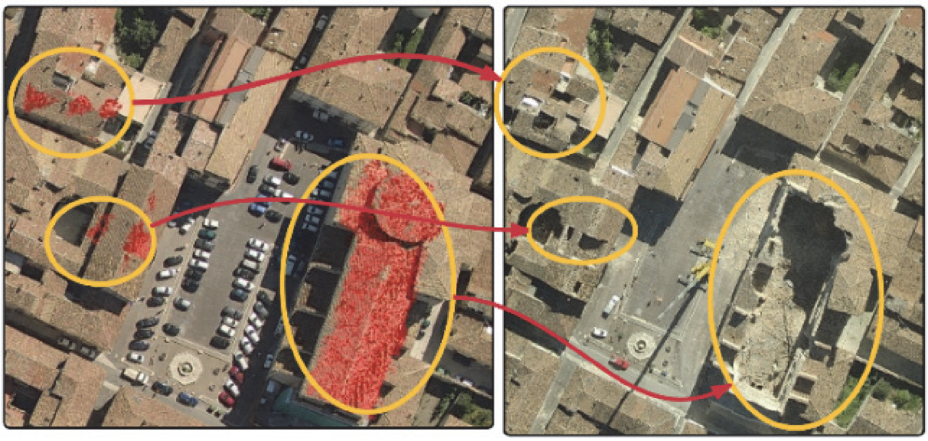


Figure 5-10. The detected missing pre-event segments using composite segment-based approach are projected as red points over a pre-event image (left), and outlined in the post-event (right) images with yellow circles.

Table 5-1. The results of missing pre-event 3D segments detected by all three approaches

Method	Detected damage (not detected)	Missing elements due to:	
		Occlusions	Mis/no matches
VBA	17 (6)	3	8
SBA	20 (3)	1	3
CSBA	20 (3)	1	2

5.4.4 Results of change classification:

For the change classification process, the missing elements identified by the segment-based approach were considered (cf. Table 5-1).

Case 1: Element missing in the post-event data due to occlusion: only one segment that was very small (area < $5m^2$) was identified as missing in the post-event data due to occlusion (cf. Table 5-1), and it was classified correctly and removed from the further classification process.

Case 2: Element missing due to damage and mis/no matches in 3D point generation: The results of the change classification process are provided in Table 5-2. In total all 20 elements that are missing due to damage were classified correctly. Among the three non-damaged missing elements two were correctly classified as mis/no matches in 3D point generation. The remaining one was misclassified as damage related to structural hole as it was affected by shadow in the post-event data.

Table 5-2. Results of classification of missing 3D segments detected by segment-based approach

	Predicted	
Actual	**Damage**	**Non-damage**
Damage	20	0
Non-damage	1	2
Recall =100% ; precision= 95.23% and accuracy= 95.6%		

The final outcome of the overall process is the report for each building that shows the area and boundary map of each missing elements of the building due to damage. For example, Figure 5-11 depicts the area and outline of each missing elements for a building shown in Figure 5-3.

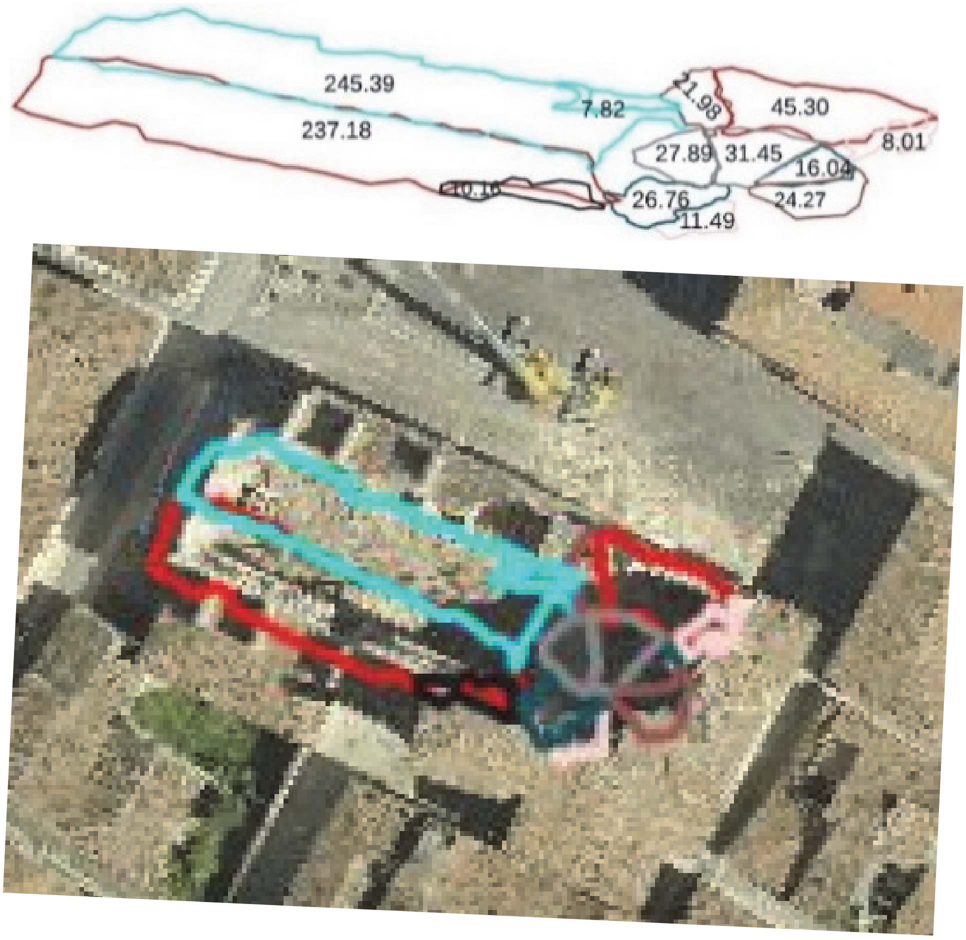


Figure 5-11. Outline of each missing segment of a building and their areas are annotated in m^2 (top) and superimposed on the corresponding building in post-event (bottom).

5.5 Discussion

The pre- and post-event point clouds derived from the oblique images are noisy. However, the noise level varies from place to place as they are susceptible to the radiometric characteristics of the surface. Therefore, the places corresponding to very noisy 3D points were not segmented well, and because of this three (out of 23) damaged elements were not detected by both segment-based approaches (cf. Table 5-1). The voxel-based approach has not detected six damaged elements. This is due to its limitations that are highlighted in the cons of the voxel-based approach (cf. 0). The change classification (cf. 5.3.1.3) is a straightforward approach which correctly classified the missing elements due to earthquake damages or man-made changes. The classification of mis/no matches failed just in one case, in correspondence of a shadowed building element that was wrongly classified as structural hole.

5.6 Conclusion and outlook

In the presented chapter, automated methods were developed to identify the structural damages by utilizing both 2D and 3D information derived from multi-temporal, pre- and post-event oblique images. The methods focussed on identifying structurally deformed elements due to damage using the pre- and post-event images and 3D point cloud. The developed methods produced significant results, particularly, the segment-based approaches detected 87% (20 out of 23) of geometrically deformed elements, and all of them were correctly classified as damaged by the proposed change classification approach. However, while the proposed methods can identify the structurally deformed elements due to damage, they cannot infer the type of structural deformation, such as dislocation, inclination, complete collapse or pancake collapse. Generally, any missing structural element in the pre-event data will emerge as a new element (at least debris) in the post-event epoch. These newly emerged post-event elements can be identified by detecting the missing elements from the post- to pre-event data using any one of the proposed change detection methods. A semantic analysis by mapping of corresponding missing pre-event and the newly emerged post-event elements would help to infer the specific reason of structural deformation. This would be the logical extension of this work.

The planar segmentation was adopted to derive segments in the segment-based change detection approaches. It often failed to provide accurate segmentation for very noisy and non-planar regions in the 3D point cloud, which hindered the assessment for those regions. However, numerous point-cloud segmentation methods have been developed which utilize the contextual information, the image-radiometric and points cloud-geometric information in combination, and which provide better segmentation for noisy and non-planar regions (cf. Anh and Bac (2013)). Adopting such segmentation coupled with our proposed composite segment-based approach can yield better assessment.

5.7 References of Chapter 5

Anh, N., Bac, L., 2013. 3D point cloud segmentation: A survey. *6th IEEE Conference on Robotics, Automation and Mechatronics (RAM),*, 12-15 Nov. 2013, pp. 225-230.

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.

Gerke, M., Kerle, N., 2011. Automatic structural seismic damage assessment with airborne oblique pictometry imagery. *Photogrammetric Engineering & Remote Sensing,* 77 (9), pp. 885-898.

Katz, S., Tal, A., Basri, R., 2007. Direct visibility of point sets. *ACM Transactions on Graphics (TOG),* 26 (3), pp. 24.

Khoshelham, K., Oude Elberink, S., Sudan, X., 2013. Segment-based classification of damaged building roofs in aerial laser scanning data. *IEEE Geoscience and Remote Sensing Letters,* 10 (5), pp. 1258-1262.

Miura, H., Midorikawa, S., Kerle, N., 2013. Detection of building damage areas of the 2006 Central Java, Indonesia, earthquake through digital analysis of optical satellite images. *Earthquake Spectra,* 29 (2), pp. 453-473.

Murtiyoso, A., Remondino, F., Rupnik, E., Nex, F., Grussenmeyer, P., 2014. Oblique aerial photography tool for building inspection and damage assessment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XL-1, pp. 309-313.

Sui, H., Tu, J., Song, Z., Li, Q., 2014. A novel 3d building damage detection method using multiple overlapping UAV images. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-7, pp. 173-179.

Uprety, P., Yamazaki, F., 2012. Detection of building damage in the 2010 Haiti earthquake using high-resolution SAR intensity images. *Journal of Japan Association for Earthquake Engineering,* 12 (6), pp. 21-35.

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, pp. 61-78.

Vosselman, G., 2012. Automated planimetric quality control in high accuracy airborne laser scanning surveys. *ISPRS Journal of Photogrammetry and Remote Sensing,* 74, pp. 90-100.

Wenjing, J., Huaifeng, Z., Xiangjian, H., Qiang, W., 2006. A comparison on histogram based image matching methods, *IEEE International Conference on Video and Signal Based Surveillance -AVSS '06*, pp. 97-97.

6 Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning*

* This chapter is based on the article:
Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning. In: Proceedings of GEOBIA 2016 : Solutions and synergies, 14-16 September 2016, Enschede, Netherlands / edited by N. Kerle, M. Gerke and S. Lefevre. Enschede: University of Twente Faculty of Geo-Information and Earth Observation (ITC), 2016. ISBN: 978-90-365-4201-2. 5 p

Abstract

Automated damage assessment based on satellite imagery is crucial for initiating fast response actions. Several methods based on supervised learning approaches have been reported as effective for automated mapping of damages using remote sensing images. However, adopting these methods for practical use is still challenging, as they typically demand large amounts of training samples to build a supervised classifier, which are usually not readily available. With the advancement in technologies local and detailed damage assessment for individual buildings is being made available, for example through analysis of images captured by unmanned aerial vehicles, monitoring systems installed in buildings, and through crowdsourcing. Often such assessments are being done in parallel, with results becoming available progressively. In this chapter, an online classification strategy is adopted where a classifier is built incrementally using the streaming damage labels from various sources as training samples, i.e. without retraining it from the scratch when new samples stream in. The Passive-Aggressive online classifier is used for the classification process. Apart from the classifier, the choice of image features plays a crucial role in the performance of the classification. The features extracted using recently reported deep learning approaches such as Convolutional Neural Networks (CNN), which learns features directly from images, have been reported to be more effective than conventional handcrafted features such as gray level co-occurrence matrix and Gabor wavelets. Thus in this study, the potential of CNN features is explored for online classification of satellite image to detect structural damage, and is compared against handcrafted features. The feature extraction and classification process is carried out at an object level, where the objects are obtained by over-segmentation of the satellite image. The proposed online framework for damage classification achieves a maximum overall accuracy of about 73%, which is comparable to that of batch classifier accuracy (74%) obtained for the same training and testing samples, however at a significantly lesser time and memory requirements. Moreover, the CNN features always significantly outperform handcrafted features.

6.1 Introduction & related works

Very high resolution satellite images, which are usually made available within a few hours after a disaster event such as an earthquake, serve as an ideal data source for rapid damage assessment over large areas for fast response actions (Kerle and Hoffman, 2013a). Though many automated methods have been proposed for damage mapping using satellite imagery, in practice operational damage mapping continues to be based on manual interpretation of satellite images, which is time- and labour- intensive. There are many reasons for automated methods not yet being adopted for automatic processing, including the limited spatial resolution of satellite imagery compared to increasingly available aerial alternatives, and the complexity of the scene. However, the primary reason is that most of the automated methods are based on supervised learning approaches that require a large volume of training samples to build accurate models which are often not readily available (Dong and Shan, 2013). Also adopting a pre-trained supervised model that was previously developed for a different geographic area or image data type typically has limited transferability. In such cases at least a small number of training samples reflecting the study area is required to fine-tune the pre-trained model. Either for constructing a new model or for calibrating the existing model, a significant number of study-area-specific training samples is required. However, training samples representing the distribution of damage characteristics of the study area, are usually not readily available. Moreover, manual construction of a new training dataset with a large number of samples is not feasible at the time of emergency. With the advancement in technologies, local and detailed damage assessments for individual buildings are made available from various sources. For example, it is becoming increasingly common to assess damages locally after an earthquake. UAV images substantially exceed satellite images in terms of spatial resolution and multi-perspective coverage of individual buildings (Vetrivel et al., 2015a). Therefore, automated damage estimation based on UAV images can be more accurate and reliable. Furthermore, damage information for specific locations is made available on-line from many other sources such as buildings equipped with automated monitoring systems (Sdongos et al., 2014) and crowdsourcing (Ghosh et al., 2011; Sdongos et al., 2014). Local damage estimates from such sources could be used to construct the study-area-specific training data for building a required supervised classifier. However, two major challenges need to be addressed when constructing a supervised classifier with such training data:

1) Handling streaming training samples: In the considered scenario damage assessment results from the aforementioned sources arrive at different points in time, also depending on whether data processing takes place locally or remotely. Therefore, a classifier is needed that can be trained dynamically, i.e., when new samples arrive the classifier should learn without retraining from scratch, and reclassify the image if required. This kind of learning is referred to as incremental or on-line learning (i.e. learning without having access to all the samples at once) (Crammer et al., 2006). Conventional batch-learning methods such as Random Forests or Support Vector Machines (SVM) are not suitable for this kind of learning (Narayanan et al., 2016b). Many on-line learning algorithms have been developed, and have been shown to perform similarly to batch-learning methods (Hoi et al., 2014; Wang et al., 2014).

2) Data-specific feature for damage classification: Even when a large number of training samples is available, it is critical to choose the features and a representation strategy that is suitable for the specific data, study area and application. For example, in earlier work (cf. Vetrivel et al. (2016b)) we examined various image features for identifying damage using images from various geographic locations. It was observed that specific features are performing well for specific study areas. Hence, selection of appropriate features specific to the study area is crucial for improved assessment. Recent research revealed that supervised feature learning methods such as Convolutional Neural Networks (CNN) could learn the data-specific features and their representation directly from the image pixel values (Bengio et al., 2013). These features are found to be far superior to conventional handcrafted features, which are described in the later section (Bengio et al., 2013).

Generally, the regions corresponding to heavy damage are determined through the identification of damage patterns corresponding to rubbles piles, debris and spalling in an image (Kerle and Hoffman, 2013a). The recognition process of those damage patterns can be performed by analysing features extracted either at pixel or region level (Dong and Shan, 2013; Kaya et al., 2010; Miura et al., 2013). However, the pixel level analysis is not meaningful for very high spatial resolution images, particularly in the context of damage assessment, as the evidences are identified based on the characteristics of their radiometric distribution pattern, which can be captured more precisely at a region- or object-level. Therefore, super-pixels or segments derived from object-based image analysis approaches are considered as the primary entity for performing feature extraction and classification. Moreover, this kind of segment-based (super-pixels) approach has been demonstrated as a more efficient approach in several studies compared to pixel-based methods, particularly in applications dealing with very high resolution images (Blaschke, 2010; Blaschke et al., 2014).

Another challenging task is how to compile information from various sources to label each super-pixel as damaged or undamaged. However, this is not the focus of this chapter. Instead we center on how to build an incremental classifier if the labels for super-pixels are made available from streaming data sources. Hence, in this study the labels for super-pixels are manually annotated and they are synthetically framed as a stream of training labels obtained from various sources to carry out the online-classification.

To the best of our knowledge online classification with CNN features has not yet been tested for remote sensing applications, particularly for damage assessment. Thus the objective is to develop a framework to use online-learning and CNN together to build an incremental classifier with data-specific features for automated satellite image-based damage classification from streaming samples.

6.2 Methodology

The proposed methodology comprises three pipeline processes. As a first step, the satellite image is split into super-pixels using an over-segmentation approach. In the second step, feature extraction is carried for the super-pixels, and as a step 3 an online classifier is constructed based on the extracted features, by considering them as streaming training samples.

Step 1: Over-segmentation of image

Super pixel construction is a mandatory pre-processing step in many image processing applications. For that a range of methods has been reported (Achanta et al., 2012a; Aksoy and Akcay, 2005b; Salem et al., 2013b). Among them Simple Linear Iterative Clustering (SLIC) (cf. Achanta et al. (2012a)) is widely used and reported to be effective for obtaining objects in uniform size, which is suitable to create rectangular image patches, the input format required by common CNN implementation for extracting features.

Step 2: Feature extraction

Many kinds of feature extraction techniques, such as statistical, filtering and morphological operations, have been reported for image processing applications (Zhang and Tan, 2002). Among them filtering is recognized as the most effective approach (Arivazhagan et al., 2006; Tian, 2013b). For example, many popular features such as Gabor-, Sobel-, Gaussian- and wavelet-features are based on filtering techniques. Such filtering-based features have proved to be effective for many image processing applications, particularly for image classification (Arivazhagan et al., 2006; Tian, 2013b). These filters are designed based on standard mathematical functions and they are referred to as hand-crafted features. However, these features are not especially designed for specific data types or applications. It is challenging to choose appropriate features, i.e. designing filters with appropriate weights that give the best image representation for a specific application (Vetrivel et al., 2016b). CNN is one of the deep learning approaches where the filters' weights are learned directly from the images chosen for a specific application, instead of using a mathematical function (Krizhevsky et al., 2012). It has been reported that filters that are directly learned from images outperform conventional handcrafted features (Antipov et al., 2015; Chen et al., 2015; Krizhevsky et al., 2012) .

In this study, both hand-crafted- and CNN features are examined and compared for the damage classification capacity.

a) **Hand-crafted features:** Two kinds of hand-crafted texture features that have been widely reported as effective features for damage classification are considered. They are 1) features based on gray-level co-occurrence matrix (GLCM) and 2) Gabor wavelet features. Details about the extraction of GLCM- and Gabor wavelet- features can be found in Preethi and Sornagopal (2014) and Arivazhagan et al. (2006), respectively.

b) **Deep learning features based on CNN:** CNN can be used for classification in three different scenarios:
    1) **Training from scratch:** Designing and training of new CNN requires a large amount of training data to avoid overfitting. This method can be adopted if large number of training sample is available.

2) **Tuning a pre-trained model:** Another common approach is to adopt a pre-trained CNN model that is trained for a related domain (e.g., general image classification) using a large amount of training samples: the network weights of a pre-trained model are fine-tuned using a domain specific training samples. This approach also requires relatively many training samples.
3) **Pre-trained model without tuning**: If only a small amount of training sample is available for the designated application (here damage classification), then a common approach is to extract the features using the pre-trained model and use them to perform classification based on any supervised classifier such as SVM. The activations of one of the fully connected layers in CNN architecture are considered as the CNN features for a given image patch (super-pixels). For more details about the layers in CNN architecture refer to Zeiler and Fergus (2014) and Krizhevsky et al. (2012).

The third approach is most suitable for our application where it is usually common to obtain only small amount of training samples from the aforementioned sources i.e., few hundreds to thousands of samples depending on the nature of the disaster event. Moreover, this approach has been reported to be effective for image classification in various domains, including remote sensing applications such as land cover classification using very high resolution images (Hu et al., 2015). Hence, this approach is adopted in this study where the features for super-pixels are obtained using the pre-trained model and using these features an independent supervised classifier is built for performing the final classification.

Step 3: online classification
Many online-learning algorithms have been proposed and among them the widely used Passive-Aggressive (PA) algorithm is adopted here for building the online classifier (Crammer et al., 2006). The classifier is built incrementally by providing one sample at a time, where it predicts the label and confidence rate of the provided unseen sample. Further, the classifiers gets updated for each wrong and low confidence prediction. The details of the PA algorithm can be found in Crammer et al. (2006).

The overall work flow is depicted in Figure 5-1.

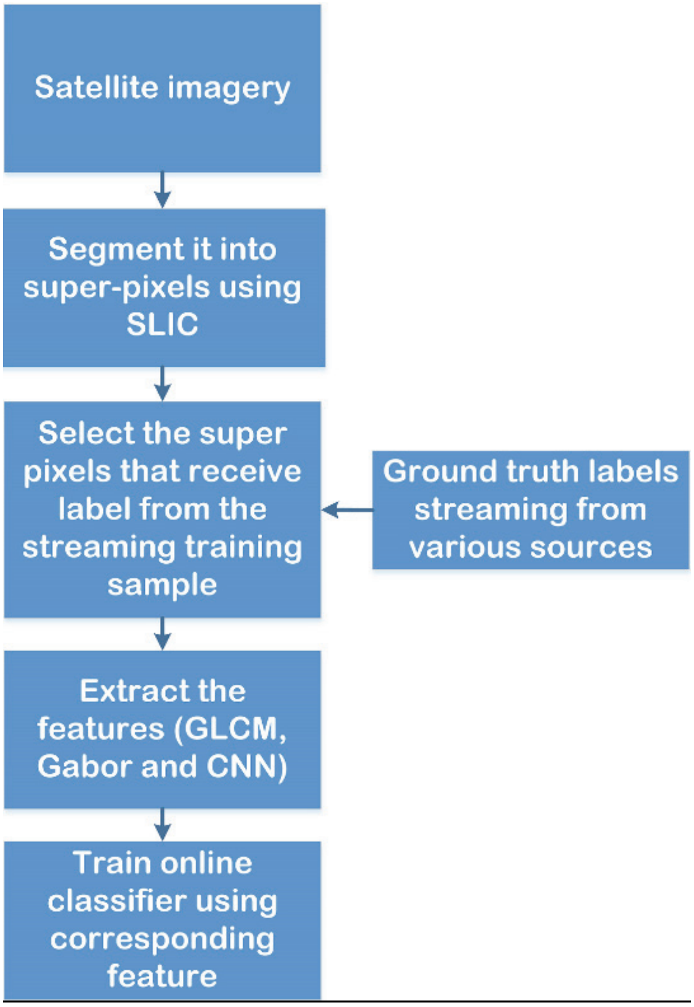


Figure 6-1. Overall workflow

6.3 Experiments

6.3.1 Data used:

A subset of a Geoeye satellite image with 50 cm nominal ground resolution of Port-au-Prince captured after the 2010 Haiti earthquake was considered for this experiment. The damaged and undamaged regions in the image were manually delineated using a polygon and annotated as damaged and undamaged, respectively, to generate the training samples for building the classifier. These polygons are considered as the streaming ground truth information for analyzing the proposed online classifier.

6.3.2 Experimental steps and implementation details:

Step 1: The super-pixels were generated for the selected image subset using SLIC method. As stated earlier, we need super-pixels with more uniform size and shape. To achieve this, the parameter 'm' in SLIC that controls the compactness of a super-pixel was empirically determined as 40.

Step 2: The manually delineated polygons annotated with damage labels were overlaid on the super-pixels of the image. The super-pixels having at least 50% of their area overlapping with a polygon were assigned with corresponding polygon's label, and considered for training and testing the classifier. In total, 2553 super-pixels were labelled, of which 70% of the samples were considered for building the classifier, while the remaining 30% samples were used for testing.

Step 3: Three kinds of features (GLCM, Gabor and CNN features) as described in the methodology section were extracted for the selected super-pixels. The CNN features were extracted using a pre-trained CNN model 'imagenet-caffe-alex' developed by Krizhevsky et al. (2012). This model demands the input of image patch with size 227x227x3. Hence the super-pixels were converted into rectangular patches and then scaled to above said dimension. In general the images are scaled to larger size using an interpolation technique that significantly degrades the quality of the image. This might have an impact on the quality of the CNN features. Hence, in this study the image patches based on super-pixels were resized to aforementioned dimensions using two approaches: 1) images resized based on interpolation techniques and 2) images resized based on zero padding. CNN features from image patches obtained based on these two approaches are compared as well.

In total four different features – GLCM, Gabor, CNN_pad (image resized by zero padding) and CNN_nopad (image resized by interpolation technique) are analyzed independently by fitting the PA online classifier.

6.4 Results

The results of the online classification are shown in Figure 6-2, which depicts the cumulative error rate for each feature. Also this figure implicitly shows the number of times the classifier gets updated. The overall accuracy estimated based on their error rates for GLCM, Gabor, CNN_pad and CNN_nopad are 66.5%, 70.0%, 71.0% and 74.5%, respectively.

It is evident that the online learning algorithm does not need to be re-trained from scratch every time a sample streams in, leading to significantly superior time and memory efficiency. However, this might adversely affect the accuracy of the model. Hence, we intend to determine whether the accuracy diminishes in the online learning setting, and if so, how significant would that be. To this end, we compare the online classifier with canonical batch classifier SVM, using all aforementioned features. The SVM classifier was trained using the same 70% of the samples. Subsequently, both SVM and PA classifiers were evaluated using the remaining 30% of samples and the accuracies are reported in Table 6-1.

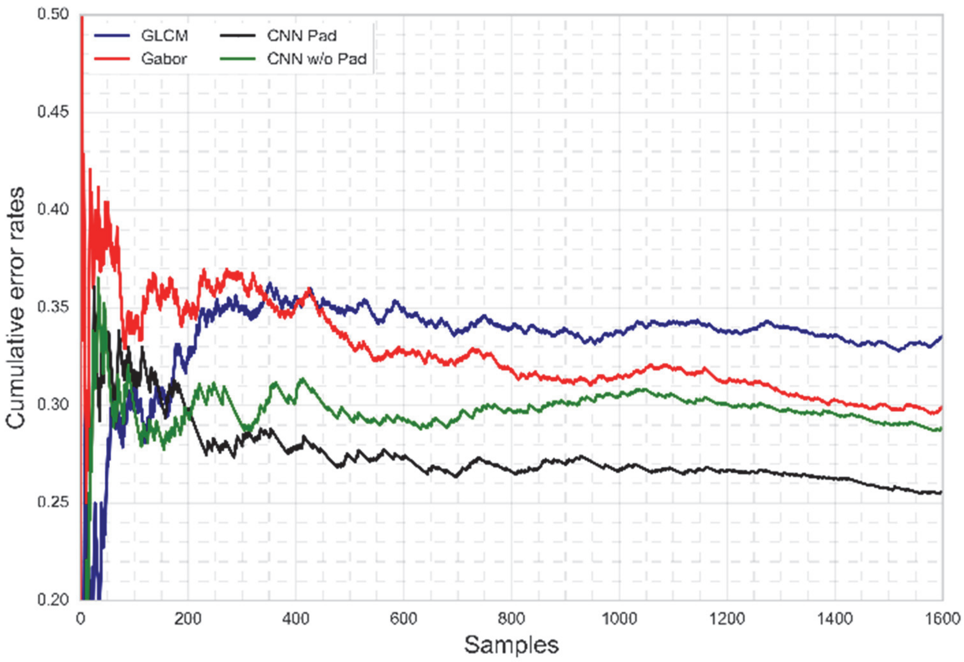


Figure 6-2. The cumulative error plot of PA online classifiers when associated with GLCM, Gabor and CNN features

Table 6-1 Overall accuracy of online- and batch- classifiers when associated with different features

Classifier	Overall accuracy in %			
	GLCM	Gabor	CNN_nopad	CNN_pad
PA	57.3	59.5	68.1	73.2
SVM	63.4	63.7	70.9	74.7

6.5 Discussions, conclusions and future work

A framework was developed for automated mapping of building damage from satellite imagery, using an online classifier that dynamically learns from streaming training samples based on local damage assessments from different sources. In this study, the potential of CNN features from pre-trained model, in combination with online PA classifier was explored for automated damage classification of satellite imagery. Further, the accuracy retrieved from CNN features was compared against the standard hand-crafted features GLCM and Gabor by fitting them using the PA classifier. The results show that CNN features are performing better than both handcrafted features. Moreover, the cumulative error graph (cf. Figure 6-2) shows that CNN features generalize better with fewer training samples than the handcrafted features. For example, the error curve (cf. Figure 6-2Figure 6-2) of CNN tends to be smoother (shows better generalization) after being trained with 1000 samples, while the error curves of handcrafted features fluctuates, indicating that it could not learn fast and generalize well compared to CNN features. Also the error graph shows that the choice of image resizing approach used to resize the super-pixel based image patch in a dimension required for CNN feature extraction have a significant impact on the performance of the final classification accuracy. For example, CNN features extracted from the image patch resized using an interpolation technique were found to be significantly inferior to the CNN features extracted from image patch resized using zero padding (cf. Figure 6-2).

The features evaluated in online learning settings were also evaluated using the canonical batch mode classifier SVM. The results show that the selected online classifier performs slightly inferior to the batch classifier, irrespective of the features (cf. Table 6-1). However, the difference in accuracy is only around 1% concerning CNN features. There is significant difference between the accuracies of online and batch classifiers when employing the handcrafted features. This is because the online classifiers based on the handcrafted features could not learn and generalize well. The overall conclusion is that the choice of features has significant impact on the performance of the online classifier. Also it is observed that compared to handcrafted features the CNN features are more effective and consistent for online classification, as they yield similar accuracy when used in the batch learning setting (cf. Table 6-1).

Though CNN features were reported to be effective for many remote sensing application such as land cover classification, the maximum classification accuracy obtained for damage classification using satellite imagery was around 73% (Table 6-1), signalling a continuing limitation of automated damage classification with satellite imagery (Kerle, 2010). This inferior accuracy might be due to many reasons such as insufficient number of training samples used for building the classifier, and complexity of the scene. However, the major reason would be the limitation of the satellite imagery where the considered image resolution is around 50 cm. With this level of spatial resolution even manual interpretation is difficult, and similar problems have previously been reported (Kerle, 2010). The available pre-event satellite imagery could be of help and is usually available. In such a case, the classifier can be trained with CNN features from both epochs, which could provide better results and we intend to explore this in the future. On the other hand, the training samples considered in this study are from single satellite imagery of a single geographic location, where the characteristics of the streaming samples are not highly variable. However, the usefulness and potential of online learning can be realized when there is high variability in the streaming samples in terms of characteristics of the study area, weather condition and spatial resolution of the images. Hence, as a future work we intend to explore the potential of the proposed framework by applying it to much better spatial resolution satellite or aerial imagery obtained for different geographic locations varying in aforementioned characteristics.

6.6 References of Chapter 6

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

Aksoy, S., Akcay, H.G., 2005. Multi-resolution segmentation and shape analysis for remote sensing image classification, *Proceedings of 2nd International Conference on Recent Advances in Space Technologies, RAST 2005.* , pp. 599-604.

Antipov, G., Berrani, S.-A., Ruchaud, N., Dugelay, J.-L., 2015. Learned vs. hand-crafted features for pedestrian gender recognition. *Proceedings of the 23rd ACM international conference on Multimedia*, Brisbane, Australia, pp. 1263-1266.
Arivazhagan, S., Ganesan, L., Priyal, S.P., 2006. Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters,* 27 (16), pp. 1976-1982.
Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35 (8), pp. 1798-1828.
Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65 (1), pp. 2-16.
Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic object-based image analysis – towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing,* 87, pp. 180-191.
Chen, M., Zhang, L., Allebach, J.P., 2015. Learning deep features for image emotion classification, *IEEE International Conference on Image Processing (ICIP)* pp. 4491-4495.
Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., Singer, Y., 2006. Online Passive-Aggressive algorithms. *The Journal of Machine Learning Research,* 7, pp. 551-585.
Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.
Ghosh, S., Huyck, C.K., Greene, M., Gill, S.P., Bevington, J., Svekla, W., DesRoches, R., Eguchi, R.T., 2011. Crowdsourcing for rapid damage assessment: The global earth observation catastrophe assessment network (GEO-CAN). *Earthquake Spectra,* 27 (S1), pp. S179-S198.
Hoi, S.C., Wang, J., Zhao, P., 2014. Libol: A library for online learning algorithms. *The Journal of Machine Learning Research,* 15 (1), pp. 495-499.
Hu, F., Xia, G.-S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing,* 7 (11), pp. 14680-14707.
Kaya, G.T., Ersoy, O.K., Kamasak, M.E., 2010. Spectral and spatial classification of earthquake images by support vector selection and adaptation. *International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 194-197.
Kerle, N., 2010. Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it? *International Journal of Applied Earth Observation and Geoinformation,* 12 (6), pp. 466-476.
Kerle, N., Hoffman, R.R., 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural hazards and earth system sciences (NHESS),* 13 (1), pp. 97-113.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, pp. 1097-1105.
Miura, H., Midorikawa, S., Kerle, N., 2013. Detection of building damage areas of the 2006 Central Java, Indonesia, earthquake through digital analysis of optical satellite images. *Earthquake Spectra,* 29 (2), pp. 453-473.
Narayanan, A., Liu, Y., Chen, L., Liu, J., 2016. Adaptive and scalable android malware detection through online learning. *arXiv preprint arXiv:1606.07150.*, pp.
Preethi, G., Sornagopal, V., 2014. MRI image classification using GLCM texture features, *International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, pp. 1-6.
Salem, M., Ibrahim, A.F., Ali, H.A., 2013. Automatic quick-shift method for color image segmentation, *8th International Conference on Computer Engineering & Systems (ICCES)*, pp. 245-251.
Sdongos, E., Tsertou, A., Georgakopoulos, N., Loupos, K., Amditis, A., Joram, N., Ellinger, F., Lindner, B., Edwan, E., Ehlerding, A., Camarinopoulos, S., Kallidromitis, V., Bairaktaris, D., Bairaktaris, E., Naundrap, J., Kerle, N., Gerke, M., Vierhuß-Schloms, K., Markus, M., 2014. A novel & practical approach to structural health monitoring -The RECONASS vision: Local positioning, sensor networks, secure communications and remote sensing at the service of structural monitoring to assess construction damage and related needs, *IEEE Workshop on Environmental Energy and Structural Monitoring Systems (EESMS)* pp. 1-6.
Tian, D.P., 2013. A review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering,* 8 (4), pp. 385-396.
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, pp. 61-78.
Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2016. Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach. *Remote Sensing,* 8 (3), pp. 231.
Wang, J., Zhao, P., Hoi, S.C.H., 2014. Cost-sensitive online classification. *IEEE Transactions on Knowledge and Data Engineering,* 26 (10), pp. 2425-2438.
Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, pp. 818-833.
Zhang, J., Tan, T., 2002. Brief review of invariant texture analysis methods. *Pattern Recognition,* 35 (3), pp. 735-747.

7 Segmentation of UAV-based images incorporating 3D point cloud information[*]

[*] This chapter is based on the article:
Vetrivel, A., Gerke, M., Kerle, N. and Vosselman, G. (2015) Segmentation of UAV - based images incorporating 3D point cloud information. In: PIA15+HRIGI15 : Joint ISPRS conference proceedings, 25-27 March 2015, Munich, Germany. ISPRS Archives, Vol XL-3/W2. : open access. ISPRS, 2015. pp. 261-268.

Abstract

Numerous applications related to urban scene analysis demand automatic recognition of buildings and distinct sub-elements. For example, if LiDAR data is available, only 3D information could be leveraged for the segmentation. However, this poses several risks, for instance, the in-plane objects cannot be distinguished from their surroundings. On the other hand, if only image based segmentation is performed, the geometric features (e.g., normal orientation, planarity) are not readily available. This renders the task of detecting the distinct sub-elements of the building with similar radiometric characteristic infeasible. In this chapter, the individual sub-elements of buildings are recognized through sub-segmentation of the building using geometric and radiometric characteristics jointly. 3D points generated from Unmanned Aerial Vehicle (UAV) images are used for inferring the geometric characteristics of roofs and facades of the building. However, the image-based 3D points are noisy, error prone and often contain gaps. Hence the segmentation in 3D space is not appropriate. Therefore, we propose to perform segmentation in image space using geometric features from the 3D point cloud along with the radiometric features. The initial detection of buildings in 3D point cloud is followed by the segmentation in image space using the region growing approach by utilizing various radiometric and 3D point cloud features. The developed method was tested using two data sets obtained with UAV images with a ground resolution of around 1-2 cm. The developed method accurately segmented most of the building elements when compared to the plane-based segmentation using 3D point cloud alone.

7.1 Introduction and related work

Automatic detection of individual building and recognition of its distinct sub-elements from remote sensing data are crucial for many applications including 3D building modelling, building level damage assessment and other urban related studies (Dong and Shan, 2013; Sun and Salvaggio, 2013). Generally, the buildings and its elements possess unique geometric characteristics. Hence, the 3D geometric features are being used as the fundamental information in building detection and categorisation of its sub-elements (Rottensteiner et al., 2014; Xiong et al., 2013). 3D point clouds are well suited to infer the geometric characteristics of the objects. Particularly, the multi-view airborne oblique images are a suitable source to generate 3D points cloud for building analysis as they can provide information of both the roofs and facades of the building (Liu and Guo, 2014). Unmanned Aerial Vehicles (UAVs) are attractive platforms which can capture the images with suitable characteristics such as multi-view, high overlap and very high resolution to generate very dense 3D point cloud in minimal time and cost (Colomina and Molina, 2014).

Generally, the building detection process from 3D point clouds has been carried out through identifying planar segments as most elements of general buildings are planar surfaces (Dorninger and Pfeifer, 2008). Planar segments with its geometric features could help to detect and delineate buildings in the scene. However, an accurate segmentation of individual elements of the building is not always feasible, especially with the geometric features from image-based 3D point cloud. This is due to various reasons such as 1) presence of low-textured planar surfaces might lead to sparse 3D point cloud generation with significant gaps. In such case, a single planar segment might get fragmented into multiple small segments, or even partly missed, leading to an inaccurate segmentation; 2) Outliers or random errors which are inherent in image based 3D point clouds, especially when the image block configuration is not optimal might also leads to artefacts or inaccurate segmentation of building elements (Rupnik et al., 2014a); 3) Regions affected by poor visibility (e.g., only visible in single images due to occlusions) will have no 3D points and those areas cannot be segmented; 4) 3D points belonging to non-planar objects will not be segmented by plane-based methods and it is difficult to recognize the complex objects even using other methods such as model-driven approach from sparse and erroneous 3D point clouds (Xiong et al., 2014c); 5) Objects that share the same plane geometry, e.g., windows in the roof and façade plane, might not get segmented as individual entity, hence leading to under segmentation.

The segmentation based on radiometric features alone might delineate the building regions that possess similar spectral or textural characteristics. However, elements of different category with similar spectral characteristics cannot be differentiated, e.g., roof and façade of the building with same surface characteristics and colour might be segmented as a single element. Also the segments found based on spectral features cannot be categorised into roofs, facades, etc., without inferring its geometric characteristics. Hence, it is obvious that both geometric and spectral features are important for an accurate segmentation and recognition of distinct elements of the building.

Many studies used radiometric features such as colour along with geometric features and shape descriptors for recognition of objects in 3D point clouds through segmentation (Aijazi et al., 2013; Strom et al., 2010). However, the image-based 3D point cloud might be erroneous and incomplete with missing 3D points for some regions. Hence, performing the segmentation in image space by utilizing the geometric information from 3D point cloud could be an alternative strategy.

Previously many studies have been reported for image segmentation by using the combination of 2D radiometric and 3D geometric features e.g., segmentation of depth images (RGB-D) (Mirante et al., 2011; Yin and Kong, 2013). The surface normal, gradient of depth, and residuals of plane fitting are the widely used geometric features in depth image segmentation (Enjarini and Graser, 2012; Hulik et al., 2012). Spectral and spatial features such as colour, texture, edges and shape are widely used image-based features for segmentation (Tian, 2013a). Among them texture features from GLCM are often reported as key features to infer the radiometric characteristics of the surface (Rampun et al., 2013). Numerous segmentation approaches are used in practice such as region-based approach (e.g., region growing, split and merge), clustering-based approach (e.g., k-means, mean shift), and graph-based approach (e.g., graph-cut) (Boykov and Funka-Lea, 2006; Narkhede, 2013). However, the choice of segmentation approach depends on the application and kind of features available for segmentation. Region-based approaches are often preferred for segmentation based on multiple image and 3D-features as it implicitly utilizes the spatial connectivity as a constraint (unlike clustering methods). In contrast to graph-based approaches region growing is computationally cheap and multiple features can be combined straightforward.

Another aspect concerns the question whether a more data- or a more model-driven approach should be pursued. The key question is to which extent assumptions about the object structure and properties can be made. While model-driven methods help to mitigate the effect of insufficient observations by applying strong assumptions (knowledge) about the object, they might generalize quite strongly. If such knowledge is not available, a data-driven method should be used, being aware of the fact that uncertainties and errors in the observed information might lead to wrong results.

The objective of this research work is to develop a methodology to identify the distinct segments of buildings by 1) detecting the buildings from the 3D point cloud from the UAV-based oblique images and 2) performing a sub-segmentation within the building area in image space using both the spectral and corresponding geometric features from 3D point cloud. For both steps we aim to use as less assumptions (model knowledge) as possible, hence we are pursuing a strong data driven approach. The motivation for this is that one main application of our method is building damage assessment and for this task only vague assumptions should be made to avoid any kind of misinterpretation.

It is also important to note that so far we do not exploit multi-image observations for the segmentation except for the 3D point cloud information which is derived from image matching. Here again the damage mapping context justifies this decision: in many circumstances some parts of a damaged building are only well visible in single images. In this case still we want to be able to derive segmentation information.

7.2 Methods

The methodology for image segmentation includes two processes, 1) building delineation from a 3D point cloud to define the region of interest for performing image segmentation and 2) image segmentation using the spectral information from the image and 3D geometric features from the 3D point cloud.

7.2.1 Building delineation from 3D point cloud

The building delineation is carried out by finding the connected 3D planar roof segments from the 3D point cloud. A straightforward, quite simplistic approach is used, which, however, turned out to be quite successful, see result section. We only briefly describe this method here, because actually it is just a pre-processing step which allows restricting the processing area for the main step – the segmentation.

- The 3D points are segmented into disjoint planar segments using the plane-based segmentation method as described in Vosselman (2012).
- The 3D points are classified into terrain and off-terrain points using the method proposed by Axelsson (2000a) which is implemented as part of the software *lastools (http://lastools.org).* The height normalized 3D points are computed by differencing the height of each off-terrain 3D points to its closest terrain 3D point.
- The planar segments that are above certain height (T_H) and have surface normal z-component (nz) greater than threshold (Tz) are classified as roof segments.
- A connected component analysis is used to identify the spatially connected roof segments of a single building.
- A convex hull is used to define a 2D boundary of the connected roof segments that gives an approximate 2D boundary of the building.
- All 3D points that lie within the defined boundary are registered as the 3D points of the building.

7.2.2 Segmentation

The image segmentation process is carried out based on feature similarity between the spatially connected pixels.

It is a scenario where the 3D planar segments which are derived for detecting buildings from the 3D point cloud are available in addition to the image for segmentation. In this study, an image segmentation algorithm based on region growing concept is developed by utilizing both image spectral and 3D geometric features from the planar segments for finding the distinct segments in the building.

The success of region growing based image segmentation highly depends on three key elements,

a) ***Selection of seed points**:* The mid points of 3D planar segments (which are already identified as segments in 3D space) are taken as the seed points for region growing in image space. Here, we assume that at least a small region of all elements of the building will have 3D points.

b) Features used for pixel similarity analysis:

- ***Spectral features:*** In this study colour features and gray level co-occurrence matrix (GLCM) based texture features are considered to measure the pixel similarity for region growing. A small experiment is conducted to identify the radiometric features that show maximum variation between the pixels belonging to surfaces with different radiometric characteristics. The identified feature is then used in the region growing process.
- ***Geometric features:*** The 3D points are projected onto the image and the geometric properties such as normal vector and XYZ coordinate of each projected 3D points are assigned to the corresponding image pixel.

c) ***Criteria for region growing:*** Each image pixel will have a feature vector that represents the spectral characteristics of the pixel and may have geometric features in addition.

Three criteria are used for region growing:

1. The distance between the feature vector of a new pixel to the mean feature vector of the region being grown (Spectral distance) $<T_{SD}$.
2. The dot product of normal vector of a new pixel and the plane-normal of the region being grown (Normal difference) $<T_{angle}$.
3. The distance between the 3D point corresponding to the new pixel to the plane of the region being grown (point to plane distance) $<T_{distance}$.

The image pixels that do not have 3D features will be considered for region growing based on first criteria alone.

The global definition of spectral distance threshold T_{SD} is not appropriate for segmenting elements of the building corresponding to varying surface characteristics. For example, the pixels corresponding to a rough surface show high spectral variation between them, hence a high T_{SD} is required to avoid over-segmentation whereas a low T_{SD} is suitable for smooth surfaces to avoid under-segmentation. Therefore, instead of a global threshold, all seed points are assigned with an adaptive local threshold for region growing. The local threshold for each seed point is computed as the maximum spectral difference between the pixels corresponding to the 3D points that lie within a certain distance from the seed point in the 3D planar segment. Always, the region growing process is initiated by choosing the seed point corresponding to lowest local threshold value in the lists, in order to segment the smoother regions first to avoid under-segmentation.

Procedure for segmentation of individual elements in the building:

a) Data preparation for image segmentation:

1. Individual buildings in the scene are delineated from the 3D point cloud using the procedure described earlier.
2. Select one delineated building and an appropriate image where the building is visible for segmentation. We are not posing any requirements for image selection, since this decision should be made by the actual application, e.g. the image where a certain damage region is best visible.
3. The 3D points of the planar-segments of the delineated building which are visible in the selected image (camera view) are found using the hidden point removal (HPR) operator e.g., Katz et al. (2007b) as described in Gerke and Xiao (2014). The visible points are then projected over the image.
4. The image pixels that correspond to the projected 3D points are assigned with their plane-normal vector and XYZ value.
5. A majority filter is used to assign the 3D features for pixels that do not have corresponding 3D points from their adjacent pixels that have 3D points.
6. The boundary of the building in image is defined by constructing a convexhull for the projected 3D points which forms the region of interest (ROI) for segmentation.
7. The spectral feature such as colour and texture are derived for each pixel.
8. The midpoints of all 3D planar segments are considered as the seed points and each seed point is assigned with four parameters: a) normal vector of the plane, b) distance of the plane to the origin and c) local spectral distance threshold (T_{SD}), and d) feature vector of the seed point as mean spectral feature vector.

b) Image segmentation:

1. The seed points are sorted by local spectral distance threshold.
2. Remove the topmost seed point (i.e. the one with lowest T_{SD}) in the list and initiate region growing using this seed point.
3. Consider the un-segmented neighbouring pixels to the pixels in the region as new pixels for growing.
4. Grow the region by adding the new pixels to the region if they satisfy the growing criteria (refer to (c) under section 2.2) and they lie within the ROI.
5. Update the mean spectral feature vector of the region based on the newly added pixels.

6. Continue steps 3 to 5 until no new pixel is added to the region.
7. Compute the boundary of the new region using a boundary tracing algorithm.
8. Find the seed points that lie within the boundary of the obtained region and remove them from the list.
9. Continue steps 2 to 8 until the seed point list becomes empty.
10. Find the boundary of the regions with significant size (number of pixels) that remain un-segmented.
11. Consider the midpoint of the un-segmented regions as seeds for region growing and perform steps 2 to 9.

The overall workflow is depicted in Figure 7-1.

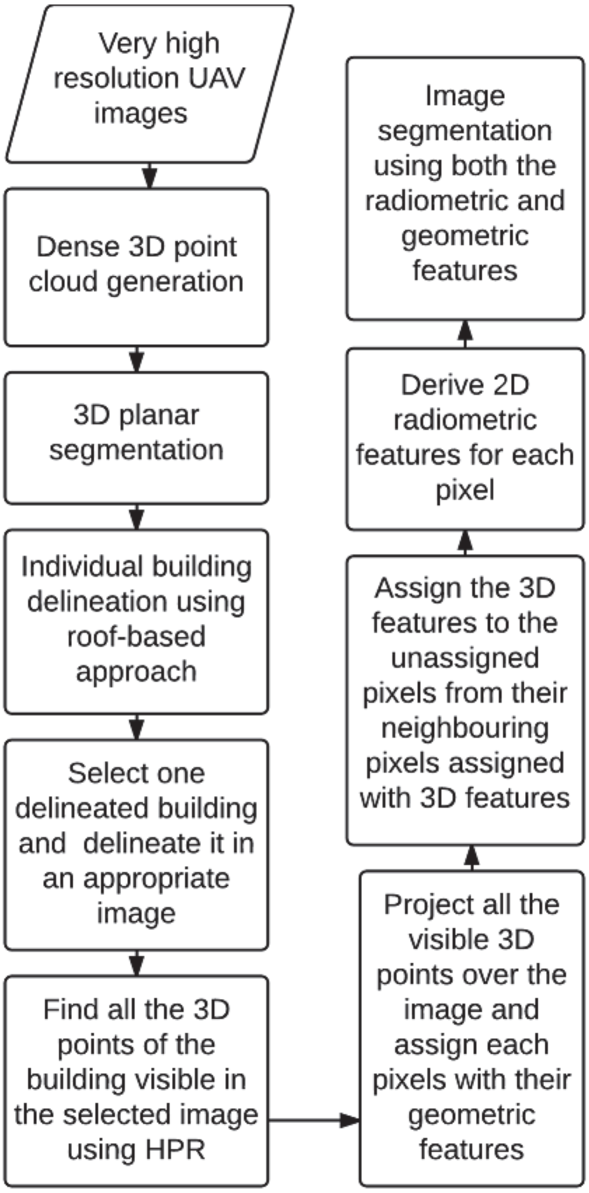


Figure 7-1 . Overall workflow

7.3 Experimental results

The proposed methodology was tested on two data sets captured by a UAV platform. One important aspect of this kind of image analysis task is the question on how far thresholds and parameters are transferrable. Therefore – besides the standard evaluation of the method – this issue is checked further. It is done by fixing threshold values for the first data set and using the same values for the second.

7.4 Data set 1 and results

The UAV images captured over a small region around the Church of Saint Paul in Mirabello, after the earthquake in 2012, were considered. The images were captured by a VTOL (vertical take-off and landing) UAV from various heights, positions and views (oblique and nadir). The average GSD of the captured images is around 1 cm. A dense 3D point cloud of the scene was generated from 152 images with an average point density of 650 points per m^2 by automatic orientation of the images, followed by dense matching using the software pix4Dmapper (http://pix4d.com). The selected region contained six buildings. Among them the larger one comprised of various complex sub-components was considered for testing the developed segmentation method. The selected building consists of different segments such as roofs composed of planar faces with different orientations and different radiometric characteristics, façades painted with different colour, windows in the façade, non-planar objects on the roof, balconies, etc.

7.4.1 Building delineation in 3D point cloud and in image of data set-1:

The 3D point cloud was segmented into disjoint planar segments. The thresholds T_H = 3 meters and T_Z = 0.6 were used to filter out the roof segments through the procedure described in section 2.1. All six buildings in the 3D point cloud of the scene were detected and delineated with close approximate to the actual boundary of the building. The major objective of this research is to segment the building into its various sub-components in image space, once it has been delineated in the 3D point cloud. Hence, detailed information about the conducted experiments, results and analysis related to building delineations from the 3D point cloud is not in the focus of this chapter. An example for building delineation from the 3D point cloud and the delineation of the same in the image is shown in Figure 7-2 and Figure 7-3. The planar segments that are obtained from the 3D point cloud and lie within the boundary of the delineated building were projected onto the image. Their geometric features were assigned corresponding image pixels. Figure 7-5 shows that segments are not accurately delineated from 3D planar segments. Also many portions of the building do not have projected 3D points, particularly the façade regions contain sparse 3D points hence these portions have radiometric features alone for segmentation.

7.4.2 Radiometric features and various threshold values used in segmentation

The colour features such as red, green, blue, hue, and saturation, and GLCM texture features such as mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation were considered. The potential of each feature in separating the pixels of different elements of the building was analysed. Five small image regions corresponding to various elements of the building with different radiometric characteristics were considered as shown in Figure 7-3. The above mentioned radiometric features were derived for each region. A silhouette value ((Wang et al., 2009) which gives the measure of how well each pixel in one cluster matches with the pixels in the other clusters was used to identify the features that show maximum variation (high silhouette value) between the pixels corresponding to different clusters. The GLCM features showed a higher silhouette value than the colour features (c.f. Figure 7-4). Particularly, the contrast and homogeneity features produced higher silhouette values than when used independently than used in combination with other GLCM features. Therefore, contrast and homogeneity of GLCM features were used as the radiometric features for image segmentation. The adaptive local spectral threshold method (c.f. section 2.2) provided better results than a global threshold. However, in few regions, an over-segmentation was observed which was then resolved by adding a constant to the local threshold value. As we have the radiometric features as additional constraint for segmentation, the geometric constraints were relaxed by setting higher threshold values for T_{angle} (0.9) and $T_{distance}$ (0.75 m) to achieve better results even with erroneous 3D point measurements.

The obtained segmentation result for the above mentioned threshold values is shown in Figure 7-6. Based on visual analysis, it was inferred that the segmentation obtained in image space based on both radiometric and geometric features is more accurate than the segmentation in 3D object space without using radiometric features. The developed segmentation algorithm delineated all planar surfaces in the building with close approximate to their actual boundary. The non-planar objects and regions that do not have 3D points were segmented using the radiometric features alone. However, in such cases, over- or under-segmentation was observed.

For example, c.f. Figure 7-6, where the rooftop element and small portion of ground were segmented as single segment because of radiometric similarity and absence of 3D information. This clearly implies that both geometric and radiometric features are essential for accurate segmentation. The same building was segmented in another image with smaller scale and different orientation (Figure 7-7 a). The segmentation was largely similar (Figure 7-7 b). However, the segmentation in larger scale image is more accurate. This slight performance difference may be due to the variation in texture representation between different scales.

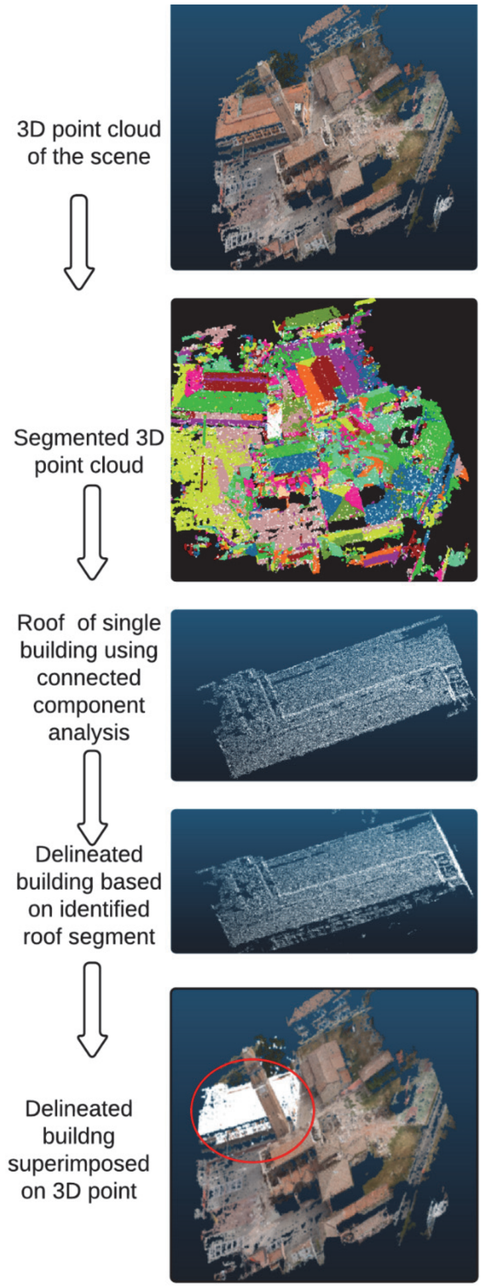


Figure 7-2 Building delineation from 3D point cloud

Figure 7-3 Building delineated in image and annotated regions (R1 –R5) are used for feature significance analysis as described in section 3.1.2

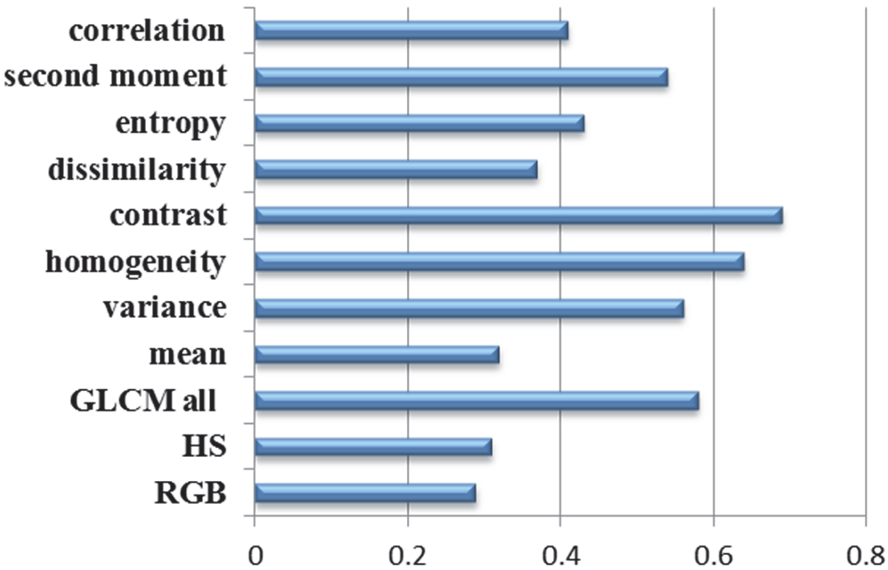


Figure 7-4 Silhouette value for analysing the feature significance in differentiating the image regions with different radiometric characteristics

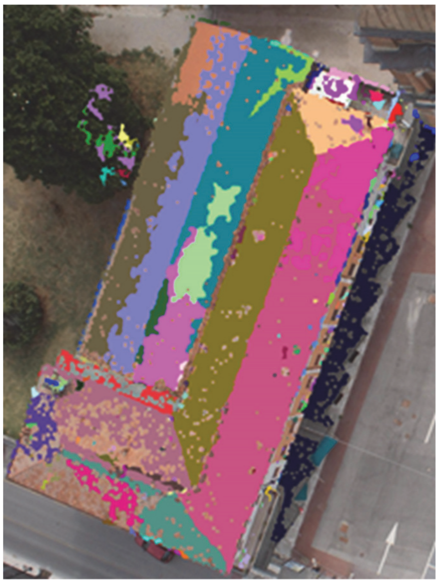


Figure 7-5 3D planar segments of the delineated building from 3D point cloud are projected over the image

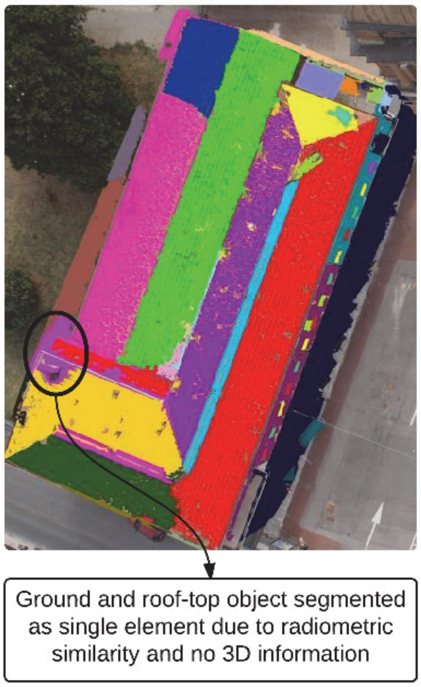
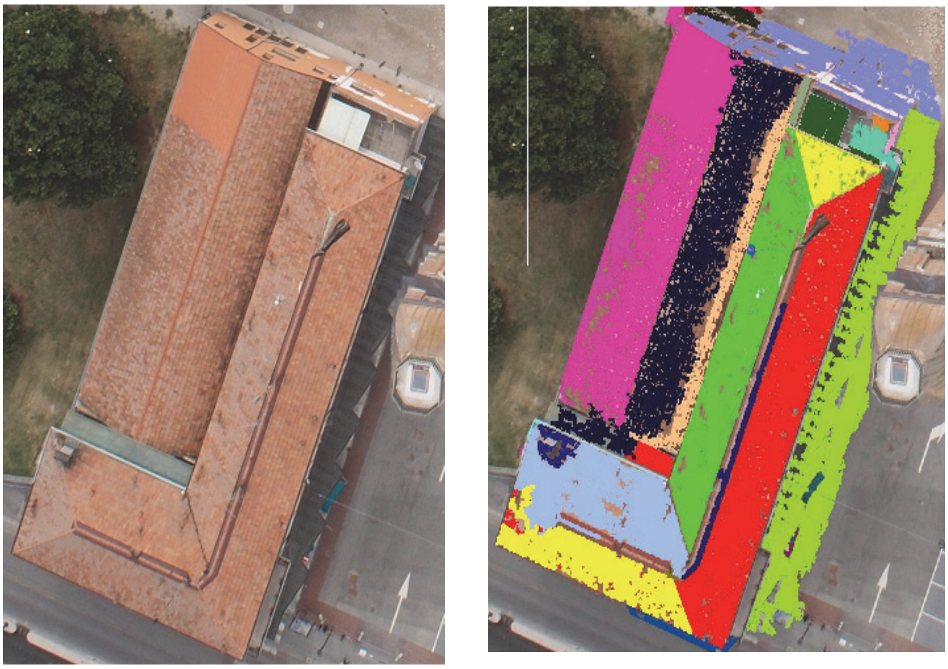


Figure 7-6 Delineated building segmented using radiometric and 3D geometric features



(a) (b)

Figure 7-7 (a) Building in image at smaller scale and different orientation compared to Figure 3, and (b) corresponding segmentation

The transferability of the thresholds to other datasets will be demonstrated with the following experiment.

7.4.3 Data set 2 and results

The developed segmentation algorithm was tested with the 3D point cloud generated from the UAV images of small urban area in the municipality of Nunspeet in The Netherlands (Hinsbergh et al., 2013). The images are captured in nadir view with an average GSD of 1.5 cm and the 3D point cloud was generated with an average point density of 250 points per m^2. The buildings in this region are less complex compared to the selected building from dataset 1. For example, the individual elements in the building are highly homogeneous and show high contrast with their neighbouring elements in terms of radiometric characteristics. Moreover, the buildings in the selected region are more identical and mainly made of planar surfaces. Among them two buildings that show different structure were considered for testing the segmentation algorithm (Figure 7-8 a & b). The selected buildings have gabled roofs with different kinds of windows on them, such as flat windows that lie in the same roof plane and windows extruded above the roof. The façades are single planar surfaces with uniform colour and texture.

The 3D planar segments obtained from 3D point cloud were projected over the image. Many of the 3D segments were more accurately segmented when compared to the 3D segments obtained for the building in dataset 1 (Figure 7-8 c & d). However, over-segmentation was observed in the façade and few places in the roof (Figure 7-8 d). The flat windows over the roofs were not identified as separate segments in the 3D segmentation.

The image segmentation using the texture features along with the projected 3D features was carried out following the same procedure and thresholds used for the segmentation of building in dataset1. The segmentation results are shown in Figure 7-8 e & f.

The obtained results were better than the plane-based 3D point cloud segmentation where the over segmented regions in 3D space such as façades were well segmented in the image space (c.f. Figure 7-8 d & f). Most of the windows and small non-planar above roof elements were also segmented as separate segments. However, in few places over-segmentation was observed due to the variation in radiometric characteristics within the same element. For example, the dirt in the corner of the segment resulted in over-segmentation even though they are geometrically recognized as single planar segment (c.f. annotated region in Figure 7-8 b and same region in Figure 7-8 d & f). This is due to the weakness in the segmentation criteria where the geometric constraints are relaxed to a certain extent when the radiometric characteristics are similar but not the other way around. However, the radiometric constraint has to be relaxed when there is strong hold on geometric characteristics. For example in the above case, the segmentation based on geometric features results in uniform shape whereas the consideration of radiometric features results in a non-uniform shape. In such instance the radiometric constraint can be relaxed. This kind of analysis can be carried out even in post-processing.

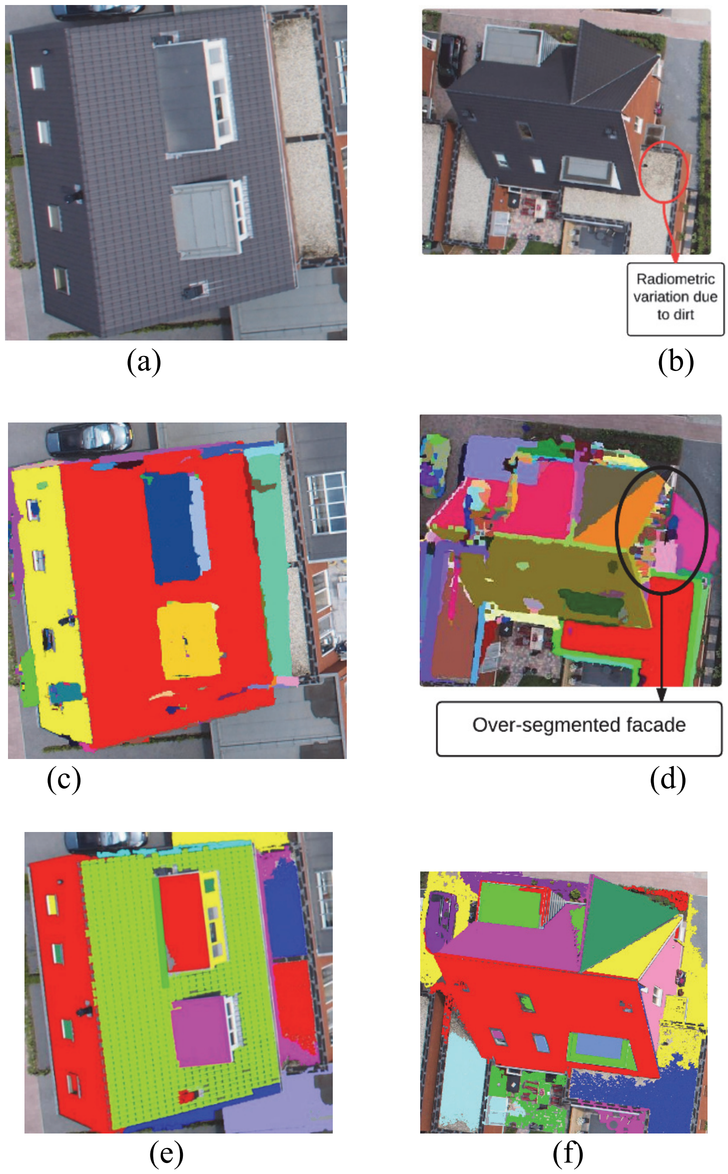


Figure 7-8. (a) & (d): Buildings in UAV image for segmentation, (c) & (d): projected 3D segments over the images of (a) & (b) respectively, and (e) & (f): finally segmented images using the developed method

7.5 Discussion and conclusion

The overall results indicate that the radiometric features complement the 3D geometric features and a combination of the two produced significantly superior segmentation compared to the 3D geometric features based segmentation alone. The radiometric features seem to be advantageous in identification of single segments even though there is significant error in geometric measurements. The sub-segmentation of planar objects also might lead to over-segmentation, when the face contains shadows, dirt, etc., refer to Figure 7-8 b, d and e. This is however, the correct behaviour since on purpose we chose this data driven approach. In the actual application – like damage mapping – those segments might give valuable information for the interpretation.

In this study, the 3D features such as normal orientation and planarity derived from plane-based segmentation in 3D space were used as geometric features in combination with radiometric features for segmentation in image space. The plane-based features alone are not sufficient in all cases. For example, plane-based features cannot accurately segment the curved surface which leads to over-segmentation. In such cases, other 3D features could be of help, such as the curvature feature which can be computed based on local neighbourhood 3D points. The inclusion of more 3D features such as curvature likely will improve the segmentation accuracy.

7.6 References of chapter 7

Aijazi, A.K., Checchin, P., Trassoudaine, L., 2013. Segmentation based classification of 3D urban point clouds: A super-voxel based approach with evaluation. *Remote Sensing,* 5 (4), pp. 1624-1650.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing,* 33 (B4/1, PART 4), pp. 111-118.

Boykov, Y., Funka-Lea, G., 2006. Graph cuts and efficient ND image segmentation. *International Journal of Computer Vision,* 70 (2), pp. 109-131.

Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 92, pp. 79-97.

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.

Dorninger, P., Pfeifer, N., 2008. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. *Sensors,* 8 (11), pp. 7323-7343.

Enjarini, B., Graser, A., 2012. Planar segmentation from depth images using gradient of depth feature, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012* IEEE, pp. 4668-4674.

Gerke, M., Xiao, J., 2014. Fusion of airborne laserscanning point clouds and images for supervised and unsupervised scene classification. *ISPRS Journal of Photogrammetry and Remote Sensing,* 87, pp. 78-92.

Hinsbergh, W.H.M.v., Rijsdijk, M.J., Witteveen, W., 2013. UAS for cadastral applications: testing suitability for boundary identification in urban areas. *GIM International,* 27, pp. 20-25.

Hulik, R., Beran, V., Spanel, M., Krsek, P., Smrz, P., 2012. Fast and accurate plane segmentation in depth maps for indoor scenes, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012* pp. 1665-1670.

Katz, S., Tal, A., Basri, R., 2007. Direct visibility of point sets. *ACM Trans. Graph.,* 26 (3), pp. 24.

Liu, J., Guo, B., 2014. Reconstruction and simplification of urban scene models based on oblique images. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 1, pp. 197-204.

Mirante, E., Georgiev, M., Gotchev, A., 2011. A fast image segmentation algorithm using color and depth map, *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, pp. 1-4.

Narkhede, H., 2013. Review of image segmentation techniques. *International Journal of Science and Modern Engineering,* 1 (8), pp. 54-61.

Rampun, A., Strange, H., Zwiggelaar, R., 2013. Texture segmentation using different orientations of GLCM features, *Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*. ACM, p. 17.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J.D., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing,* 93 (0), pp. 256-271.

Rupnik, E., Nex, F., Remondino, F., 2014. Oblique multi-camera systems - orientation and dense matching issues. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3/W1, pp. 107-114.

Strom, J., Richardson, A., Olson, E., 2010. Graph-based segmentation for colored 3D laser point clouds, *IEEE/RSJ International Conference onIntelligent Robots and Systems (IROS), 2010* pp. 2131-2136.

Sun, S., Salvaggio, C., 2013. Aerial 3D building detection and modeling from airborne LiDAR point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6 (3), pp. 1440-1449.

Tian, D., 2013. A Review on Image Feature Extraction and Representation Techniques. *International Journal of Multimedia and Ubiquitous Engineering,* 8 (4), pp. 385-396.

Vosselman, G., 2012. Automated planimetric quality control in high accuracy airborne laser scanning surveys. *ISPRS Journal of Photogrammetry and Remote Sensing,* 74, pp. 90-100.

Wang, K., Wang, B., Peng, L., 2009. CVAP: Validation for cluster analyses. *Data Science Journal,* 8, pp. 88-93.

Xiong, B., Oude Elberink, S., Vosselman, G., 2014. Building modeling from noisy photogrammetric point clouds. *ISPRS Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-3, pp. 197-204.

Xiong, X., Adan, A., Akinci, B., Huber, D., 2013. Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction,* 31, pp. 325-337.

Yin, J., Kong, S., 2013. Hierarchical image segmentation algorithm in depth image processing. *Journal of Multimedia,* 8 (5), pp. 512-518.

8 Accurate roof face delineation by synergistic use of UAV images and derived point clouds for automatic 3D reconstruction of buildings to perform detailed structural damage assessment

8.1 Introduction

Rapid assessment of the damage state of individual buildings after disaster events such as earthquakes provides critical information for stakeholders involved in response and recovery actions. As discussed in Chapter 1, the required level of damage information varies for each stakeholder. For example, those involved in the search and rescue process require only the information about collapsed/heavily damaged buildings, though very rapidly. Other actors, such as insurance companies, require a less time critical but very detailed damage assessment, precisely to the level of cracks on a building. In practice, remote sensing images are increasingly being used as an alternative to conventional field survey, and they constitute the best data source for initial and large-area damage assessment (Dell'Acqua and Gamba, 2012; Dong and Shan, 2013). Particularly, very high resolution (less than 15 cm ground sampling distance (GSD)) oblique view images have been found to be very useful, as they capture both the roofs and facades of the building with finer details than nadir-view satellite or aerial images (Gerke and Kerle, 2011a; Rasika et al., 2006; Saito et al., 2010). Numerous methods have been reported for building level damage assessment (Miura et al., 2013; Nex et al., 2014; Tong et al., 2012b), but most of them are focussed on the identification of completely collapsed or heavily damaged buildings (Dong and Shan, 2013). However, for practical use, a detailed building damage assessment (at least of physical damages) is required, of a type suitable for all stakeholders mentioned above (Fernandez Galarreta et al., 2015b). The prerequisites to perform such detailed damage assessment, for both manual and automated approaches, include: 1) identification of various kinds of damage evidences, such as debris, spalling, cracks, holes, broken elements, hanging and inclined parts of the building; 2) building face topology (e.g. a geometric model), as it is important to infer the type and position of the elements along with the damage evidences, since the damages have variable impact on different elements of the building. For example, damages to elements such as a roofs have less impact than damages to structural elements such as columns or beams in the facades; 3) semantic integration of those damages along different elements on different sides of the building to infer the overall damage state of the building, i.e. to arrive at one overall building damage score. Images alone can only partially satisfy the aforementioned prerequisites. This is because the building topology and the damage evidences that require 3D geometric information, such as holes, debris quantity, or hanging and inclined elements, cannot be inferred directly only from images. Hence a damage assessment only based on them will be incomplete, ambiguous and uncertain. However, a semantically rich 3D model of the buildings, with the corresponding images mapped to each portion of the 3D model, would be the best representation for performing such very detailed damage assessment. This has been partly demonstrated by Fernandez Galarreta et al. (2015a) and Saedi (2016). However, those demonstrations are based on synthetically generated 3D building models. Though the 3D models along with images are recognized as desirable data representation for detailed and more accurate damage assessment, it has not yet become operational. This is because of the challenges that subsist in the automated construction of such 3D models of buildings from remote sensing images for damaged environment. Addressing these challenges is the focus of this Chapter.

3D point clouds are an ideal data source for automatic 3D modelling of buildings (Rottensteiner et al., 2014; Sun and Salvaggio, 2013; Xiong et al., 2014d). Particularly photogrammetric 3D point clouds from images, such as acquired by Unmanned Aerial Vehicles (UAV), seem to be advantageous for generating detailed 3D building models for the reasons mentioned in Chapter 1. The images captured with the characteristics of the UAV will help to obtain high quality point clouds in terms of point density and good coverage of the individual buildings, better than point clouds from any manned airborne platform, which forms an ideal base for the detailed 3D modelling of buildings. However, the generation of such 3D models of the buildings in structurally damaged environments and particularly based on photogrammetric point clouds is quite challenging. This is because accurate delineation of independent roof faces of single buildings is the elementary requirement for 3D reconstruction, but often it is difficult to obtain them from photogrammetric point clouds (Vetrivel et al., 2015a). The difficulties are mainly because of two reasons:

1) Limitations in the photogrammetric point cloud: i) the presence of gaps in the 3D point cloud caused by poorly-textured or reflective surfaces, as well as by partial occlusion; ii) artefacts due to outliers and random errors, which are inherent in the process of photogrammetric 3D point cloud generation; iii) objects close to buildings such as trees with dense leaves that possess a geometry similar to building roofs, meaning that building elements cannot be differentiated using geometric features alone. All the aforementioned problems affect the accurate recognition of the individual roof faces of the building using point clouds, thereby leading to inaccurate building's roof face delineation (Vetrivel et al., 2015a). For example, Xiong et al. (2014b) reported that large errors inherent in the photogrammetric 3D point cloud cause a single planar roof face to be recognized as multiple smaller segments. This, consequently, affects the accurate recognition of roof topology of the building. For example, refer to Figure 8-1 depicting the above limitations of photogrammetric point cloud for accurate delineation of individual roof faces.

Figure 8-1 The roof faces delineated based on planar segmentation of photogrammetric 3D point cloud are projected over the image in varying colors.

2) Characteristics of the damaged scene: in particular urban damaged scenes are typically very complex and cluttered, and hence a 3D point cloud that only provides geometric information is not sufficient for handling such complexities (Vetrivel et al., 2015a). This is due to the geometric constraints used to infer the 3D objects in the erroneous point cloud are generally relaxed to some extent (Vetrivel et al., 2015c). With such relaxed geometric constraints even distinguishing roof segments of intact buildings from those of completely collapsed structures that result in debris heaps is ambiguous, as both possess similar geometry.

The problems mentioned above can be significantly mitigated if image-based features are used along with the 3D point cloud. This hypothesis is based on the previous studies which have demonstrated the pertinence of synergistic use of image and 3D point cloud features for building roof detection and delineation (Gilani et al., 2016; Hermosilla et al., 2011). However, they are mostly based on image features combined with LiDAR point clouds. So far, very few studies have used the combination of images and photogrammetric 3D point clouds for building roof detection (Rau et al., 2015b; Vetrivel et al., 2015a). However, they are not primarily concerned about the delineation accuracy of each individual roof faces as it is not critical for their application of building detection, unlike in the case of 3D reconstruction. Moreover, the ambiguity in recognition of debris heaps that possess similar geometry as roofs in 3D point cloud can also be resolved with the aid of image-based radiometric features, since they are recognized as strong indicators for those kind of damages (Radhika et al., 2012; Vetrivel et al., 2015a). For example, Vetrivel et al. (2016b) demonstrated that image-based texture features have the potential to identify image areas corresponding to damage evidences such as debris/rubble piles and spalling.

In summary, it is expected that the combined use of image and 3D point cloud features will help to overcome the aforementioned limitations in obtaining independent roof faces of the building with an accuracy required for an automatic building 3D reconstruction. To the best of our knowledge, no methods have been proposed yet addressing all aforementioned challenges in automatic 3D reconstruction of buildings with the combined use of images and their derived point cloud, particularly for damaged environments. Thus, the objective of this chapter is to develop a methodology for the automatic and accurate delineation of independent roof faces and thereby 3D reconstruction of the buildings in the damaged environment, by synergistically using the photogrammetric 3D point cloud and images of the UAV. Afterwards the corresponding images will be mapped to the reconstructed 3D model. This can serve as an input to any damage classification system (e.g., Saedi (2016)), facilitating further detailed damage assessment either in a manual or automated fashion. The overall methodology is a framework which comprises numerous methods, including image and point cloud segmentation, classification, damaged area detection, accurate roof delineation and 3D reconstruction of buildings. The proposed framework as a whole is novel. The background, justification and novelty specific to each method in the framework are provided in the respective sections.

8.2 Methodology

Photogrammetric 3D point clouds and images are considered for accurate roof face delineation for 3D reconstruction of buildings in damaged environments. The 3D point cloud- and image-features can be incorporated and processed either in object- or image-space. We prefer to process in image-space, since the photogrammetric 3D point clouds (representation in object-space) are erroneous, sparse and incomplete, with missing 3D points for some regions (Vetrivel et al., 2015c). However, even in image-space, the simultaneous use of 3D point cloud- and image radiometric- features at pixel level is not desirable. This is because of the varying density of point cloud, with the number of gaps leaving many image portions (pixels) with no corresponding 3D points. Hence, super-pixels or segments derived from object-based image analysis are considered as the primary entity for the incorporation of 3D features in image-space for further processing. Moreover, this kind of segment-based (super-pixels) approach has been demonstrated as being superior to pixel-based approaches by several studies, particularly in applications dealing with very high resolution images (Blaschke, 2010; Blaschke et al., 2014). The proposed methodology comprises six processes in a pipeline, which are explained in detail later in this section. As a first step, nadir-view images are selected as they are suitable for the delineation of roof faces of the buildings and less complex than oblique-view images. A selected image is segmented into super-pixels using an over-segmentation approach based on texture features. In the second step, the super-pixels are classified into roof, terrain, damaged (spalling or debris), vegetation and other category by utilizing either or both radiometric and 3D point cloud features. In the third step, the identified roof segments are refined such that each roof segment should be composed of only one planar surface. In the fourth step, the refined roof segments with similar radiometric and geometric features are merged to obtain the independent roof faces of the building. Finally, based on the delineated roof faces, a roof topology graph is constructed, whereby 3D reconstruction is performed. Then, the images are mapped to the corresponding portions of the reconstructed 3D model that can facilitate further detailed assessment. The detailed description of the above processes is provided below.

Step 1: Over-segmentation of image

The super-pixels construction is a mandatory pre-processing step in many image processing applications. For that numerous methods have been reported and among them methods such as SLIC, Quickshift and multi-resolution image segmentation are widely being used (Achanta et al., 2012b; Aksoy and Akcay, 2005a; Salem et al., 2013a). However, it is still challenging to obtain super-pixels from images designated for a specific application. This is because the characteristics of images vary based on the chosen application and it is critical to choose the appropriate features that can derive meaningful super-pixels for the actual problem. For example, here our objective is to derive super-pixels for very high resolution aerial images (GSD of 1-2 cm) captured over damaged areas. In general, objects in urban environments are highly heterogeneous in terms of geometry and radiometric characteristics. In particular, the damaged environment is cluttered and possesses a high degree of heterogeneity in its surface characteristics. Moreover, the radiometric heterogeneity of an image area is directly related to the spatial resolution of the image as nuances of an object captured in images increase with increasing spatial resolution. For example, consider Figure 8-2, which depicts the same object in two different spatial resolutions. The same area in the high spatial resolution image (right image in Figure 8-2) shows a higher degree of spectral heterogeneity than in lower spatial resolution (left image in Figure 8-2). Therefore, it is obvious that the direct true colour features (red, green and blue spectral bands) of an image are not adequate for consistent segmentation of very high resolution images, particularly for the application of damage assessment. Texture features are often reported as superior to spectral features for segmentation of images in many applications including complex urban scene segmentation of remote sensing images (Aguilar et al., 2012; Vetrivel et al., 2015c). Particularly, several studies have demonstrated that Gabor wavelets features are effective for urban scene segmentation of remote sensing images (Jiangye et al., 2014; Jiao and Deng, 2016). Also, Vetrivel et al. (2016b) demonstrated the ability of Gabor features in recognizing the damaged areas from very high resolution aerial images. Taking into account these inferences, in this study, the Gabor wavelet features are adopted to carry out the image segmentation.

Figure 8-2. Subset of airborne oblique image with average GSD of 14 cm (left) and subset of UAV image with average GSD of 1 cm (right). Both depict the same church in Mirabello, Italy after the 2012 earthquake

The Gabor wavelet comprises a set of filters, where each filter is tuned to extract information from the image at a specific frequency and orientation. In general, a large number of filters are used to construct the Gabor filter bank which is determined by the number of frequency scales and orientations considered for the feature extraction. The detailed procedure for Gabor wavelets filter generation can be found in Arivazhagan et al. (2006). However, the dimensionality of the Gabor features will be high (for e.g., 5 scales x 8 orientations = 40 filters), and some of the features might be highly correlated as there is a high possibility of capturing the same information at consecutive frequency scales. This can be partly addressed by eliminating the redundant information by adopting dimensionality reduction technique such as principal component analysis (PCA) (Gupta et al., 2013). However, still the dimensionality of the feature might be high. The next critical step is to choose the segmentation algorithm that is suitable to handle high dimensional features. A multi-resolution method is adopted for image segmentation as its ability to handle high dimensional features has been demonstrated in several studies (Darwish et al., 2003; Shao et al., 2014). The detailed procedure pursued for image segmentation is described below.

Procedure for image over-segmentation:
Input: *Nadir-view image*
Outcome: *super-pixels of the image*

a) Generate M x N number of 2D Gabor wavelet filters, where M and N are the number of frequencies and number of orientations considered for feature extraction, respectively.
b) Convolve the selected image with the generated filter banks to extract M x N number of feature images.
c) Apply PCA for the extracted M x N feature images to transform them to M x N principal component images.
d) Select the principal components that contributes more than 95% of total information, i.e. sort the eigenvalues corresponding to each principal component in descending order and select the top eigenvalues in the list that contributes 95% of total variance. Then select the principal components corresponding to the selected eigenvalues.
e) Perform image over-segmentation to the selected principal component images using the multi-resolution segmentation method (Benz et al., 2004).

Step 2: Classification of segments

After segmentation, the next step is to identify the roof segments of the buildings. The roofs can be identified based on their geometric characteristics, as they are always elevated and consist of horizontal or slanted planar surfaces. However, as stated before, trees and debris heaps often hinder the accurate recognition of roof segments as they may possess similar geometric characteristics. Therefore, the super-pixels are initially classified into five categories such as roofs, terrain, vegetation, damaged (spalling or debris) and others (e.g., façade elements). A multi-level, mixed unsupervised and supervised classification approach is adopted, as a specific set of features and classification strategy is required to recognize each category. For example, the vegetation can be classified based on simple thresholding of image spectral features. However, detection of super-pixels containing damage evidences such as debris and spalling requires texture features with a suitable classifier. Further, 3D information is required to distinguish between those damage evidences and to identify the roof and terrain segments. The pursued classification procedure is described below.

a) Identification of tree segments: The vegetation index (VI) = G/(R+G+B) is computed for each pixel in the selected image, where R, G and B correspond to the red, green and blue spectral bands of the image, respectively. The mean vegetation index is computed for each super-pixel. The super-pixels with mean VI greater than threshold T_V are classified as vegetation. The threshold T_V is computed using Otsu's method (Otsu, 1979). These super-pixels are eliminated from further analysis.

b) Identification of the damaged segments: The damage segments are identified using the CNN features based on the method described in Chapter 4. This method is adopted as it produced around 94% of overall accuracy in detecting the damages for the datasets used in this Chapter.

c) Classification of segments to identify the roof faces: To distinguish the roofs from other segments, the 3D information is required for which the 3D points derived from the images are considered. These 3D points are initially classified as terrain and off-terrain points, using the method proposed by Axelsson (2000b) that is implemented as part of the software *LAStools* (Rapidlasso, 2013). Subsequently, height normalization is performed, i.e. the height information is appended to each off-terrain point by differencing their Z value to the Z value of the nearest terrain point. Then the 3D points that are visible in the selected image are identified using the Hidden Point Removal method (HPR) (cf. Katz et al. (2007c)) and projected to the over-segmented image. In HPR, the hyper-parameter sphere radius R is defined using the technique described by Alsadik et al. (2014a). The segments containing the terrain 3D points are classified as terrain and removed from further processing. Further, the normal vector of each 3D point is computed based on its local neighbourhood. The 3D points with Z of normal $>T_Z$ and height $> T_H$ are classified as roof points. Then super-pixels are classified as roofs if at least 50% of their 3D points are roof points. This substantial relaxation is provided as some of the roof super-pixels might be under-segmented, and also because there might be misclassifications due to errors in the normal vector computation. The remaining super-pixels containing non-roof points are classified as other category.

Step 3: Roof segment refinement

The above classified roof super-pixels might be inaccurate as these are initially obtained using image radiometric features alone, in which over/under segmentation is inevitable. For example, any adjacent objects (e.g., two adjacent faces in gabled roof) in images possessing similar radiometric texture will lead to under-segmentation. Also, as stated earlier, the super-pixels belonging to debris regions might also be misclassified as roof, since they both possesses similar geometry. Hence, a roof segment refinement step is carried out to address these issues. This refinement step is to make each super-pixel to contain only one 3D planar segment. This is based on the assumption that a single roof face often consists of a single planar segment. The pursued refinement procedure is described below.

Procedure:

Inputs: List of super-pixels that are classified as roof, and 3D points that lie within each super-pixel.

Outcome: Refined super-pixels such that each super-pixel will contain only one 3D planar segment.

a) Identification of super-pixels to be refined

Identify and count the planar segments that lie within each super-pixel in the list. This is done by segmenting the 3D points corresponding to each super-pixel using a planar segmentation method. The super-pixels containing more than one 3D planar segment are considered for the refinement process. For example, consider the top polygon in Figure 8-3 as super pixel that is considered for the refinement process, as it comprises three different 3D planar segments.

b) Super-pixel refinement process

Select a super-pixel that needs to be refined. Identify the 3D planar segments within the super-pixel and list them according to their perimeter in descending order. There is a possibility that identified planar segments could belong to the same planar surface but are recognized as separate segments in object-space due to high error or gaps in the 3D point cloud. Hence, as a first step such planar segments are identified and merged to a single planar segment. If more than one planar segment persists after this merging process, then the boundary of each planar segment within the

super pixel is delineated and treated as a new super-pixel. Therefore, the final super-pixels will contain only one 3D planar segment. The overall refinement process is described below.
Select the largest 3D planar segment in the list and identify other planar segments that are coplanar to it. This is done by fitting an infinite plane to the selected planar segment and select the other 3D segments as coplanar to it, if their 3D points lie within a distance of T_D to the defined infinite plane (point to plane distance). The identified coplanar segments are considered as parts belonging to the same roof face and they are merged based on the following criteria:

i) Merge two planar segments that are considered to be part of same 3D surface if no other planar segments lie in-between. For example, consider Figure 8-3: if the planar segments corresponding to red and blue 3D points are found to be coplanar then they can be merged straightforwardly, since there is no other 3D segments in between them (refer to case 1 in Figure 8-3). The assumption here is that those segments within the super-pixel are recognized as separate segments in object-space due to gaps in the 3D point cloud.

ii) In case there are any planar segments in-between them, merge all of them, if additionally two more conditions are satisfied (refer to case 2 and 3 in Figure 8-3):
   - The in-between segments should also be the roof segments. The planar segments are identified as roof if their Z of plane normal > T_Z.
   - The average point to plane distance of the 3D points of the in-between segments and the plane defined for the 3D points of two coplanar segments should be less than T_{LD}.

   The above criteria are defined based on the assumption that any roof segment lying in-between two segments of the same roof surface should also be the part of that surface. It is assumed to be segmented as separate roof parts in object-space due to the high error in the 3D point measurements. Since it is anticipated as highly noisy region, set the above offset threshold T_{LD} much higher than the offset value T_D, which is used earlier for planar segmentation.

After the above merging process, the number of 3D planar segments in the super-pixel is analysed. If it contains only one planar segment, then it is considered as refined super-pixel and the refinement process is again continued for subsequent super-pixels in the list. Alternatively, if the super-pixel still contains more planar segments then it would be due to either one of the two possible scenarios: 1) the super-pixel may correspond to a debris region which possesses uneven distribution of the 3D points, leading to a higher number of planar segments, or 2) the super-pixel could be actually composed of different planar surfaces. In the former case, the super-pixel is classified as debris and removed from the super-pixel list. In the latter case, the boundaries of the planar segments within the super-pixel are derived and split into new super-pixels (see Figure 8-3, case 3). Among the newly derived super-pixels identify the ones corresponding to roof segment based on their normal vector as pursued earlier and add them to the super-pixel list.

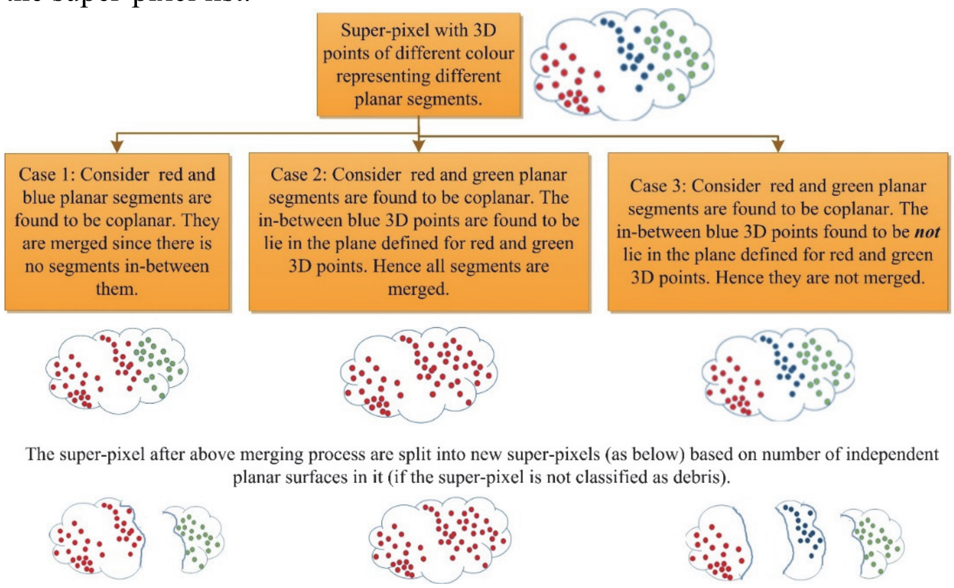


Figure 8-3. Criteria for refining the super-pixel if any two planar segments within it are coplanar

Step 4: Independent roof face construction

The refined super-pixels obtained from the above processes are considered for the independent roof face delineation process. This is achieved by merging the adjacent super-pixels possessing similar radiometric and geometric characteristics. The critical step is to decide whether two super-pixels are considered to be radiometrically and geometrically similar or dissimilar. The pursued approach to define the radiometric and geometric similarity between super-pixels is described below.

a) Radiometric similarity: Each super-pixel is assigned with a mean radiometric feature (MR) and a standard deviation (SD) computed based on the radiometric features (Gabor texture features) of the collection of pixels belonging to it. The two adjacent super-pixels are considered to be radiometrically similar if the distance between their MR is less than their SDs.

b) Geometric similarity: The adjacent super-pixels are considered to be geometrically similar if the angle between the normal vectors of their 3D planar segments is less than T_{Angle} and also if the distance between those two planar segments is less than $T_{Distance}$. However, from our experiments, these criteria are found to be inadequate for handling noisy super-pixels (i.e. super-pixels containing highly erroneous 3D points). This is because a large angle deviation is often observed between the normal vectors of the plane defined for the noisy super-pixels belonging to the same planar surface. Therefore, an additional criteria is defined later in this section for analysing the geometrical similarity between the noisy super-pixels. However, the critical step is the identification of noisy super-pixels. In the plane segmentation, the 3D points that lie beyond the allowed plane offset are left unsegmented. The number of these unsegmented points is expected to be high in the noisy super-pixels as the 3D points corresponding to them are unevenly distributed. Thereby, the super-pixel is considered as noisy if the ratio of the unsegmented to segmented 3D points in the super-pixel is greater than threshold P. Further, the additional criteria to identify the geometrical similarity between such noisy super-pixels is defined as: among the considered two super-pixels, select the noisier one and find the average point to plane distance (T_{PD}) of the unsegmented 3D points to the plane fitted for the segmented 3D points (this gives an approximate local error level). By taking this value as the offset threshold, find whether the 3D points of second super-pixel lie near the plane of the first super-pixel and label them as newly-segmented points. If the number of newly-segmented points is greater than number of segmented points already in the second super-pixel, then both super-pixels are considered to be coplanar and hence they are considered as geometrically similar.

Based on the above criteria, the adjacent super-pixels possessing similar radiometric and geometric characteristics are merged. Thereby the individual roof faces are delineated for the 3D reconstruction process.

Step 5: 3D reconstruction

The roof segments obtained using above roof refinement method are used to reconstruct the 3D model based on the method by Xiong et al. (2014a). In this method, the roof surfaces are constructed by deriving the boundary lines of the roof faces based on computing contours using the 2D α-shape algorithm and then generalized into polygons with fewer edges and regular angles. Finally, the outer boundary lines of polygon are projected onto ground as walls.

Step 6: Mapping of images to the 3D reconstructed model

The image corresponding to each segment in the 3D model is automatically mapped onto the model segment based on the following criteria,

1. The angle between the normal orientation of the segment in 3D model and the direction of the optical axis of the camera should be within the threshold T_{angle}.
2. The 3D segment of the model must lie within the boundary of the camera view and may not be occluded by other objects.
3. From the selected images based on the above two criteria, select one where the distance between the segment and the position of the camera is minimal to obtain better spatial resolution.

8.3 Results and discussion

8.3.1 Data used

A small region around the '*Church of Saint Paul*' in Mirabello captured by an UAV was considered for evaluating the developed methods. The images of the selected region were captured by a UAV from various heights, positions and views (oblique and nadir). The average GSD of the captured images is around 1 cm. A 3D point cloud of the scene was generated from 152 images with an average point density of 650 points per m^2. The selected region contains six buildings and among them only one was damaged.

8.3.2 Over-segmentation of the image

The multi-resolution image segmentation algorithm in eCognition was used for over-segmenting the image into super-pixels. In order to highlight the need for the proposed texture based segmentation, the selected image was segmented using both RGB values and the texture features independently. The segmentation parameters (compactness, scale and shape) were arbitrarily defined in order to portray that the derived texture features are effective and less sensitive to these parameter values compared to raw RGB images. For example, the over-segmented images obtained using both RGB and PCA-Gabor features with arbitrarily chosen parameter values are presented in Figure 8-4 and Figure 8-5, respectively. The segmentation based on RGB was very poor, yielding results that cannot be used for our purpose, i.e. the incorporation of 3D point features for obtaining geometrically uniform roof segments. Alternatively, the segmentation using the proposed PCA-Gabor based features provided relatively uniform segments, which is highly desirable for carrying out the further processes in the pipeline. The texture features were more useful than the RBG values in the image segmentation, as the Gabor features were found to effectively differentiate between the regions based on their surface patterns, irrespective of their intensity. For example, consider Figure 8-6 (taken from chapter 2), where the RGB image depicts the scene that contains three different types of building roofs, annotated as A, B and C, and a damaged region annotated as D. The Gabor feature images that are depicted in Figure 8-6 are the feature images corresponding to different frequencies and orientations. The roof segments A, B and C were clearly differentiated by those Gabor wavelet features. For example, for roof A, the Gabor feature 3 showed a strong signal, whereas B and C were highlighted by other Gabor features, as depicted in Figure 8-6. In all the feature images, the damaged region annotated as D in Figure 8-6 was found to show similar characteristics. This is because man-made objects have a dominant orientation; hence, the respective feature corresponding to that orientation shows a clear peak. Conversely, the damaged region has a gradient orientation in many directions; hence, they possess similar characteristics in most of the feature images corresponding to different orientations. The visual assessment indicates that Gabor features have the potential to differentiate the objects in the scene based on their dominant frequency and orientation characteristics.

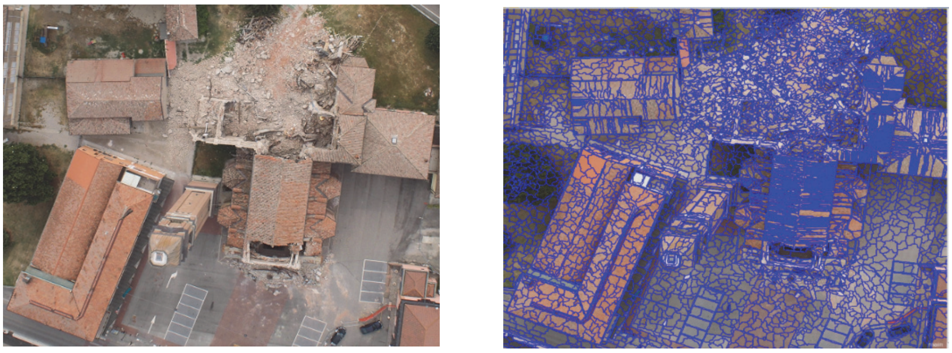


Figure 8-4. Selected nadir image (left) and super-pixels based on RGB using multi-resolution segmentation in eCognition for scale (50), compactness (0.5) and shape (0.5) (right).

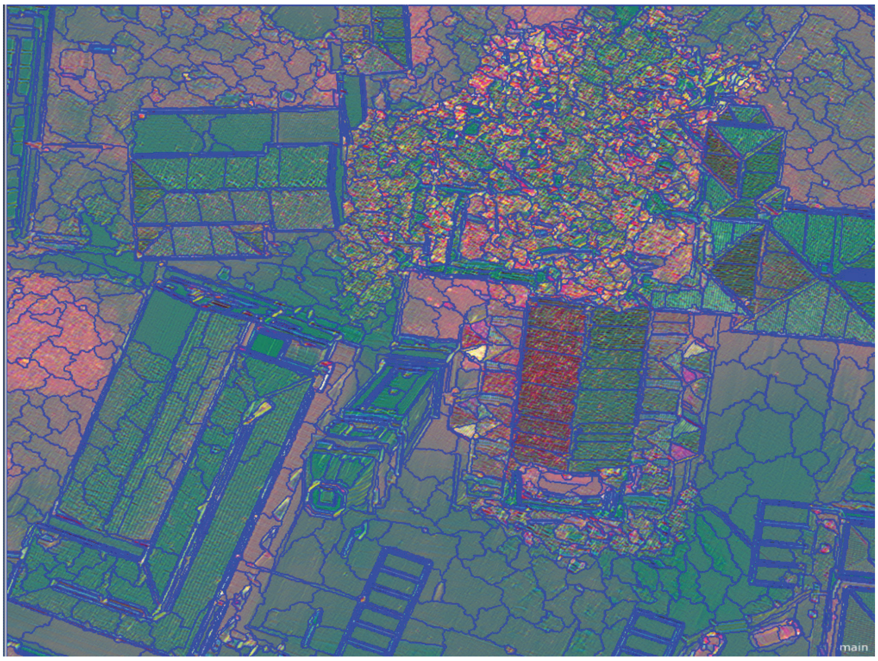


Figure 8-5. Super-pixels based on Gabor-PCA features of image depicted in Figure 8-4, using multi-resolution segmentation with same scale, compactness and shape values used in RGB based segmentation as in Figure 8-4.

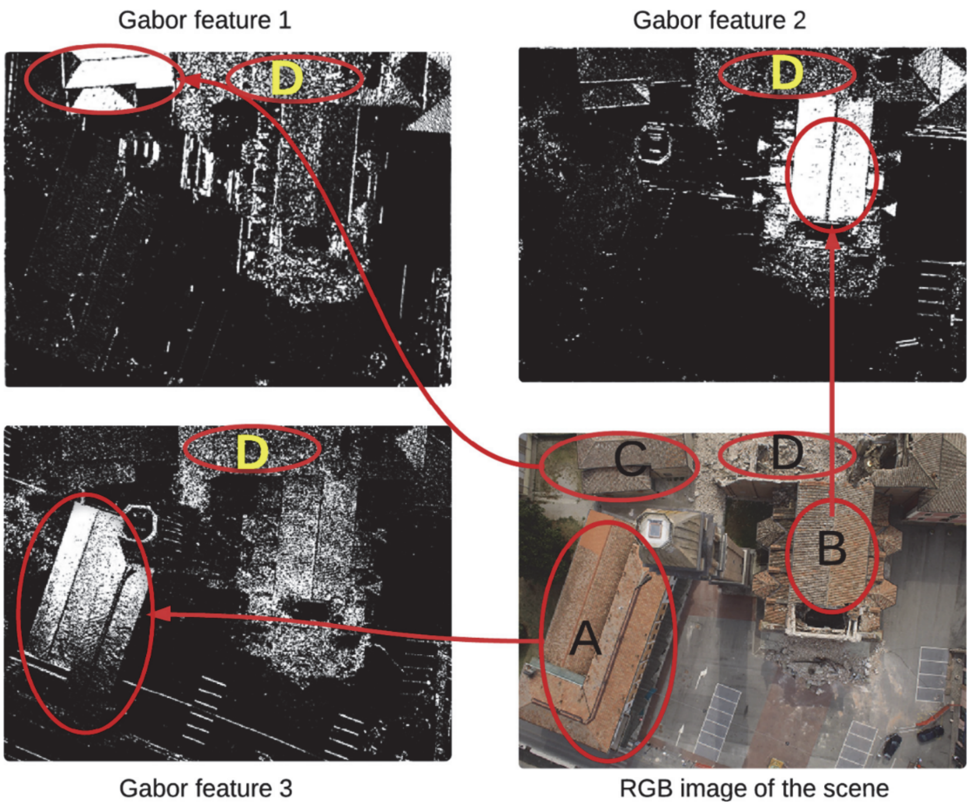


Figure 8-6. Gabor feature images extracted for different orientation and frequency parameters.

After image segmentation, the 3D point features were incorporated corresponding to each super-pixel, and the independent roof faces of the building were obtained using the procedure described earlier and the sample results are shown in Figure 8-7. Subsequently, the 3D reconstruction of the building and mapping of images to each element of the reconstructed 3D model was performed, and the results are shown in Figure 8-8 and Figure 8-9, respectively. From the results, even based on direct visual assessments it is clearly evident that the delineated roof faces based on the proposed method using both image and point cloud features were more accurate and geometrically regular compared to the roof faces obtained based only on planar based segmentation of 3D point cloud (cf. Figure 8-7 and Figure 8-1). The reconstructed 3D models were close enough to the original building if their roof faces were accurately delineated. For example, consider building B in Figure 8-6, which was not completely reconstructed. This is because that building contains a multi-layer roof, but in our method, the horizontal or slanted elements which are below roofs were considered as other elements of the building, e.g., extended balconies or staircases. Hence, the second layer roof elements were eliminated in this case and thus leading to inaccurate modelling of this building. However, this has to be addressed i.e., the roof detection method has to be improved to detect multi-layer roofs by recognizing other elements that are similar to roof structures. Moreover, in our case, the 3D point clouds were generated based on oblique-view images, where both top and side view information are typically available. Hence, in addition to roofs other lateral elements, such as façades, balconies and staircases, can also be used for 3D modelling, which would lead to more accurate models. This would be one of the crucial extensions of this work. Also, the presence of large gaps in the point cloud due to aforementioned reasons leads to inaccurate modelling. For example, see the building in Figure 8-7 which is marked with red circles indicating incomplete roof faces due to gaps in the 3D point cloud. Because of this, the same building in Figure 8-8 was not accurately modelled in the portions corresponding to gaps in 3D point cloud. Also, a significant structure – the tower in Figure 8-2 – was not detected and modelled in Figure 8-7 and Figure 8-8. This is because the top horizontal element of the tower was smaller and also much smoother. Hence, no significant number of 3D points corresponding to that particular portion was generated to be detected as a planar surface. Thus the tower was missing in the results of roof delineation and 3D reconstruction, even though the side portions of the tower were visible in the 3D point cloud. This is one of the examples for the aforementioned limitation in the roof-based 3D reconstruction approach.

The proposed framework consists of a number of methods and each comprises several tuneable parameters, whose definition is described in Table 8-1. These thresholds are defined based on the domain knowledge. For instance, the roof segment height is defined as greater than 3m from ground based on the fact that in general the roof heights will be at least two times greater than the average height (1.5 m) of a human being.

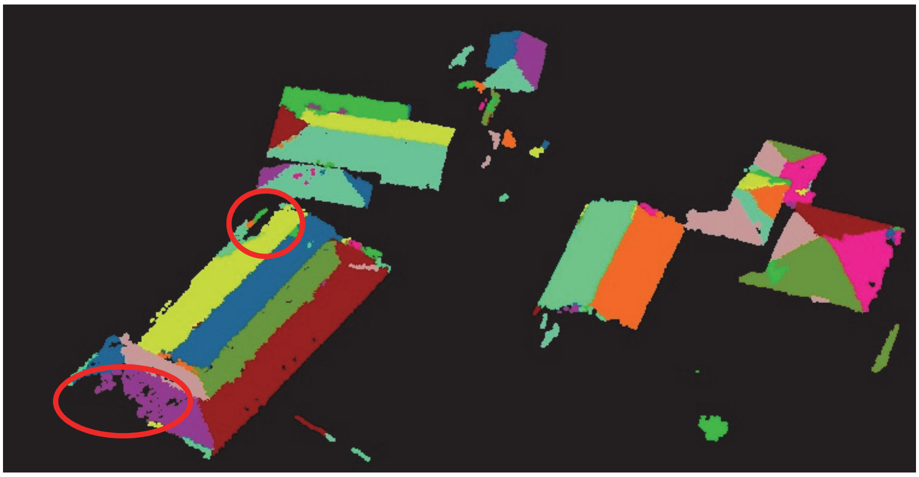


Figure 8-7. 3D point clouds depicting the delineated independent roof faces after roof segment refinement process. The red circles indicate the incomplete roof segments because of gaps in the 3D point cloud

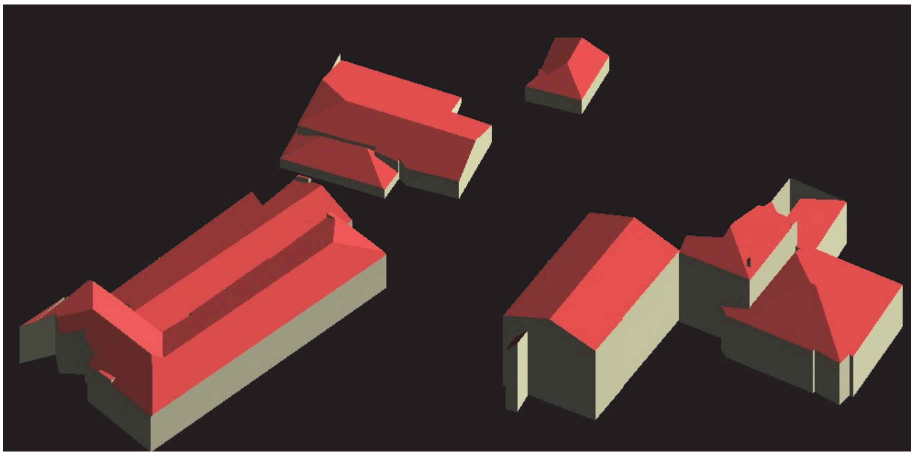


Figure 8-8. 3D reconstruction of building based on the 3D points of refined roof segments

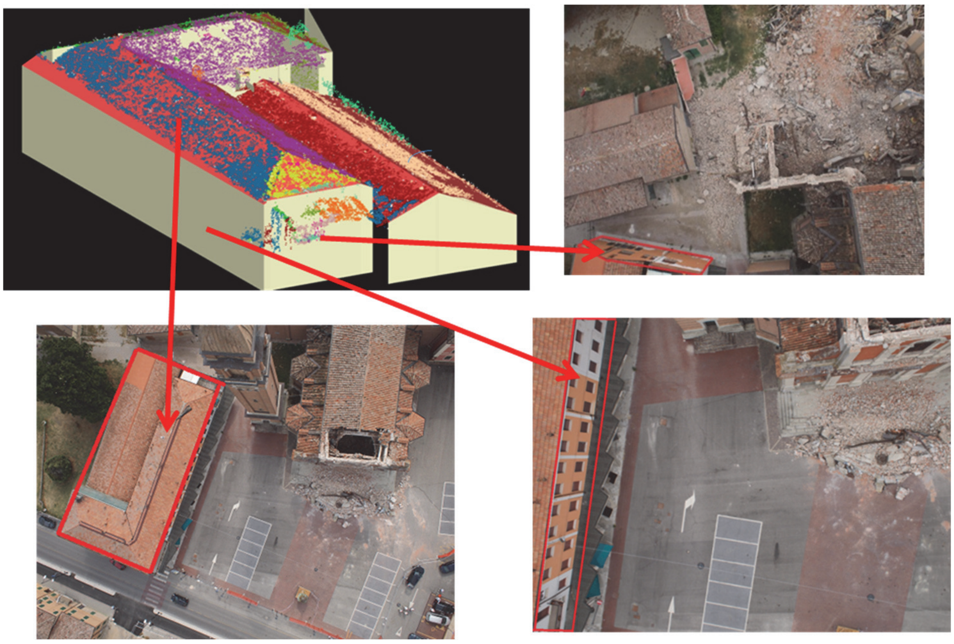


Figure 8-9. The images mapped to each segment of the model to provide radiometric information along geometric information for damage assessment

Table 8-1. The definition of various parameters used in the methods developed in this study.

Section	**Sub heading**	**Parameter definition**	**Description**
Over-segmentation of image	Procedure for image over-segmentation:	M= 5 and N=8	M and N are the number of frequencies and orientations used for generating the Gabor filter banks
Classification of segments	c) Classification of segments to identify the roof faces	T_Z = 0.4 degree, T_H= 5m	T_Z and T_H are the thresholds, where Z of normal of a planar segment is greater than T_Z and average height greater than T_H are considered as roof segments
Roof segment refinement	Super-pixel refinement process	T_D =0.5, T_{LD} = 1.5m	TD is the allowed offset for the plane and the points. If the distance

			between the 3D points that are lying between two coplanar segments will be considered as the points of those planes, if the point to plane distance is less than T_{LD}
Independent roof face construction	b) Geometric similarity	T_{Angle} = 5 degree $T_{Distance}$ = 1.5 m	Adjacent super-pixels are considered to be geometrically similar, if the angle between the normal vectors of their 3D planar segments is less than T_{Angle} and also if the distance between those two planar segments is less than $T_{Distance}$.

The major contribution of the proposed framework is the method for accurate roof face delineation. The other methods in the framework, such as vegetation detection, damage detection and 3D reconstruction, were adopted from previous chapters or from other published scientific articles. Hence the results corresponding to accurate roof face detection was given more focus. Pertaining to this, the proposed method for accurate roof face delineation was examined in another data set, based on images captured by manned aircraft over a portion of the city L'Aquila, Italy. The captured images had a spatial resolution of approximately 10 to 16 cm, which was much lower than the spatial resolution of the images of earlier dataset based on UAV. Therefore, the generated 3D point cloud of the current dataset based on images of manned aircraft was noisier than the 3D point cloud of UAV dataset. The detailed description about the dataset can be found in the data description Section (5.2) of Chapter 5. For the dataset based on a survey with a manned aircraft, the proposed roof face delineation method was examined. The image subset corresponding to the considered study area and the corresponding segmented 3D point cloud, and the final delineated roof faces are depicted in Figure 8-10. The results show that the roof faces obtained from the proposed method were more accurate compared to the roof faces obtained from planar segmentation of the raw point cloud (e.g., compare Figure 8-1 vs Figure 8-7 and Figure 8-10b vs Figure 8-10c). Though the spatial resolution of the images and quality of the 3D point clouds vary between the datasets, the threshold values defined in Table 8-1 were found to be working well for both datasets.

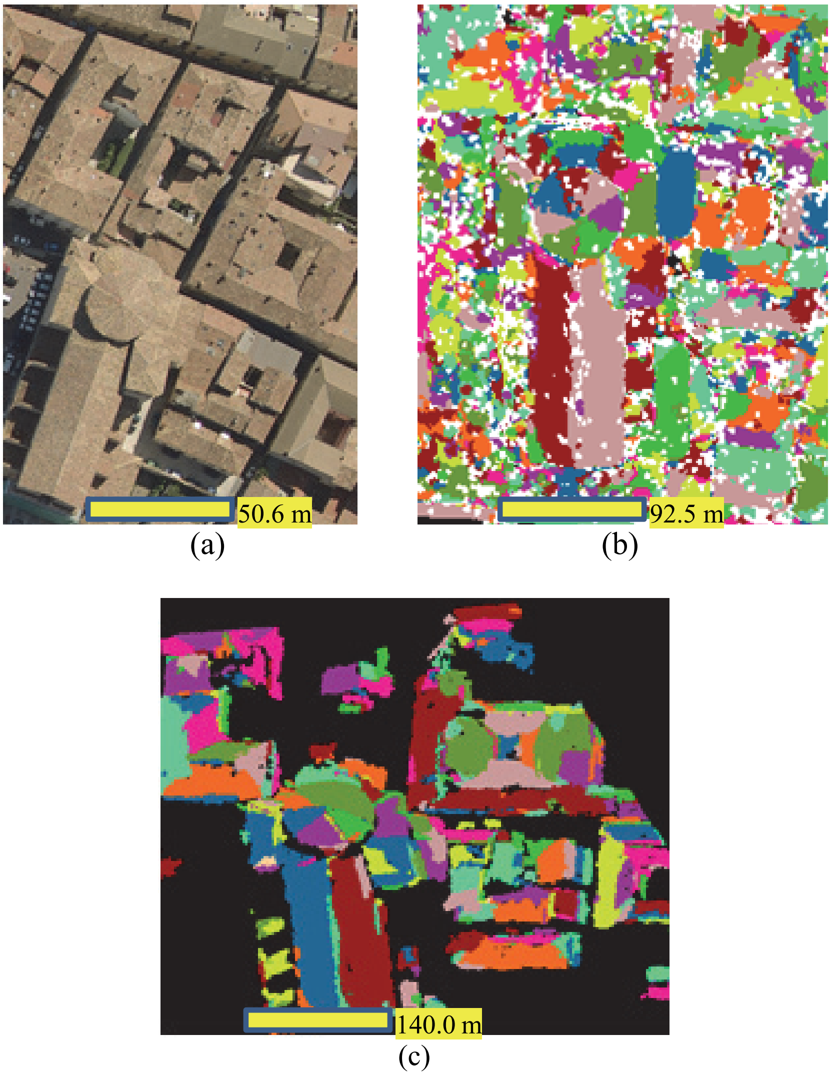


Figure 8-10 a) Image subset corresponding to the study area considered for examining the roof delineation method; b) planar segmentation based segmented 3D point cloud corresponding to the image subset; c) the final delineated roof faces based on the proposed method

8.4 Conclusion

3D models of buildings are highly desirable for a comprehensive damage assessment. Accurate delineation of roof faces from a 3D point cloud is the minimal requirement for 3D modelling of buildings. Often 3D point clouds derived from images are quite noisy, which hinders the accurate delineation of roof faces as discussed earlier. To address this, a framework was developed for the delineation of accurate roof faces of building from noisy 3D point clouds and images for 3D reconstruction of the buildings. There are several independent tasks within the framework such as image segmentation, damage detection, roof face delineation and 3D reconstruction. The framework was tested using the real world dataset: UAV images of Mirabello city, Italy captured after 2012 earthquake and the 3D point cloud derived from them. All 6 buildings in the point cloud were detected and reconstructed as a 3D model by the developed framework. The quality of the 3D models depends on the accuracy of the roof face delineation, where the model of the buildings with accurate roof face delineations was visually close enough to the shape of the original buildings. The gaps in the 3D point cloud hinder the accurate roof face delineation thereby affecting the quality of the reconstructed model. Also the images corresponding to the structural elements of the 3D model are mapped using the automated procedure. This representation, i.e., the 3D model mapped with images, is sufficient to satisfy the requirement for deriving the comprehensive damage classification of an individual building based on the recently reported classification systems, e.g., Saedi (2016). The major contribution of this research was the method for accurate roof face delineation. Hence, the framework up to roof face delineation was examined for another data set obtained with a manned aircraft, which was inferior to the UAV dataset in terms of the spatial resolution of the images and the quality of 3D point cloud. The roof faces delineated by the proposed method were more accurate than the roof faces derived from plane-based segmented 3D point cloud for this dataset as well. The values of the thresholds associated with the methods defined based on domain knowledge were the same for both datasets, and were found to be working well regardless of their varying characteristics. However, the generalization of these thresholds has to be examined further by considering several datasets with varying characteristics.

8.5 References of Chapter 8

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.,* 34 (11), pp. 2274-2282.

Aguilar, M.A., Vicente, R., Aguilar, F.J., Fernández, A., Saldaña, M.M., 2012. Optimizing object-based classification in urban environments using very high resolution Geoeye-1 imagery. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.,* I-7, pp. 99-104.

Aksoy, S., Akcay, H.G., 2005. Multi-resolution segmentation and shape analysis for remote sensing image classification, *Recent Advances in Space Technologies, 2005. RAST 2005. Proceedings of 2nd International Conference on*, pp. 599-604.

Alsadik, B., Gerke, M., Vosselman, G., 2014. Visibility analysis of point cloud in close range photogrammetry. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.,* II-5, pp. 9-16.

Arivazhagan, S., Ganesan, L., Priyal, S.P., 2006. Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters,* 27 (16), pp. 1976-1982.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing,* 33 (part B4/1), pp. 111-118.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing,* 58 (3), pp. 239-258.

Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65 (1), pp. 2-16.

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic object-based image analysis – towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing,* 87, pp. 180-191.

Darwish, A., Leukert, K., Reinhardt, W., 2003. Image segmentation for the purpose of object-based classification, *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. IEEE International*, pp. 2039-2041.

Dell'Acqua, F., Gamba, P., 2012. Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE,* 100 (10), pp. 2876-2890.

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.

Fernandez Galarreta, J., Kerle, N., Gerke, M., 2015a. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. *Natural Hazards and Earth System Science,* 15 (6), pp. 1087-1101.

Fernandez Galarreta, J., Kerle, N., Gerke, M., 2015b. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. *Natural Hazards and Earth System Sciences,* 15 (6), pp. 1087-1101.
Gerke, M., Kerle, N., 2011. Automatic structural seismic damage assessment with airborne oblique pictometry imagery. *Photogrammetric Engineering & Remote Sensing,* 77 (9), pp. 885-898.
Gilani, S., Awrangjeb, M., Lu, G., 2016. An Automatic Building Extraction and Regularisation Technique Using LiDAR Point Cloud Data and Orthoimage. *Remote Sensing,* 8 (3), pp. 258.
Gupta, R.P., Tiwari, R.K., Saini, V., Srivastava, N., 2013. A simplified approach for interpreting principal component images. *Advances in Remote Sensing,* 2, pp. 111-119.
Hermosilla, T., Ruiz, L.A., Recio, J.A., Estornell, J., 2011. Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data. *Remote Sensing,* 3 (6), pp. 1188.
Jiangye, Y., DeLiang, W., Rongxing, L., 2014. Remote Sensing Image Segmentation by Combining Spectral and Texture Features. *IEEE Transactions on Geoscience and Remote Sensing,* 52 (1), pp. 16-24.
Jiao, J., Deng, Z., 2016. Individual Building Rooftop and Tree Crown Segmentation from High-Resolution Urban Aerial Optical Images. *Journal of Sensors,* 2016, pp. 13.
Katz, S., Tal, A., Basri, R., 2007. Direct visibility of point sets. *ACM Transactions on Graphics (TOG)*, 26 (3).
Miura, H., Midorikawa, S., Kerle, N., 2013. Detection of building damage areas of the 2006 Central Java, Indonesia, earthquake through digital analysis of optical satellite images. *Earthquake Spectra,* 29 (2), pp. 453-473.
Nex, F., Rupnik, E., Toschi, I., Remondino, F., 2014. Automated processing of high resolution airborne images for earthquake damage assessment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XL-1, pp. 315-321.
Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics,* 9 (1), pp. 62-66.
Radhika, S., Tamura, Y., Matsui, M., 2012. Use of post-storm images for automated tornado-borne debris path identification using texture-wavelet analysis. *Journal of Wind Engineering and Industrial Aerodynamics,* 107, pp. 202-213.
Rapidlasso, 2013, Homepage lastools. http://lastools.org, (Accessed 15 Jun, 2014).
Rasika, A., Kerle, N., Heuel, S., 2006. 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, 8-11 May 2006, Enschede, the Netherlands. Enschede: ITC, 2006. 8 p.*
Rau, J., Jhan, J., Hsu, Y., 2015. Analysis of Oblique Aerial Images for Land Cover and Point Cloud Classification in an Urban Environment. *IEEE Transactions on Geoscience and Remote Sensing,* 53 (3), pp. 1304-1319.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J.D., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing,* 93, pp. 256-271.

Saedi, Y., 2016. Comprehensive remote sensing based building damage classification, *Faculty of Geo-Information and Earth Observation (ITC)*. University of Twente, Enschede.

Saito, K., Spence, R., Booth, E., Madabhushi, M., Eguchi, R., Gill, S., 2010. Damage assessment of Port-au-Prince using Pictometry, *8th international workshop on remote sensing for disaster management*. Tokyo.

Salem, M., Ibrahim, A.F., Ali, H.A., 2013. Automatic quick-shift method for color image segmentation, *Computer Engineering & Systems (ICCES), 2013 8th International Conference on*, pp. 245-251.

Shao, P., Yang, G., Niu, X., Zhang, X., Zhan, F., Tang, T., 2014. Information extraction of high-resolution remotely sensed image based on multiresolution segmentation. *Sustainability,* 6 (8), pp. 5300.

Sun, S., Salvaggio, C., 2013. Aerial 3D building detection and modeling from airborne LiDAR point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6 (3), pp. 1440-1449.

Tong, X., Hong, Z., Liu, S., Zhang, X., Xie, H., Li, Z., Yang, S., Wang, W., Bao, F., 2012. Building-damage detection using pre-and post-seismic high-resolution satellite stereo imagery: A case study of the May 2008 Wenchuan earthquake. *ISPRS Journal of Photogrammetry and Remote Sensing,* 68, pp. 13-27.

Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2015a. 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, pp. 61-78.

Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2015b. Segmentation of UAV-based images incorporating 3d point cloud information. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XL-3/W2, pp. 261-268.

Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2016. Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach. *Remote Sensing,* 8 (3), pp. 231.

Xiong, B., Elberink, S.O., Vosselman, G., 2014a. Building modeling from noisy photogrammetric point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 2 (3), pp. 197.

Xiong, B., Oude Elberink, S., Vosselman, G., 2014b. Building modeling from noisy photogrammetric point clouds. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.,* II-3, pp. 197-204.

Xiong, B., Oude Elberink, S., Vosselman, G., 2014c. A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing,* 93, pp. 227-242.

9 The application of developed methods for building damage assessment based on the EU-FP7 project RECONASS

9.1 Background

As mentioned in Chapter 1, this research is embedded in the research project RECONASS, which aimed to develop a near real-time damage assessment system, referred to as RECONASS system. Remote sensing and a wireless sensor network (WSN) based damage assessment systems are the two major subsystems of the RECONASS system. The remote sensing subsystem was solely developed by our research group at ITC. The objectives of this subsystem were already described in Chapter 1. In order to accomplish those objectives, this subsystem was designed to carry out three major tasks based on UAV images and 3D point clouds derived from them: 1) building detection, 2) automatic detection of externally visible damage evidences such as spalling, debris, rubble piles, broken and inclined elements, cracks and debris volume quantification, and 3) 3D modelling of buildings to systematically integrate the damages identified along different portions of a building to derive a comprehensive building level damage label based on the damage classification scheme. The methods for accomplishing these tasks have been developed as part of the remote sensing subsystem in RECONASS. All methods developed in this PhD thesis were part of this subsystem. In order to demonstrate all subsystems within RECONASS system, a 3-story building equipped with RECONASS sensors was constructed in Sweden. A pilot experiment was conducted by deliberately damaging the building based on two massive blasts. The first blast was an explosion of 400 Kg of TNT outside the building, and the second was a smaller blast of 16 Kg of TNT inside one of the rooms of the building. The remote sensing-based subsystem is evaluated using the data captured by a camera on the UAV after the two blasts. Apart from methods reported in this thesis, we have also developed several other methods in order to accomplish the objectives of RECONASS. For example, methods such as identification of inclined elements, debris volume quantification, crack detection and final damage classification to derive the building level single damage label were developed and delivered to European commission as technical reports. For the sake of completeness and to provide a broad picture of the remote sensing subsystem in RECONASS, the results of these methods are also presented along with the results of the methods reported in this thesis.

9.2 Demonstration of the methods in the remote sensing subsystem

The system is demonstrated in three scenarios based on the availability of the data: 1) when both pre-event CAD model and post-event images and point cloud are available for assessment; 2) when pre- and post-event images and 3D point cloud are available for assessment; 3) when only post-event images and 3D point cloud alone available for assessment. The descriptions about the demonstration strategy and their results are reported together in the below respective subsections

9.2.1 Data description

In the RECONASS pilot experiment a hexacopter Aibotix Aibot X6 V2 UAV was used, i.e. a rotary wing system with 6 rotors. The chosen UAV can lift a payload with a total weight of up to 2 kg, hence it is ideal to carry a professional DSLR camera. In our case a Canon D600 with a Voigtländer 20mm fix zoom lens was employed. The UAV also comes with a GNSS receiver in order to geotag the images.

Figure 8-11 Aibotix Aibot X6 V2

The image acquisition after the explosion was planned using a circular waypoint layout, that is, the drone flies a circle in a given height. The center of this circle was above the area of interest (building) and the circle radius and nick angle of the camera were chosen in way to optimally cover the building, see Figure 8-12 for a birds-eye view of the scene and the cameras. In the experiment we chose a flying height above ground of 55m and 65m, respectively, which resulted in an average ground sampling distance (GSD, pixel size at the object) of 1.5 cm.

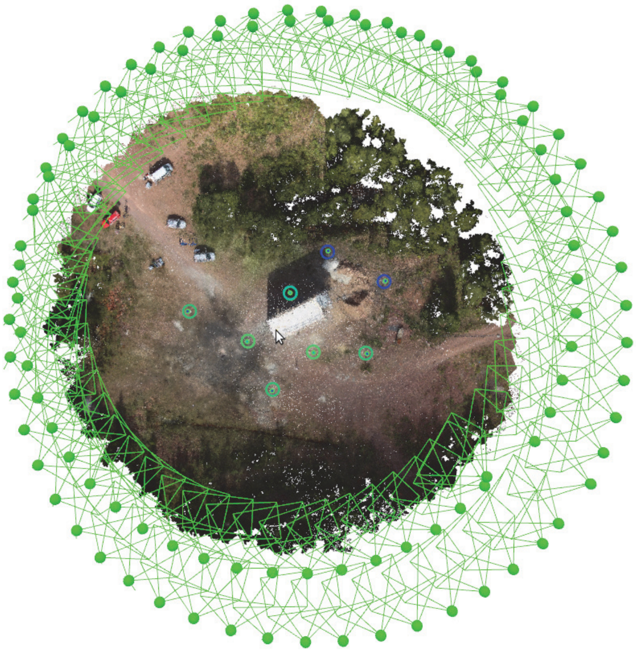


Figure 8-12 Circular layout of images, also indicating the ground control points (green circles)

In total 103 images were captured. Since the BIM or CAD model of the building of interest maintains a local coordinate system we preferred to generate 3D point cloud in that system to facilitate the direct comparison of the CAD model and 3D point cloud for damage assessment. This was one of the experiments conducted for demonstrating the potential of the developed methods which is reported later in this chapter. However, the GNSS receiver on board the UAV provides image locations in a global coordinate frame. Hence a procedure to co-register the images with the BIM model was developed, to generate a 3D point cloud in the local coordinate system. The overall procedure is described below[3]:

- Before the detonation took place some corners of the building were surveyed with a professional GNSS receiver.

[3]This co-registration procedure was solely developed and implemented by Prof. Markus Gerke

- Using both the GNSS-coordinates of the corners and the corresponding coordinates in the local system, the 6 transformation parameters were computed.
- After the initial image orientation using the GCPs on the ground, the image locations and tie points were transformed to the local model, defined by the BIM model by applying the 6 parameter-transform.
- As a last step the sparse point cloud in the local building coordinate system was fine-registered with the point cloud derived from the BIM-model by employing a variant of the ICP (Iterative Closest Point) algorithm. The image locations were then also transformed accordingly, and finally the image orientation parameters were defined in the BIM model system.
- A dense 3D point cloud was generated based on the same procedure described in earlier chapters.

The detailed procedure of this co-registration process can be found in Chapter 6 of RECONASS deliverable D4.3, available at http://reconass.eu/.

9.2.2 Demo 1: Damage assessment by comparing the CAD model and post-event point cloud and images

In this scenario, the post-event UAV images and point clouds derived from them, and pre-event 3D points of the CAD model were the inputs for the remote sensing subsystem.

The damages to the RECONASS building were identified by comparing the post-event 3D point cloud generated from UAV images to the externally visible elements of the CAD model of the monitored building. This assessment included the identification of various damage evidences and thereby classifying the building elements into different classes as described below.

1) **Broken elements**: Missing CAD elements in the post-event point cloud were identified by performing the element-wise comparison of post-event point cloud with the point cloud generated based on CAD model. The details about this method can be found in section 5.3.2.2 in chapter 5.
2) **Inclined elements:** The difference in angle of plane normal for the corresponding planar elements in both post-event point cloud and CAD model was computed. If the angle difference was greater than certain threshold (3 degree), then the respective planar element is classified as inclined.

3) **Debris:** The debris and spalling regions were identified based on recognizing the unusual radiometric using information from images based on the method described in chapter 4. The presence of these patterns on the ground surface was considered as debris and its volume was quantified. The volume quantification method is described in Chapter 5 of RECONASS deliverable D4.1, available at http://reconass.eu/.
4) **Cracks:** Cracks were identified based on the radiometric characteristics and geometrical shapes. For example, the darker region with linear shapes on intact planar segment was classified as cracks. The description about the crack detection method can be found in Chapter 6 of RECONASS deliverable D4.3, available at http://reconass.eu.
5) **Intact:** The elements with no above mentioned damage evidences were classified as intact.
6) **Occluded:** The CAD model elements invisible in the images were identified and labelled as occluded. The occlusion detection was carried out using HPR as described in chapter 5.

   The classified 3D points of the CAD model based on above mentioned classes were the output of damage detection process as depicted in Figure 8-13. From an image-based system, it is possible to map only the visually recognizable damages. Therefore, the evaluation of damage assessments were carried out by visually comparing the automated results with the damages visible in the images and the 3D point cloud and the results are discussed below.

9.2.2.1 Results of blast 1:

The classified point cloud of a CAD model representing the damage information (automatically identified by this subsystem) along with corresponding image are provided in Figure 8-13 for visual inspection. The monitored building in Figure 8-13 is represented in two different orientations in such a way that all the sides of the building are visible. The detected damages in the point cloud and images were highlighted and annotated using the letters A to H for analysis, which are briefly described below.

Damage region A: It depicts two broken elements leading to holes in the building and debris in the ground which were correctly identified by our automated methods. This was the only region affected by the blast while other regions (B-H) were detected as damages due to other reasons which are described below.

Damage regions B, G and F: These regions were detected as broken elements as there was difference in the CAD model and the actual constructed building. For example, the elements in the region B and G in CAD model were corresponding to infill walls which were not actually constructed. On the other hand, the building elements corresponding to region F were detected as missing as they were hidden by the staircase which was not part of the CAD model design. Therefore, these elements in CAD model were annotated as damage since they were missing in the post-event point cloud. The developed algorithms worked well by detecting the missing elements, though the reason was not damage. This observation makes clear that in a real case scenario the BIM or CAD model of the building must actually be well maintained and be identical to the as-built-status.

Damage region C: The elements detected as damaged in region C were not actually damaged. They were the structural elements (beams) embedded in the roof segment. In this case, damage was detected based on comparing pre-event 3D CAD model based point cloud with post-event image-based point cloud. For comparison, only the visible CAD model elements were considered. While the visibility analysis is conducted for the elements in CAD model, the structural elements (beams) in region C were marked as visible by our HPR algorithm as they were embedded in the visible roof segment. Hence, they were considered for comparison with post-event point cloud. However, these elements were absent in the post-event point cloud as they were not visible in the images captured by UAV. Hence these elements were wrongly classified as damaged.

Damage region D: This intact element was wrongly detected as inclined because of the presence of noise leading to large variations in the Z of normal of the 3D points corresponding to this region.

Damage region H: This region is detected as damaged due to the gaps that exist in the generated post-event point cloud. The reasons for occurrence of these gaps is not obvious since the criteria for generation of 3D point cloud were satisfied (e.g., these regions are visible in more than three cameras and the surface characteristics of these regions were similar to the other regions which had corresponding 3D points).

There were no visible cracks in the exterior building elements after blast 1. Overall, the developed algorithm detected 1) all the missing elements due to damage, and 2) the debris region. The wrongly detected damage regions D and H alone can be considered as false positives which resulted from the limitations in the generated point cloud, while other wrongly detected damaged regions B, C, G and F cannot be considered as false positives as they were detected as damage because of the difference in the CAD model and the constructed building.

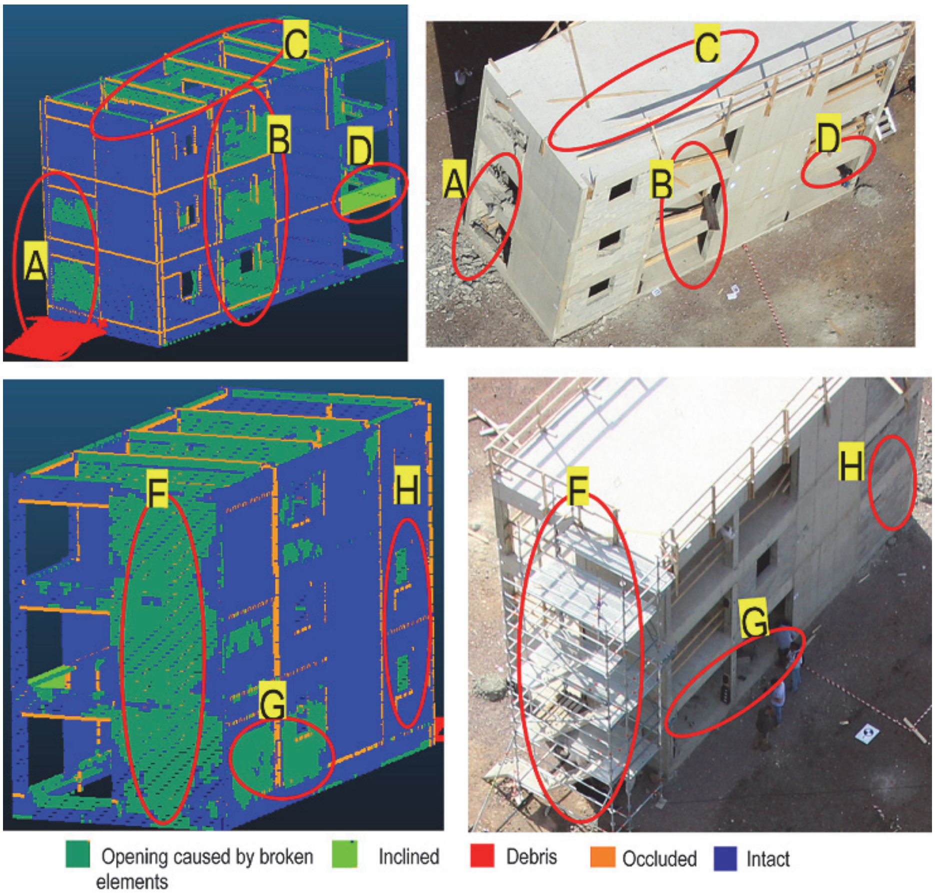


Figure 8-13. The classified point cloud of CAD model depicting the element-wise damage information (left-top and left bottom) of blast-1 identified automatically by the developed methods and the corresponding images (right-top and right bottom) for visual inspection. The damaged regions were annotated using alphabets A-H which are briefly described in the text.

9.2.2.2 Results of blast 2:

Similar to above results, the buildings are depicted in two different views in order to make all the sides of the buildings visible. The detected damages after the blast were highlighted and annotated in the respective point cloud and images using alphabets A to H for analysis, as depicted in Figure 8-14 and it is described below.

Damage region A: All broken elements and debris regions were detected by our method.

Damage regions B, C, F, G and H: The same reasons as described in section 0.

Damage region D: The element annotated as D was correctly identified as inclined, the angle difference between the element in the CAD model and post-event point cloud was estimated as 4.4 degree.

Damage region I: The presence of minor crack in this region was identified by our method.

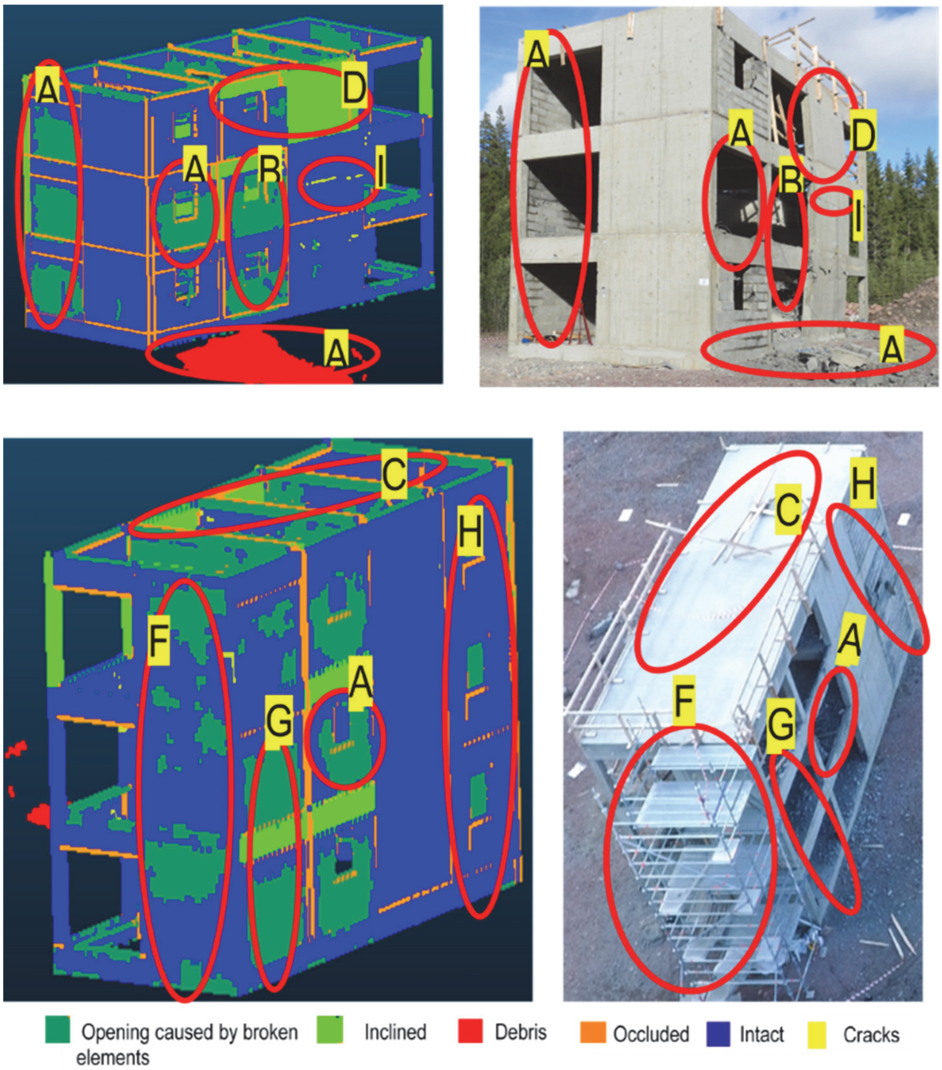


Figure 8-14. The classified point cloud of CAD model depicting the element-wise damage information (left-top and left bottom) of blast-2 identified automatically by the developed methods and the corresponding images (right-top and right bottom) for visual inspection. The damaged regions were annotated using alphabets A-H which are briefly described in the text.

Overall, concerning both blast 1 and blast 2, a total of eight externally visible damaged regions was identified by visual inspection: five missing infill walls, one inclined façade and two debris regions on the ground. All these damaged regions were identified by our method, however with a few false positives (cf. section 6.3.1.1).

9.2.3 Demo 2: Damage assessment by comparing the pre- and post-event point cloud and images:

This is the scenario where the oblique images of pre- and post-event are available. The 3D point clouds of pre- and post-event were generated independently and compared for identifying the missing elements. The detailed procedure of this approach can be found in chapter 5. The detected damaged regions by this approach were annotated in Figure 8-15, where all missing elements due to damage were correctly identified, and there was no false positive detection. Even the damage to the wooden fence on the roof top (region A in Figure 8-15) was correctly identified, portraying the robustness of this approach.

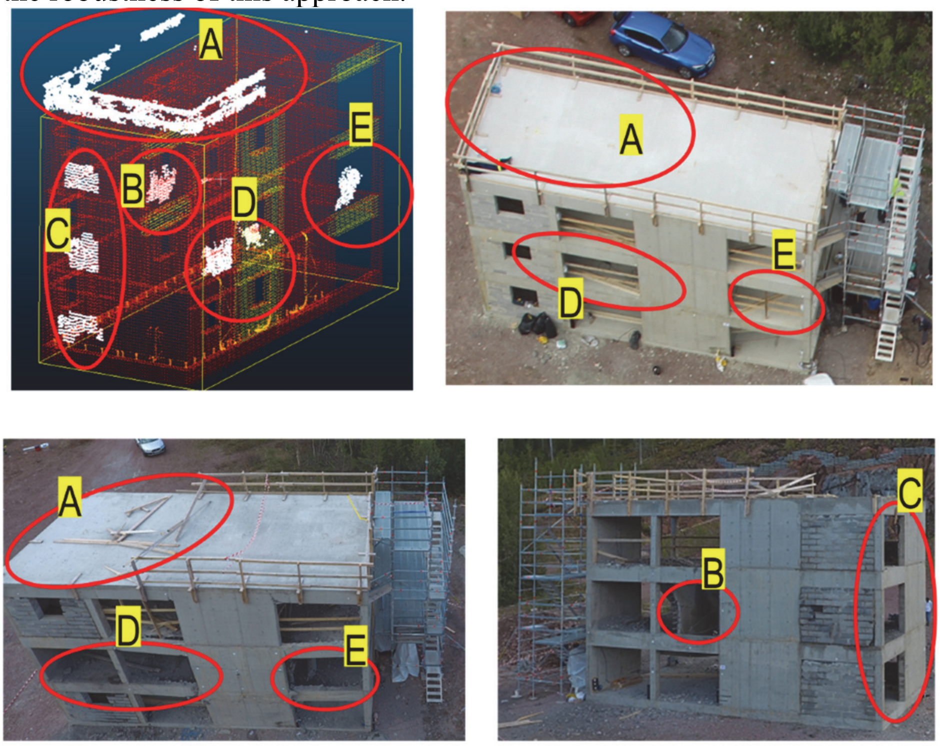


Figure 8-15. The detected missing element by comparison of pre- and post-event point cloud were highlighted using alphabets A-E in the point cloud of CAD model for reference (left-top) and the corresponding elements are highlighted in pre-event image (right-top) and post-event images (left- and right-bottom). The annotated alphabets A-E are briefly explained in the text.

9.2.4 Demo 3: Damage assessment from post-event data alone

This is the scenario where only the post-event images and point cloud are available. The missing elements are assumed to often create an opening to the building and further this appears as a gap in the 3D point cloud. However, for several reasons gaps in the 3D point cloud can correspond to a normal and desired feature (e.g., architectural elements). Also, due to the 3D information generation process, gaps can also be created in case of partial building occlusion (e.g., by vegetation) or image matching problems. Therefore, as a first step, the gaps in the point cloud are detected and the gaps with debris or spalling evidences around them are classified as gaps due to structural damage, since any deformation in the concrete surface creates a sign of spalling or debris around the deformed region. The detailed procedure for the gap based damage detection is provided in RECONASS deliverable D4.1, and can be found in Vetrivel et al. (2015). The debris and spalling regions were detected (annotated using red polygon in Figure 8-16) by the improved version of damage detection framework reported in Chapter 4. The damaged regions detected by this approach were annotated using alphabets A-E in Figure 8-16, where only the gap depicted in region A was damage, while other gaps B-E (the openings due to other reasons) were detected as damages, since debris and spalling evidences are found around them (cf. Figure 8-16). This shows that, although the accurate mapping of the damage evidences such as debris and spalling from the post-event images is feasible (can help to infer the severity of the damage), it is still challenging to interpret the actual damaged elements based on these evidences. This highlights the importance of pre-event data as a reference for accurate damage mapping.

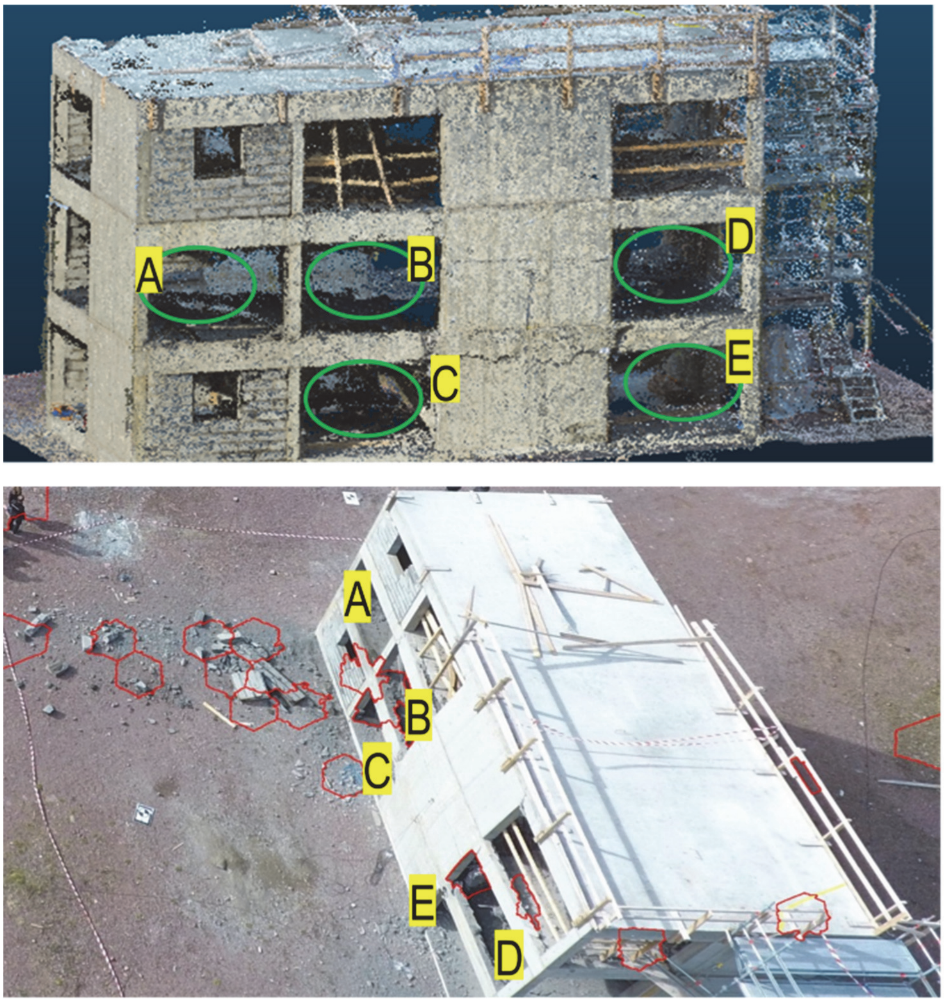


Figure 8-16 Point cloud of the building after blast-2 highlighted and annotated the openings detected as damage by the method based on post-event data alone (top). The image corresponding to the point cloud in which the debris and spalling regions detected by our method were highlighted using red polygons. The annotated alphabets A-E are briefly explained in the text

Finally, a damage classification system was developed within the RECONASS framework for deriving a building level damage based on systematic aggregation of various aforementioned damage evidences from different sides of the building (cf. Chapter 4 of RECONASS deliverable D4.3, available at http://reconass.eu/). However, the results of the damage classification are not provided and discussed here as it is beyond the scope of this thesis.

9.3 Conclusion

In this chapter, the potentials of methods developed in earlier chapters for damage detection and 3D modelling of the building were demonstrated in several aspects using datasets from the pilot experiments conducted in Sweden. The major inference from the demonstration results is that the level of assessment and accuracy depends on the kind of data available for the assessment. In particular, the availability of both pre- and post-event data facilitates more detailed and accurate assessment than assessments based on post-event data alone. Overall, the developed methods are found to be working well when demonstrated using the data sets from the pilot experiment. No major issues arose while transferring the methods developed and tested based on several real-world datasets as mentioned in earlier chapters to the new data from this pilot project.

10 Synthesis

This research focused on developing methods for the automated extraction of information from remote sensing images that could aid stakeholders involved in disaster management to carry out fast response and recovery actions. As emphasized several times in this thesis, a stakeholder involved in different phases of disaster management needs damage information with different levels of abstraction in a specific format. For example, a specific stakeholder may require the locations of collapsed buildings in 2D map format, while another stakeholder may require damage information along every element of the building annotated on a 3D model of the building. The fundamental information required from remote sensing data for producing these kinds of damage representations are: 1) automated delineation and 3D modelling of buildings, and 2) automated recognition of various kinds of damage evidences required for damage assessment such as spalling, openings in a building due to damage and debris/rubble piles mapping and quantification. The methodologies for performing the aforementioned tasks have been developed in this thesis that are especially (but not only) suitable for oblique view images (for the reasons mentioned earlier) either from manned or unmanned aerial platforms and 3D point clouds derived from them. The developed methodologies and their results are briefly summarized and discussed below.

Any severe structural damage often creates an opening (hole) to the buildings. Among several above-mentioned damage evidences, we initially focused on developing methodology for the identification of the structural openings (damage) caused by the disaster event. This is because we use photogrammetric 3D point clouds in addition to the images for various tasks including building detection and 3D reconstruction. In general, the photogrammetric point clouds are highly noisy and often contain gaps that can be due to several reasons, such as the architectural design or partial building occlusion (e.g., by vegetation), or image matching problems. It is essential to distinguish the gaps in 3D point cloud due to the aforementioned reasons from the gaps created due to damage, in order to effectively carry out other tasks that are based on 3D point clouds (particularly, accurate delineation of individual elements of the building and 3D modelling of the building). The methodology for mapping the gaps created due to damage was developed and reported in Chapter 2. The methodology includes three steps in which automatic building delineation from aerial images and photogrammetric 3D point cloud was the first step. While conducting this research in 2014, methods were already available for building delineation from 3D point cloud which were mostly based on LiDAR technology (Rottensteiner et al., 2014; Sun and Salvaggio, 2013). However, it was challenging to directly adopt these methods to accurately delineate the buildings from photogrammetric 3D point clouds as it is often inferior in quality compared to LiDAR point clouds. Hence, a method suitable for building delineation from noisy photogrammetric point cloud was developed. Also, it was often difficult to distinguish between densely-leafed trees and the roof segments of buildings from the noisy photogrammetric point cloud. These issues were largely addressed in the developed method by using spectral information from images in addition to 3D point clouds. Overall, the developed methodology detected 96% of buildings in the urban areas selected for the experiment. However, the building delineations were not accurate compared to the actual boundary of the building, but it was sufficient for the desired gap based damage detection process. The second step was the gap detection process. The method developed for gap detection was based on a voxelization of the 3D point cloud where the empty voxel which were visible from the number of camera views sufficient to generate 3D point was marked as a gap. This method worked well, detecting all gaps in significant sizes in the point cloud. The third step was gap classification. The gap was classified as an opening due to damage if any damage evidence such as spalling or debris were found around them. Hence, a method for the detection of these damage evidences was required. In literature, several methods for detecting these kinds of damage evidences from remote sensing images were already reported at the time of conducting this research in 2013 (Dong and Shan, 2013). Most of the methods were based on a supervised classification approach where predominantly textures were used as feature for building the classifier (Dong and Shan, 2013; Ma and Qin, 2012b; Radhika et al., 2012). In most cases, statistical textures such as grey level co-occurrence matrix based features were used. However, the texture features such as HoG and Gabor wavelets were reported to be more efficient than GLCM features in several classification based applications in computer vision and remote sensing (Radhika et al., 2012; Ruiz et al., 2004b; Stavrakoudis et al., 2011). Hence, it was hypothesised that these features can be useful for damage detection application as well. Thus, a methodology for mapping the spalling, debris and rubble piles from images was developed based on these two kinds of texture features –Gabor and HoG using the supervised classifiers SVM and Random Forests. The classifier based on Gabor features with Random Forests performed best, identifying 95% of the damaged regions in the considered study area. However, the generalization capability of the developed classifier was very poor, with the quality measures decreasing by around 30% when tested on another independent data set. Since the method produced an accuracy of 95% for the study area, the gap classification process was successful, with all 21 detected gaps correctly classified.

The mapping of damage evidences such as spalling, debris and rubble piles was highly desired as they are the strong indicators of severe structural damage. As mentioned above, the methods based on texture feature reported in Chapter 2 were found to be useful for detecting these damage evidences. However, they strongly suffered from generalization problem, which was mainly due to the varying characteristics of image (various views and scales), scene and damage pattern. In order to improve the classification accuracy, a state-of-the-art (in 2013) Bag-of-words (BoW) approach was adopted to develop a methodology for automated mapping of earlier mentioned damage evidences (cf. Chapter 3). Three kinds of texture features, HoG, Gabor and SURF, and three different classifiers (SVM, Random Forests and Adaboost) were independently used for constructing the BoW-based classifiers. The developed classifiers were tested with four different data sets that varied greatly in terms of image and scene characteristics. The results of the developed BoW-based method were compared with a conventional global representation approach (reported in Chapter 2), using the same set of image features and classifiers for all datasets. The BoW framework outperformed the conventional global feature representation approach in all scenarios (i.e. for all combinations of feature descriptors, classifiers and datasets), and produced an average accuracy of approximately 90%. Particularly encouraging was an accuracy improvement by 14% (from 77% to 91%) produced by BoW over global representation (reported in Chapter 2) for the most complex dataset, which was used to test the generalization capability. Owing to its effectiveness, the BoW framework was adopted with interest by the research community for disaster damage detection using remote sensing data. For instance, the BoW framework developed in this research with same set of features – HoG, Gabor and SURF was used by Tu et al. (2016b) for detecting the damaged roof tops using the very high resolution airborne image captured after 2014 earthquake in Beichuan. They reported an overall accuracy of 90% using HoG in BoW framework. Tu et al. (2017) used BoW as a feature for detecting damaged regions based on change detection approach by comparing the pre- and post-event data.

Though the method based on a BoW framework (reported in Chapter 3) produced 90% accuracy it still lacks the generalization capability which was described in detail in Chapter 4. In the meantime, the deep learning features such as CNN features became state-of-the-art, outperforming the BoW features in many domains, including related applications in remote sensing. These features were highly recognized by the research communities particularly, for its generalization capability. A methodology for identifying the damaged regions based on CNN features was developed and reported in Chapter 4. The methodology was tested for images of both manned and unmanned aerial platforms from several geographic locations that are highly varying in image and scene characteristics. The methods based on CNN features produced an average accuracy of approximately 93% for all datasets. Also it produced an average accuracy of 85% in the model transferability scenarios, i.e., the model was trained with one dataset and tested with another dataset. It was anticipated that inclusion of 3D point cloud features along with 2D CNN features would improve the accuracy. Pertaining to this, a framework based on multiple-kernel-learning (MKL) was proposed to integrate CNN features of images and 3D point cloud features to perform damage detection. The feature extraction was carried out at segment level, for which histogram-based 3D point cloud features were proposed. The proposed 3D features were examined independently and found to be useful in detecting the damaged regions. In conclusion, the integration of proposed 3D point cloud features and image-based CNN features based on MKL framework was found to improve the classification accuracy, particularly in the model transferability scenario where the achieved maximum accuracy improvement was around 7%. From the results, it was inferred that the CNN based model possesses a strong generalization capability. Hence, it was strongly anticipated that this model can be used as a tool for automatically producing damage maps from very high resolution images when a new event occurs, helping to circumvent the tedious manual mapping. This capability was demonstrated by Duarte et al. (2017) as they adopted the model developed in this research (without any retraining) for mapping the damaged façades from oblique airborne images and reported an overall accuracy of 83%. However, intuitively it is expected that this accuracy can be increased further by improving generalization capability of the CNN model by training it with more number of samples from distributed locations with significantly varying characteristics.

All the aforementioned methods for damage detections were based on mono-temporal post-event data. In these approaches, the damage detection methods were designed based on the primary fact that damaged regions will possess non-uniform radiometric and geometric characteristics compared to undamaged man-made objects. However, these assumptions were often found to be failing in complex urban areas, and thereby hindering the accurate mapping of damaged regions (for e.g., see L'Aquila dataset described in Chapter 5). In such cases, the presence of pre-event data could be useful. Towards this, a methodology was developed and reported in Chapter 5 to identify the damaged regions by comparing the pre- and post-event 3D point clouds and images. The proposed method detected almost all damages related to geometric deformation of the building's elements in the considered study area. The developed method was fully unsupervised in contrast to the methods reported in earlier chapters, which were largely based on supervised approach. The unsupervised methods are simpler and independent (no need of training samples) compared to supervised approaches. The major advantage of this unsupervised based multi-temporal approach is that the estimates will be more robust and reliable compared to the estimates based on mono-temporal post-event data based approaches reported in earlier Chapters.

From the experiments conducted in earlier chapters of this thesis (particularly in Chapter 4), the usefulness of site specific samples for training the supervised classifier for accurate damage detection was established. Moreover, it has become a common practice to make the damage assessment of individual buildings available from various sources (cf. Chapter 6). In order to utilize these site specific samples that are streaming online at different points in time for training the classifier, a framework for damage detection based on an incrementally learning classifier was developed and reported in Chapter 6. The suitability and potential of the proposed incremental classifier based framework for the application of damage detection was demonstrated. A noteworthy inference from the experiments related to the proposed framework is that compared to handcrafted features the CNN features are more effective and suitable for online classification, as they yield similar accuracy when used in both online and batch learning setting, whereas, handcrafted features produced significantly inferior results in online setting compared to batch learning approaches. In conclusion, CNN features were recommended for building the online classifier for the disaster damage detection application.

All methods reported in Chapter 2 to 6 were largely focussed on mapping the damages related to debris/rubble piles. However, as mentioned earlier, to perform a comprehensive damage assessment, a 3D model of the building annotated with damage evidences is required. Accurate roof delineation of the building is the minimum requirement for constructing at least a rudimentary 3D model. However, accurate delineation of independent elements of the building from photogrammetric 3D point cloud was challenging, as inferred from the experiments conducted in Chapter 2 for building delineation. It was anticipated that integrated use of 2D radiometric features from images and 3D geometric information from point cloud would help to obtain accurate roof faces of the building. To this end, a segmentation algorithm was proposed in Chapter 7 which incorporates both 3D point cloud features and 2D radiometric features from images to perform image segmentation to accurately delineate the individual elements of the building. The segmentation was carried out in image space by assigning the 3D point cloud features to the corresponding image pixels. The proposed segmentation algorithm worked well in most of the cases, where it even accurately delineated small objects such as windows on the roofs and façades. However, it failed in some image regions, particularly, it led to under-segmentation where the 3D point cloud feature information was missing or image regions corresponding to highly noisy 3D point cloud. In order to overcome these issues, another framework was proposed in Chapter 8 to automatically delineate the roof segments of the building, by synergistically using 3D point cloud and image features. The proposed method incorporates the 3D features in super-pixel level (derived based on image-based over-segmentation approach) instead of pixel level. This approach largely alleviates the under-segmentation issues and delivered accurate roof faces of the building essentially sufficient for 3D reconstruction. Also, using an automated approach, the images corresponding to each element of the building were mapped in order to facilitate further processing to produce the information in a format required by a specific stakeholder e.g., deriving building-level damage label or score.

The methods which were demonstrated to be effective based on real word datasets for damage detection, building delineation and 3D construction in Chapters 2 to 8 of this thesis, were selected. The selected methods were integrated as an automated system for damage assessment. The system was examined based on the UAV datasets of the RECONASS building which was damaged deliberately by two consecutive bomb blasts in the pilot experiments conducted in Sweden. The integrated system was found to be working well where all methods in the system were consecutively executed which produced results similar to the one obtained from the experiments conducted based on several real world datasets reported in chapters 2 to 8. As the developed modules were found to be working well when they were transferred to the new unseen datasets from the pilot experiments of RECONASS, we anticipate that these methods can be scalable as independent operational systems to assess the damages to the real word functioning buildings after any disaster event. Moreover, the methods developed in this research were not confined to damage detection application. For example, the methods developed as part of this research such as image segmentation using 2D and 3D features, the online learning framework for classification, change detection between point clouds, accurate roof segment delineation for 3D reconstruction and the 3D point cloud features proposed in Chapter 5 are generic enough to be considered in several other remote sensing applications such as object detection, 3D city modelling, cadastral mapping, and urban sprawl monitoring.

The methods developed in this research were effective and sufficient to generate quick automatic damage maps containing the information of collapsed or heavily damaged buildings based on remote sensing images. This would directly satisfy the requirements of the first responders involved in speedy response activities. However, this would be insufficient for some other actors such as insurance companies. For example, as mentioned earlier, they need very detailed and high level information in the format like: *window in first story is broken, staircase is collapsed, wide diagonal crack on the wall, roof top is inclined, chimney is broken,* etc. Conducting research to design a methodology for automatically generating these kinds of high level information by utilizing (as well improving, if required) the methods reported in this research such as mapping of damage evidences and 3D modelling of the building from remote sensing data would be the next stepping stone towards an automatic comprehensive damage assessment system.

References of Chapter 9 &10

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing,* 84, pp. 85-99.

Duarte, D., Nex, F., Kerle, N., Vosselman, G., 2017. Towards a more efficient detection of earthquake induced façade damages using oblique uav imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLII-2/W6, pp. 93-100.

Ma, J., Qin, S., 2012. Automatic depicting algorithm of earthquake collapsed buildings with airborne high resolution image, *IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), 2012* pp. 939-942.

Radhika, S., Tamura, Y., Matsui, M., 2012. Use of post-storm images for automated tornado-borne debris path identification using texture-wavelet analysis. *Journal of Wind Engineering and Industrial Aerodynamics,* 107, pp. 202-213.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J.D., Breitkopf, U., Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing,* 93, pp. 256-271.

Ruiz, L., Fdez-Sarría, A., Recio, J., 2004. Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study, *20th ISPRS Congress*, pp. 1109-1114.

Stavrakoudis, D., Theocharis, J., Zalidis, G., 2011. A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66 (4), pp. 529-544.

Sun, S., Salvaggio, C., 2013. Aerial 3D building detection and modeling from airborne LiDAR point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6 (3), pp. 1440-1449.

Tu, J., Li, D., Feng, W., Han, Q., Sui, H., 2017. Detecting Damaged Building Regions Based on Semantic Scene Change from Multi-Temporal High-Resolution Remote Sensing Images. *ISPRS International Journal of Geo-Information,* 6 (5), pp. 131.

Tu, J., Sui, H., Feng, W., Sun, K., Hua, L., 2016. Detection of damaged rooftop areas from high-resolution aerial images based on visual bag-of-words model. *IEEE Geoscience and Remote Sensing Letters,* 13 (12), pp. 1817-1821.

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, pp. 61-78.

Biography

Anand Vetrivel was born on 18 September, 1986 in Chennai, India. He received B.Tech in Information Technology from Thiagarajar College of Engineering, India, in 2008 and MSc in Environmental Remote Sensing from University of Dundee, Scotland, in 2010. He joined at ITC in 2013 as a PhD funded through EU-FP7 project RECONASS. Before joining ITC, he worked with various organizations:

07/2012 – 12/2012	**Research Associate**, Global Technology Office – Image Processing R&D, Cognizant Technology Solutions, Chennai, India.
11/2011 - 07/2012	**Algorithm Engineer** in Computer Vision/Image Processing domain, Advanced Engineering R&D, SMR Automotive Systems India Limited, Bangalore, India.
11/2010 - 08/2011	**Senior Research Fellow**, Environmental Research Lab, Thiagarajar College of Engineering, Madurai, India.
8/2008 - 02/2009	**Assistant System Engineer** (Software R&D), Tata Research Development and Design Center (TRDDC), Pune, India.

His research focuses on developing automatic methods to extract information from very high resolution images and 3D point cloud that could be used for various applications. His principle area of interests are image processing, computer vision, photogrammetry and machine learning.

Author’s publications

ISI Journal Paper:

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

Vetrivel, A., Gerke, M., Kerle, N. and Vosselman, G. (2016) Identification of structurally damaged areas in airborne oblique images using a visual-bag-of-words approach. In: Remote Sensing : open access, 8 (2016)3article no. 231, 22 p.

Vetrivel, A., Gerke, M., Kerle, N., Nex, F.C. and Vosselman, G. (2017) Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. In: ISPRS Journal of Photogrammetry and Remote Sensing, (2017) IN PRESS15 p.

International Conference Paper:

Vetrivel, A., Gerke, M., Kerle, N. and Vosselman, G. (2015) Segmentation of UAV - based images incorporating 3D point cloud information. In: PIA15+HRIGI15 : Joint ISPRS conference proceedings, 25-27 March 2015, Munich, Germany. ISPRS Archives, Vol XL-3/W2. : open access. ISPRS, 2015. pp. 261-268.

Vetrivel, A., Duarte, D., Nex, F.C., Gerke, M., Kerle, N. and Vosselman, G. (2016) Potential of multi-temporal oblique airborne imagery for structural damage assessment. In: Proceedings of the XXIII ISPRS Congress: From human history to the future with spatial information, 12-19 July 2016, Prague, Czech Republic. Peer reviewed Annals, Volume III-3, 2016 / edited by L. Halounova, ... [et al.]. ISPRS, 2016. ISSN: 2194-9050. pp. 355-362.

Vetrivel, A., Kerle, N., Gerke, M., Nex, F.C. and Vosselman, G. (2016) Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning. In: Proceedings of GEOBIA 2016 : Solutions and synergies, 14-16 September 2016, Enschede, Netherlands / edited by N. Kerle, M. Gerke and S. Lefevre. Enschede: University of Twente Faculty of Geo-Information and Earth Observation (ITC), 2016. ISBN: 978-90-365-4201-2. 5 p

ITC Dissertation List

https://www.itc.nl/Pub/research_programme/Research-review-and-output/PhD-Graduates