



# Usability of aerial video footage for 3-D scene reconstruction and structural damage assessment

Johnny Cusicanqui, Norman Kerle, and Francesco Nex

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

**Correspondence:** Johnny Cusicanqui (jfcusicanqui@gmail.com)

Received: 21 November 2017 – Discussion started: 8 January 2018

Revised: 17 April 2018 – Accepted: 8 May 2018 – Published: 8 June 2018

**Abstract.** Remote sensing has evolved into the most efficient approach to assess post-disaster structural damage, in extensively affected areas through the use of spaceborne data. For smaller, and in particular, complex urban disaster scenes, multi-perspective aerial imagery obtained with unmanned aerial vehicles and derived dense color 3-D models are increasingly being used. These type of data allow the direct and automated recognition of damage-related features, supporting an effective post-disaster structural damage assessment. However, the rapid collection and sharing of multi-perspective aerial imagery is still limited due to tight or lacking regulations and legal frameworks. A potential alternative is aerial video footage, which is typically acquired and shared by civil protection institutions or news media and which tends to be the first type of airborne data available. Nevertheless, inherent artifacts and the lack of suitable processing means have long limited its potential use in structural damage assessment and other post-disaster activities. In this research the usability of modern aerial video data was evaluated based on a comparative quality and application analysis of video data and multi-perspective imagery (photos), and their derivative 3-D point clouds created using current photogrammetric techniques. Additionally, the effects of external factors, such as topography and the presence of smoke and moving objects, were determined by analyzing two different earthquake-affected sites: Tainan (Taiwan) and Pescara del Tronto (Italy). Results demonstrated similar usabilities for video and photos. This is shown by the short 2 cm of difference between the accuracies of video- and photo-based 3-D point clouds. Despite the low video resolution, the usability of these data was compensated for by a small ground sampling distance. Instead of video characteristics, low quality and application resulted from non-data-related

factors, such as changes in the scene, lack of texture, or moving objects. We conclude that not only are current video data more rapidly available than photos, but they also have a comparable ability to assist in image-based structural damage assessment and other post-disaster activities.

## 1 Introduction

The effectiveness of post-disaster activities during the response and recovery phases relies on accurate and early damage estimations. Post-disaster structural damage assessment (SDA) is typically based on ground surveying methods, the most accurate approach for the assessment and classification of structural building damage. Main response activities, however, rely on rapidly obtaining building damage information, which is not possible through such ground-based methods. Remote-sensing techniques are an important tool for rapid and reliable SDA. Different approaches have been extensively studied; a structured review of all these approaches was presented by Dong and Shan (2013). Traditionally, data obtained from nadir perspective (i.e., vertical) platforms, such as satellites and airplanes, were used. However, the lack of an oblique perspective of this kind of data hinders the identification of damage-related features at building façades, which are critical for a comprehensive damage assessment (Gerke and Kerle, 2011; Vetrivel et al., 2015). The potential for SDA of oblique-perspective data, such as Pictometry and aerial stereoscopic imagery, has long been realized; however, means to process those data have traditionally been limited to manual/visual inspection or to single-perspective texture-based damage mapping (Mitomi et al., 2002; Ogawa and Yamazaki, 2000; Ozisik and Kerle, 2004;

Saito et al., 2010). A very promising source of oblique, and even multi-perspective, post-disaster aerial data that has matured in recent years is unmanned aerial vehicles (UAVs). In comparison to traditional platforms, UAVs are more versatile in capturing multi-perspective high-resolution imagery, are easy to transport and fly, allow easier access to destroyed areas than ground surveys, and provide an economically more attractive solution than traditional airborne approaches (Nex and Remondino, 2014). Additionally, modern photogrammetry and computer vision techniques for image calibration, orientation, and matching allow rapid processing of multi-perspective data for the generation of dense color 3-D point clouds (3DPCs) (Vetrivel et al., 2017). State-of-the-art methods based on semantic reasoning and deep learning, have been developed for highly accurate and automated post-disaster SDAs through the use of multi-perspective imagery and 3DPCs. However, the acquisition of multi-perspective imagery using UAV still cameras, labeled in this research as photos, requires a structured plan during crisis situations, one that is integrated with all humanitarian assistance regulations and that is often hindered by tight or lacking regulations and legal frameworks. To date, this poses the most severe limitation for the effective collection and sharing of this kind of data. Alternatively, aerial video footage typically obtained by civil defense forces, police, fire services, or news media is less subjected to the legal regulations. Those data optimize the endurance flight time capacity and can even be streamed straight to the operator to be available in near-real time, making them a potentially powerful information source for damage mapping (FSD, 2016; Xu et al., 2016). Video data are in essence a combination of still frames that can be easily extracted and used as aerial images for post-disaster SDA. However they differ from the ones obtained from either compact or digital single-lens reflex (SLR) cameras due to inherent artifacts (e.g., motion-blur effects) and generally lower-quality characteristics (e.g., resolution and frame redundancy) (Alsadik et al., 2015; Kerle and Stekelenburg, 2004). The effects of such video artifacts and lower-quality characteristics, and the true potential of this kind of data for SDA and other post-disaster activities, remain poorly understood. This research aims at determining the usability of aerial video data, extracted as frames, for an effective SDA, in comparison to data obtained by UAV-based still cameras, here referred to as photographs or photos. Two study cases were selected: the 2016 Taiwan earthquake, which caused severe but local damage in the city of Tainan, and the 2016 Pescara del Tronto earthquake in Italy, which instead caused more extensive and diverse (low to complete) damage.

### 1.1 Image- and video-based structural damage assessment

Dong and Shan (2013) divided image-based SDA approaches into mono- and multi-temporal. The largest group of existing methods corresponds to multi-temporal analysis of

data, where change detection is applied to a variety of pre- and post-disaster remote-sensing data. However, while for satellite data the pre-event archive has been growing, frequently allowing change detection, for airborne data, including very detailed imagery obtained from UAVs, such reference data are usually lacking. Therefore, approaches linked to multi-perspective data are based on pre-disaster assumptions that produce uncertainty (Vetrivel et al., 2016a). Two-dimensional (i.e., only using multi-perspective imagery) and 3-D (i.e., also with UAV-based 3DPCs) UAV data have been used for mono-temporal image-based SDA. Two-dimensional image features have been studied for the recognition of damage patterns, among which texture features such as histogram of oriented gradients (HoG) and Gabor were found to be the most effective (Samadzadegan and Rastiveisi, 2008; Tu et al., 2016; Vetrivel et al., 2015). Vetrivel et al. (2015) extracted Gabor wavelets and HoG texture descriptor features from UAV imagery for a supervised learning classification of damage-related structural gaps, by considering the distinctive damage textural pattern at these types of gaps' surroundings. This approach led to high accuracies, but the generalization was limited due to the sensitivity of Gabor wavelets and HoG to scale changes and clutter. An improvement on generalization was achieved later by Vetrivel et al. (2016b), who used the same texture features in a more robust framework for feature representation, visual bag of words (BoW). Pattern recognition methods based on texture features are highly capable of discriminating damaged regions; however, they rely on site-specific damage patterns that still limit an appropriate transferability. Domain-specific semantic analysis is another 2-D approach to classify structural damage. This approach refers to the use of contextual association of all retrievable features (e.g., textural, spectral, geometric) for the development of ontological classification schemes (e.g., set of rules), defined based on domain knowledge. Fernandez Galarreta et al. (2015) developed a methodology for the extraction and classification of building segments based on semantic reasoning using UAV imagery. However, the complex and subjective aggregation of segment damage-related information to building level was a main drawback to this approach. Recently, deep learning methods such as convolutional neural networks (CNNs) led to a major improvement in SDA accuracies. Vetrivel et al. (2017) tested different CNN architectures for different sites and data, and obtained highly accurate damage estimations using a pre-trained model tuned with damage and no-damage training samples from a range of disaster sites. Furthermore, 3-D data such as UAV-generated 3DPCs were used for damage estimations, based on the extraction of geometric features. Only a few mono-temporal SDA methods using 3-D data have been developed, typically based on point neighborhoods and segment-level extracted geometric features. Khoshelham et al. (2013) tested segment-based features such as point spacing, planarity, and sphericity for an accurate classification of damage building roofs. A more se-

mantic approach was developed by Weinmann et al. (2015), who extracted different geometric features from every 3-D point neighborhood and used them for the posterior classification of all 3-D points. These methods, however, did not represent the irregular pattern of damaged areas well; hence Vetrivel et al. (2017) instead used a segment-level histogram approach based on HoG, for a better representation of these patterns in every 3-D point neighborhood.

Video data have been tested for image-based 3-D scene reconstruction in different applications. For example, Singh et al. (2014) used video sequences for 3-D city modeling, and later Alsadik et al. (2015) and Xu et al. (2016) treated video data for the detection of cultural heritage objects. In general, processing pipelines were focused on an effective selection of good-quality and non-redundant video frames and their automated and accurate geo-registration (Ahmed et al., 2010; Clift and Clark, 2016; Eugster and Nebiker, 2007). These advances allow fast and reliable processing of video data in near-real time; however, less research has been done on their potential on post-disaster SDA. Mitomi et al. (2002) analyzed distinct spectral damage features in individual video frames, obtaining a reasonable building damage classifications using a maximum-likelihood classifier whose efficiency, however, was limited by the need for continuous selection of training samples. Later Ozisik and Kerle (2004) demonstrated that video data are able to improve structural damage estimations once integrated with moderate-resolution satellite imagery. However, such data require substantial work to be processed, registered, and integrated. A similar but more elaborated approach was presented by Kerle and Stekelenburg (2004), who created a 3-D environment from oriented and improved video frames. In this research, the automated classification results were poor, highlighting the quality-related limitations of video data common at that time. Video-based SDA approaches show that aerial video data require substantial work to be incorporated into conventional damage classification approaches. However, they can be a potential source of damage-related information due to their multi-perspective nature, similar to UAV images, and real-time availability. Additionally, modern cameras capture video data with an improved resolution (up to 4 K), and software resources allow easier video data processing for 3-D model generation.

## 2 Methodology

### 2.1 Study cases and data description

Two earthquake-affected urban areas were selected for this research due to their contrasting characteristics, Tainan (Taiwan) and Pescara del Tronto (Italy; Fig. 1). Both cities were affected by earthquakes of similar magnitude: 6.4 and 6.2 in February and August 2016, respectively. Tainan is a large city with regular topography where most buildings are composed

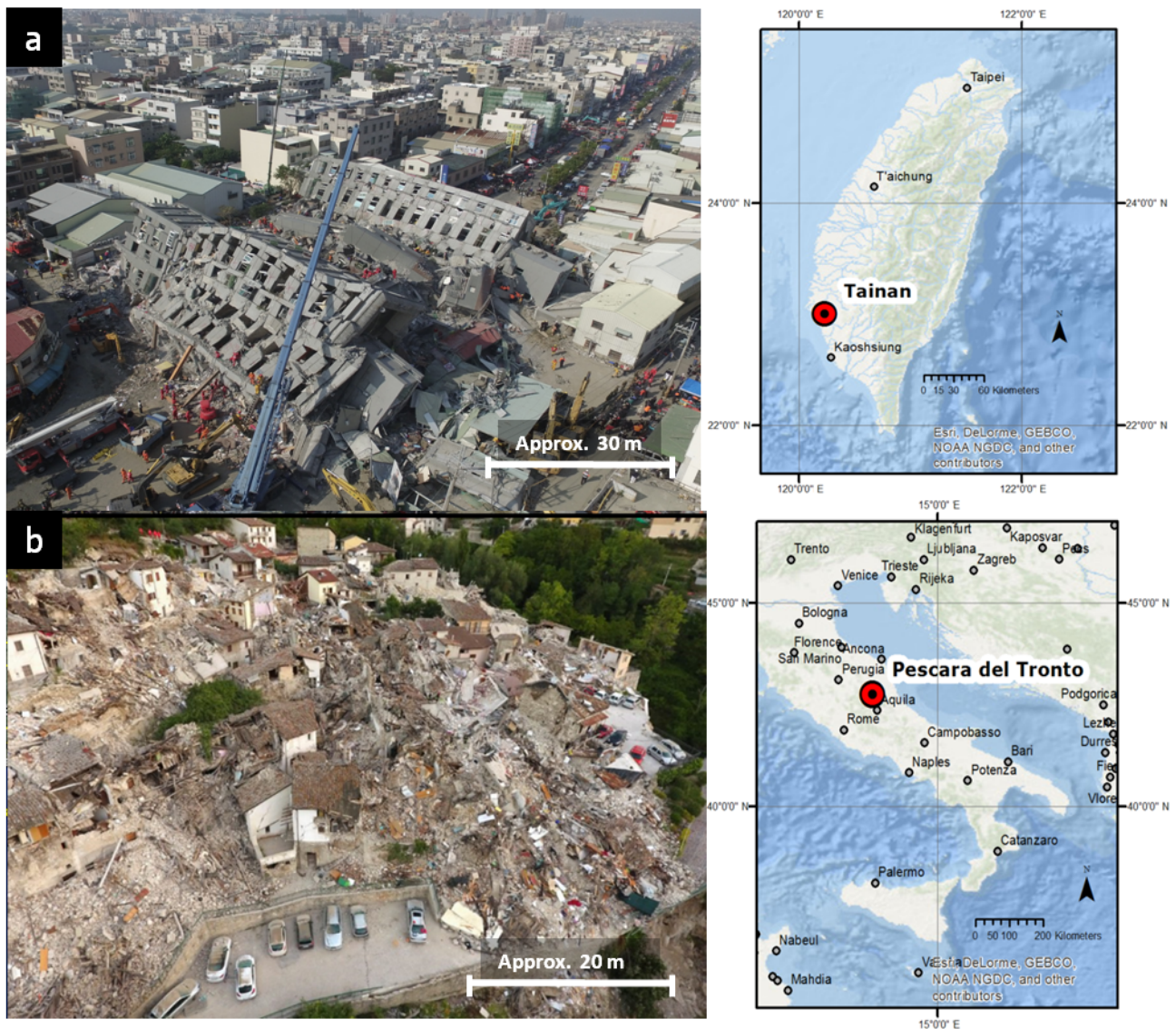
of reinforced concrete. Conversely, Pescara del Tronto is a rural city settled on one hillside of the Tronto Valley; hence topography there is hilly, and houses are mainly made of stone. These disparate characteristics give each case study different damage features: while in Tainan damage was largely confined to a single collapsed high-rise building, in Pescara del Tronto it was extensive and highly variable (i.e., intermediate damage levels can also be found). The aim of selecting these contrasting cases is to obtain a broader understanding of video data usability in SDA.

After these earthquake events UAV photos and videos were acquired. UAV-based oblique and nadir photos of six dates after the disaster were collected in Tainan for most of the damaged zone, while a more limited amount of nadir photos was obtained in Pescara del Tronto (Table 1). Photo resolution in both cases is 12 MP. The Pescara del Tronto photos exhibit some blur effects due to a low shutter speed (i.e., 1 / 60 s). Nadir photos in both study areas were obtained from larger flying heights; hence ground sampling distance (GSD) is also larger for this data type than for the oblique photos. Likewise, several aerial videos were obtained at each study site and uploaded to video-sharing websites such as Youtube, from where the most complete sequences of highest quality were downloaded. The resolution for both areas was Full HD (i.e.,  $1920 \times 1080$  pixels  $\approx 2.1$  MP). The video streams downloaded present different perspectives and characteristics. Due to the flat topography, the Tainan footage presents a more complete overview of the scene than Pescara del Tronto. However, the footage is frequently affected by extensive smoke emanating from the collapsed building and was obtained from a higher elevation (also resulting in higher GSD) than the one of Pescara del Tronto.

Additionally, for each study area precise 3-D ground control points (GCPs) were obtained. In both cases the GCPs are well distributed over the terrain; however, in Tainan GCPs were only measured using street markings, resulting in 3-D position data with very limited  $z$  variation.

### 2.2 Data preparation

Several datasets were generated from the acquired video streams and photos in relation to the following preparation parameters to be investigated: (i) data type, (ii) frame selection approach, (iii) sensor perspective, and (iv) resolution. (i) First datasets were divided according to their nature into video footage and photos for every study case. (ii) Then video frames were extracted by two approaches: random frame selection (RFS) and wise frame selection (WFS). RFS is only defined by an empirical number of randomly selected frames (defined in Pix4D, 2017), while WFS refers to a more elaborated approach that reduces blur-motion effects and frame redundancy. The latter uses as a guideline each frame's initial 3-D position (generated with Pix4D) and corresponding image quality index (IQI) value (computed with 3-D Zephyr software 3DFlow (2017)), with the aim of



**Figure 1.** Study cases: (a) Taiwan earthquake, February 2016: flat terrain, smoke presence, and highly localized damage (Weiguan Jinglong collapsed building in the middle); (b) Pescara del Tronto earthquake, August 2016: hilly terrain and more diverse, widespread damage.

discarding the most redundant (i.e., more than 80 % overlap) and lowest-quality (i.e., an IQI lower than 0.5) frames. (iii) Additionally, the Tainan photos were divided according to their perspective into oblique and nadir. Video frames by their nature can be considered oblique. (iv) Finally, extracted video frames and photos resolutions were modified to generate additional datasets; four levels were analyzed: original, medium, low, and coarse (i.e., very low) resolution. Twenty-six datasets were prepared in total, each related to a different preparation parameter to be investigated, as summarized in Table 2. Of those, two video and two photo datasets were labeled as default, since they were prepared with default parameters (i.e., random frame selection in the case of videos, oblique and nadir perspective in the case of photos, origi-

nal resolution and without refinement). Then, 3DPCs were generated from the generated datasets with Pix4d. One reference 3DPC was prepared using a refinement procedure which consisted of the use of manual tie points (MTPs) in Pix4d. Finally, the GCPs were used for the geo-registration of most photo-based 3DPCs, whereas the video and Tainan oblique-photo-based 3DPCs were registered using corresponding extracted 3-D points from the reference 3DPC.

### 2.3 Usability of video data for 3-D structural damage assessment

Once all the datasets and 3DPCs were prepared, as described in Sect. 2.2, their usability was analyzed in three sections:



**Table 1.** Data obtained at the study sites. Highlighted are some data proprieties that are especially relevant for this research.

Study Case	Data	Perspective properties	Image altitude (m)	Approximate (cm)	GSD date	Acquisition
Tainan	Photos	Nadir	4000 × 3000 p (12 MP)	87	11.64	07 and 10–14/02/16
		Oblique		42	<b>3.56</b>	
	Video	Oblique	1920 × 1080 p (2.1 MP)	50	7.66	06/02/16
	GCPs	25 GCPs acquired from street markings in Tainan (limited z variation)				
Pescara del Tronto	Photos	Nadir	4000 × 3000 p (12 MP)	166	5.91	Between 05 and
			<b>Shutter speed = 1 / 60 s</b>			09/09/16
	Video	Oblique	1920 × 1080 p (2.1 MP)	18	<b>3.8</b>	06/08/16
	GCPs	25 GCPs acquired from well-distributed marks, such as building corners in Pescara del Tronto.				

**Table 2.** Prepared datasets from video and photo data.

No.	Study site	Data type	Frame selection approach	Sensor perspective	Resolution*	Refinement	Date	Geo-registration
1	Tainan	Photo	n/a	Oblique and nadir	Original	No	07/02/16	GCPs
2	Tainan	Photo	n/a	Oblique and nadir	Original	Yes	07/02/16	GCPs
3	Tainan	Photo	n/a	Oblique and nadir	Medium	No	07/02/16	GCPs
4	Tainan	Photo	n/a	Oblique and nadir	Low	No	07/02/16	GCPs
5	Tainan	Photo	n/a	Oblique and nadir	Coarse	No	07/02/16	GCPs
6	Tainan	Photo	n/a	Oblique	Original	No	07/02/16	GCPs and reference 3DPC
7	Tainan	Photo	n/a	Nadir	Original	No	07/02/16	GCPs
8	Tainan	Photo	n/a	Oblique and nadir	Original	No	10/02/16	GCPs
9	Tainan	Photo	n/a	Oblique and nadir	Original	No	11/02/16	GCPs
10	Tainan	Photo	n/a	Oblique and nadir	Original	No	12/02/16	GCPs
11	Tainan	Photo	n/a	Oblique and nadir	Original	No	13/02/16	GCPs
12	Tainan	Photo	n/a	Oblique and nadir	Original	No	14/02/16	GCPs
13	Tainan	Video frame	RFS	Oblique	Original	No	07/02/16	Reference 3DPC
14	Tainan	Video frame	RFS	Oblique	Medium	No	07/02/16	Reference 3DPC
15	Tainan	Video frame	RFS	Oblique	Low	No	07/02/16	Reference 3DPC
16	Tainan	Video frame	RFS	Oblique	Coarse	No	07/02/16	Reference 3DPC
17	Tainan	Video frame	WFS	Oblique	Original	No	07/02/16	Reference 3DPC
18	Pescara del Tronto	Photo	n/a	Nadir	Original	No	05–09/09/16	GCPs
19	Pescara del Tronto	Photo	n/a	Nadir	Medium	No	05–09/09/16	GCPs
20	Pescara del Tronto	Photo	n/a	Nadir	Low	No	05–09/09/16	GCPs
21	Pescara del Tronto	Photo	n/a	Nadir	Coarse	No	05–09/09/16	GCPs
22	Pescara del Tronto	Video frame	RFS	Oblique	Original	No	26/08/16	Reference 3DPC
23	Pescara del Tronto	Video frame	RFS	Oblique	Medium	No	26/08/16	Reference 3DPC
24	Pescara del Tronto	Video frame	RFS	Oblique	Low	No	26/08/16	Reference 3DPC
25	Pescara del Tronto	Video frame	RFS	Oblique	Coarse	No	26/08/16	Reference 3DPC
26	Pescara del Tronto	Video frame	WFS	Oblique	Original	No	26/08/16	Reference 3DPC

Dataset 1 and 18: default photo datasets. Dataset 1 was also used for the multi-temporal application analysis together with datasets 8 to 12.

Dataset 13 and 22: default video datasets, also used for the analysis of the frame selection approach, RFS in comparison to WFS (datasets 17 and 26).

Dataset 2: dataset used as reference 3DPC. This is the only dataset that followed a refinement process with manual tie points.

\* Original resolution: 4000 × 3000 p or 12 MP (Photos) and 1920 × 1080 p or 2.1 MP (video frames); medium resolution: 1/2 (half image size);

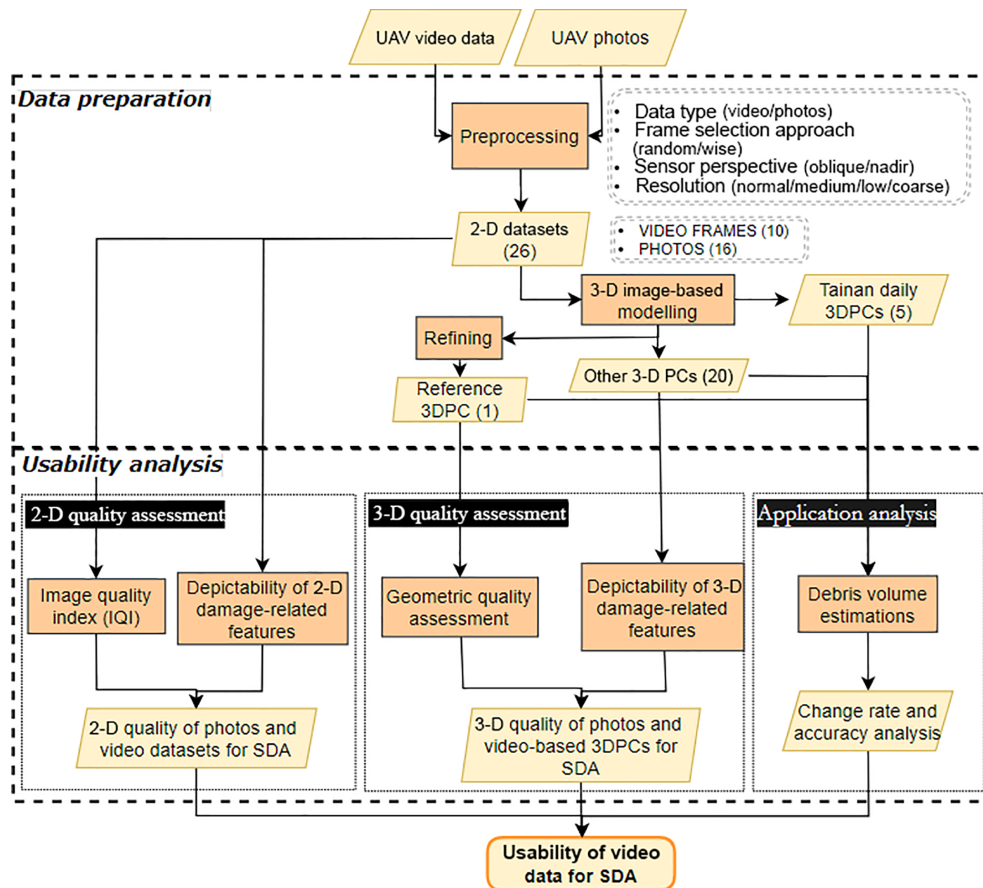
low resolution: 1/4 (quarter image size); and coarse resolution: 1/8 (one-eighth image size).

**2-D quality assessment** consisted of a direct measurement of the image quality of every dataset with the use of IQI, complemented by the assessment of their depictability of 2-D damage-related features for 2-D-based SDA.

**3-D quality assessment** consisted of the determination of the absolute geometric quality of the generated 3DPCs expressed by their internal and external accuracy, com-

plemented by their depictability of 3-D damage-related features for 3-D-based SDA.

**Application analysis** was based on the estimation of the debris volumes with the generated 3DPCs, determined by the analysis of the debris change trend and accuracy assessment.



**Figure 2.** Workflow to determine usability of video data in 2-D- and 3-D-based SDA. Dashed rectangles: main phases.

### 2.3.1 2-D quality assessment

#### Direct image quality assessment with IQI

The prepared datasets were first analyzed with their respective IQIs for their direct quality assessment. IQI series and image reference quality maps indicate possible sources of error for image-based 3-D scene reconstruction such as the presence of low-texture areas or motion-blur effects (3DFlow2017, 2017).

#### Depictability of 2-D damage-related features

This assessment was based on the classification accuracy of damaged and non-damaged segments or super-pixels extracted from the prepared datasets. A deep learning approach presented by Vetrivel et al. (2017) was tested. This is based on the “imagenet-caffe-alex” CNN model (Krizhevsky et al., 2012), which is composed of different groups of layers. Convolutional layers represent the first group and are a set of filter banks composed of image and contextual feature filters. The following group corresponds to data shrinking and normalization layers. Finally, the last group transforms all the information generated and output features with high-level

reasoning; usually this layer is connected to a loss function for the final classification. This approach uses a large amount of training samples (i.e., labeled images) to tune the weights of the CNN classification layer.

Three representative video frames and photos of the same area were chosen from both data types and study areas based on the damage presence: (i) severe/complete damage, (ii) partial damage, and (iii) no damage. Those datasets were first segmented into super-pixels using the simple linear iterative clustering (SLIC) algorithm (Achanta et al., 2010). Empirical values were tested according to the data scale with the aim of over-segmenting the scene. The extracted super-pixels were then classified into damage and non-damage categories using the CNN model. Damage classification accuracies for every dataset were determined using reference classifications created by visual inspection of the same segmented datasets. Finally, the depictability of damage-related features for the analyzed datasets was determined by comparing their damage classification accuracies.

### 2.3.2 3-D quality assessment

#### Geometric quality assessment

Point cloud geometric quality was determined by a 3DPC internal and external accuracy assessment. Internal accuracy is related to the model precision, and different indicators can measure this 3DPC attribute. Based on previous methodologies (Oude Elberink and Vosselman, 2011; Jarzabek-Rychard and Karpina, 2016; Soudarissanane et al., 2008), in this research planar fitting and completeness of the generated 3DPCs were measured. For the former a planar object (e.g., roofs) was identified and extracted as 3DPC segments from every 3DPC and study area using CloudCompare (CloudCompare, 2017). These segments were used to define a plane by a least squares fitting algorithm. The mean distance of every 3-D point within the segment to its fitted plane was measured to determine an absolute deviation for every dataset. To mitigate the effect of flying height, a relative deviation was measured by computing the ratio of the obtained mean distances and the dataset respective GSD (Table 1). Completeness, in turn, was measured from the projection of every 3DPC to two raster maps, each associated with a different perspective (i.e., vertical and horizontal) (Girardeau-Montaut, 2017). Raster cells in these maps containing the number of 3-D points were used to compute mean density and the percentage of empty cells for every 3DPC.

External accuracy instead refers to how approximate a 3DPC is to reality, which requires a reference model (Al-sadik et al., 2015; Kersten and Lindstaedt, 2012). The reference 3DPC (Table 2: dataset 2; see Sect. 2.2) was used and compared to every 3DPC. Thereby, the mean distance of the 3-D points to the reference 3-D points was computed to determine every 3DPC external accuracy.

#### Depictability of 3-D damage-related features

This assessment was based on previous investigations on the use of 3DPCs for SDA. Visual inspection was applied to 3DPCs for the identification of severely damaged buildings in Fernandez Galarreta et al. (2015). In a similar manner, here different damage-related features were identified using a mesh model generated from the studied 3DPCs. Thus, the feasibility on this visual identification was compared and interpreted. As a complement, a more elaborated approach using 3-D damage-related features based on Vetrivel et al. (2017) was applied. This consisted of a segment-level representation of different 3-D features. Three 3-D features were used: mean curvature, normal change rate, and roughness. These point-level features were estimated using a kernel (neighborhood size) of 1 m for all 3DPCs. A damaged area in Tainan was delineated to retrieve the segment-level distributions of these 3-D features for every 3DPC, including the reference 3DPC (Table 2: dataset 2). Through the use of box plots, the segment-level distributions derived from

every photo- and video-based 3DPC were compared to the ones retrieved from the reference model to determine their depictability of 3-D damage-related features. Additionally, non-damage segment distributions were retrieved from the reference 3DPC to confirm the relation of the chosen 3-D features with damage (Fig. 3).

### 2.3.3 Application analysis

UAV photos of six different dates after the Tainan earthquake on 6 February 2016 were used in this analysis: 7, 10, 11, 12, 13, and 14 February 2016. The 3DPC for the first date was generated with the Tainan photos with default preparation parameters (Table 2: dataset 1), as were the 3DPCs for the subsequent dates (see Sect. 2.2; Table 2: datasets 8 to 12). From the 3DPC of the first date the damaged area was delimited and used as a base value for the volume estimations. The volume of debris presented within this area was calculated in cubic meters and analyzed for every date using Pix4d. The same volume estimation approach was applied to the 3DPC generated with the video default dataset (Table 2: dataset 13) and the reference 3DPC (Table 2: dataset 2). Thereby, volume estimations with photo-based 3DPC were analyzed and later compared to the video-based and reference estimations to determine their similarity and accuracy.

## 3 Results

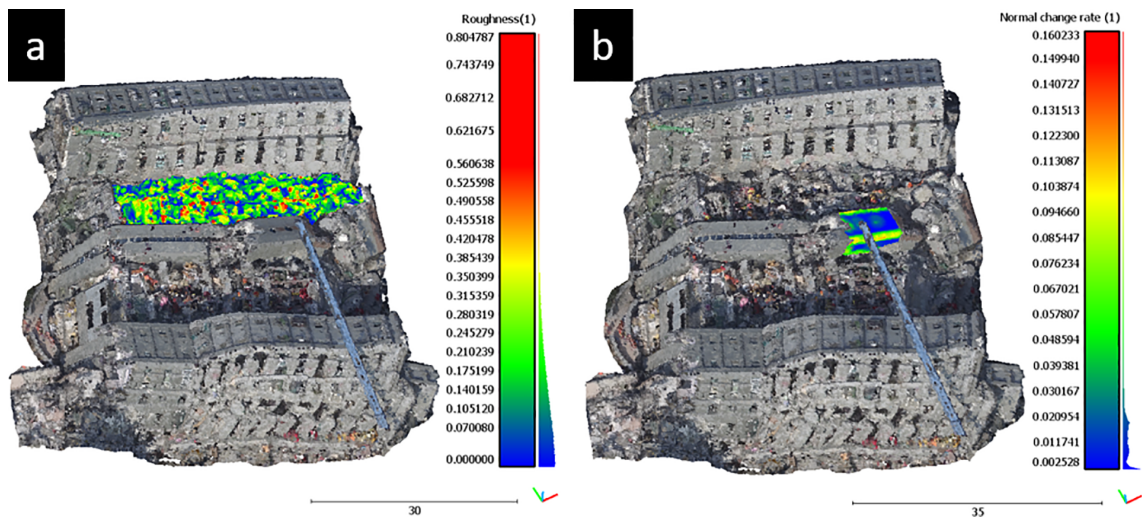
### 3.1 2-D quality assessment

#### 3.1.1 Direct image quality assessment with IQI

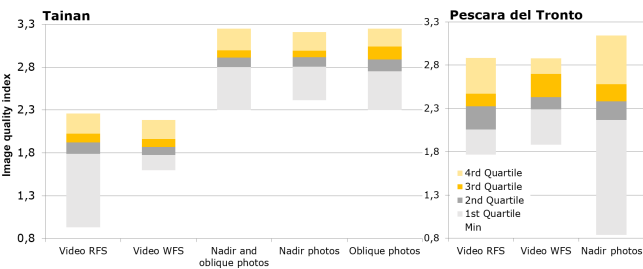
Lower IQI values for video datasets were observed in the Tainan case than in all photo datasets (Fig. 4). The WFS dataset have greatly reduced variance and in the case of Pescara del Tronto raised IQI values. Of the oblique and nadir photos, the nadir dataset showed less variance but similar IQI values in the case of Tainan. In Pescara del Tronto the nadir dataset contained substantial variance. A direct relation between IQI values with resolution was observed also when analyzing resolution-degraded datasets: the better the resolution, the higher the IQI values.

#### 3.1.2 Depictability of 2-D damage-related features

Similar extraction and classification accuracies were obtained for both data types (Table 3). Between the two study areas, higher accuracies of 84 and 82 % were achieved for the Pescara del Tronto video and photo datasets, respectively. For Tainan they were of 77 and 78 %. Errors were mainly associated with false positives (i.e., misclassified and extracted damage features), caused by texture-rich super-pixels with the presence of cars or people in Tainan and low-texture super-pixels, such as bare ground, in the case of Pescara del Tronto (Fig. 5).



**Figure 3.** Segments selected for the depictability of damage-related feature analysis. (a) Damage segment (roughness values), (b) non-damage segment (normal change rate values).



**Figure 4.** Box plot of the image quality index (IQI) computed for Tainan and Pescara del Tronto datasets.

**Table 3.** Accuracy (%) of 2-D damage-related feature extraction for every dataset and study case.

Tested data	Tainan		Pescara del Tronto	
	Video	Photo	Video	Photo
Severe damage	73	67	72	81
Partial damage	71	72	90	74
Non-damage	88	94	87	90
Mean accuracy	77	78	84	82

3.2 3-D quality assessment

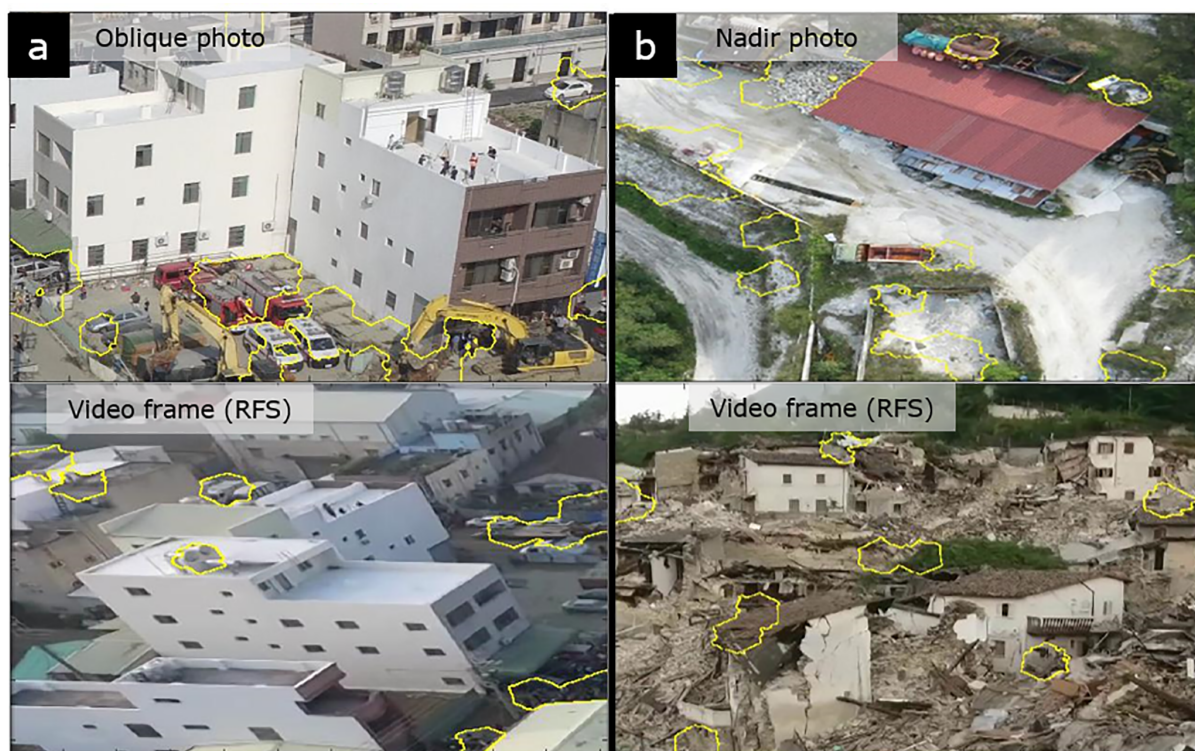
3.2.1 Geometric quality assessment

Planar fitting and completeness assessments were performed to determine internal accuracy. For the former large deviations from the chosen planar object, and thus low precision, were computed in the case of Tainan for the 3DPCs generated with the combined oblique and nadir photos (Fig. 6). The software used was not able to merge oblique and nadir

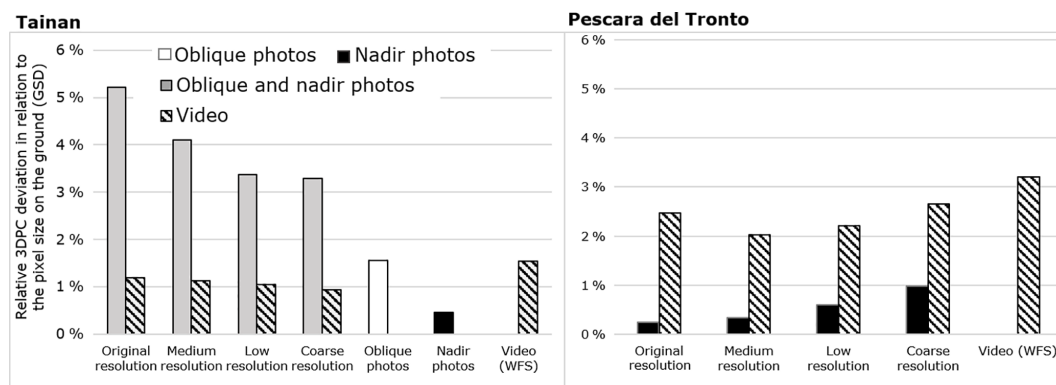
photos, resulting in a highly displaced 3DPC. Improved precision was obtained when using individually the oblique or nadir photo datasets; however in the Tainan case, RFS-video-based 3DPC is still more precise, or at least presents less relative 3DPC deviation, than the oblique-photo-based 3DPC. The 3DPCs generated with nadir photos presented the lowest deviation and were the most precise in both study cases. The 3DPC generated with the WFS video dataset presented more deviation than the one generated with the RFS video dataset in both study cases. Contrasting effects of resolution were found between the study cases. It was observed that resolution degradation can improve precision by reducing 3DPC noise in the Tainan case, while the opposite effect occurred in Pescara del Tronto from medium resolution downwards. Furthermore, the completeness analysis demonstrated that the 3DPCs generated with oblique photos were denser than the video-based 3DPCs in the Tainan case. The RFS-video-based 3DPC in turn was denser than the one generated from nadir photos in both study cases. However, considering the percentage of empty cell, the RFS-video-based 3DPCs percentages are low and outperform percentages of oblique and nadir-based 3DPC in the vertical and horizontal perspective, respectively (Fig. 7 and Table 4). The WFS video dataset produced low-density 3DPCs with large percentages of empty cells in both study cases and perspectives.

Concerning external accuracy, the combined oblique- and nadir-photo-based 3DPC also resulted in high inaccuracy. Independent oblique- and nadir-photo-based 3DPCs registered high accuracies, with mean distances of 2 and 4 cm to the reference model, respectively (Fig. 8). Video-based 3DPCs, with a mean distance of 7 cm to the reference model, resulted in lower accuracy than individual oblique- or nadir-photo-based 3DPCs. It was observed that the WFS-video-based 3DPCs removed some noise in 3DPCs (i.e., less vari-





**Figure 5.** Extraction of 2-D damage-related features using different datasets. Polygons: false positives in the extraction of 2-D damage-related features for (a) Tainan oblique photo and video frame (random frame selection, RFS) and (b) Pescara del Tronto nadir photo and video frame (RFS).



**Figure 6.** Planar fitting analysis of Tainan and Pescara del Tronto 3-D point clouds as part of the internal 3-D quality assessment. Comparison of the mean distance of 3-D points to the theoretical planar object. Larger distance implies lower 3-D point cloud precision.

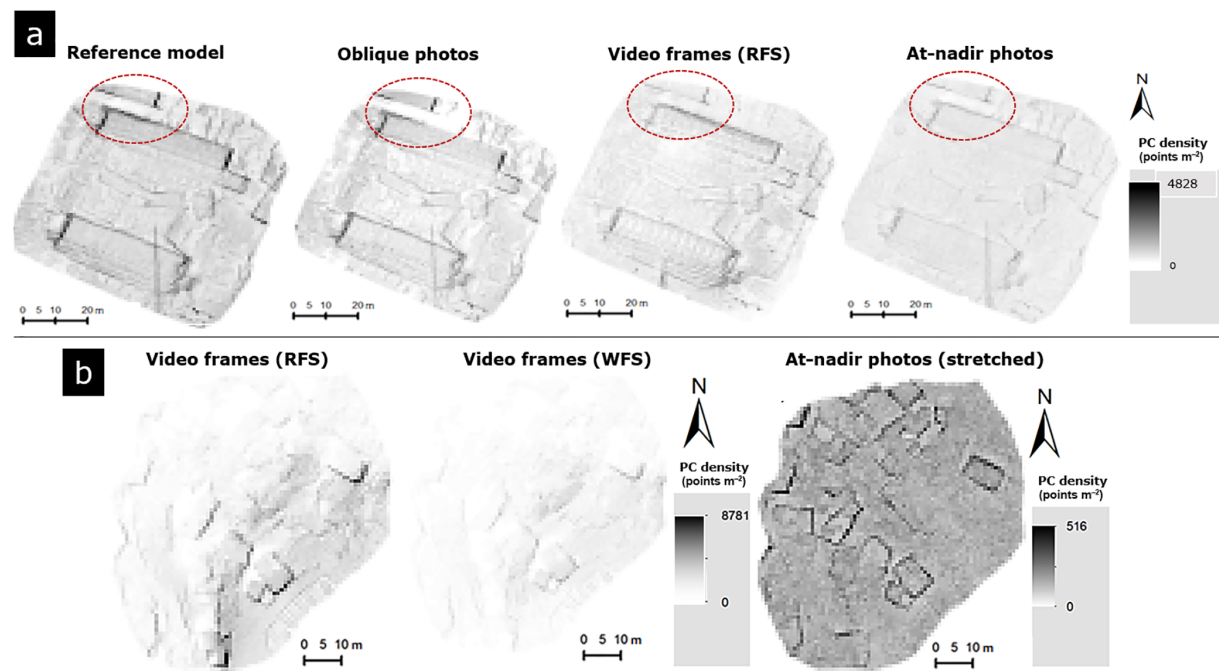
ability), but the accuracy was still the same as RFS-video-based 3DPCs.

### Depictability of 3-D damage-related features

In the case of Pescara del Tronto building deformations, cracks, and debris features could be readily identified by visual inspection in video-based mesh models. These models outperformed representations of nadir-photo-based 3DPC, and even better representations were obtained us-

ing degraded-resolution video frames (Fig. 9). In the Tainan case, cracks and spalling features were better represented in the oblique-photo-based models than in the video-based ones; however, their identification was still difficult, especially compared to the reference mesh. Other damage characteristics, such as inclined walls and spalling, were only extractable from the reference model.

The analysis of the distribution of 3-D damage-related features showed more deviating distributions and therefore less



**Figure 7.** Vertical point cloud density raster maps generated for the completeness analysis. 3-D point clouds of **(a)** Tainan and **(b)** Pescara del Tronto. Dashed red circle: noticeable difference in completeness between 3DPCs of Tainan.

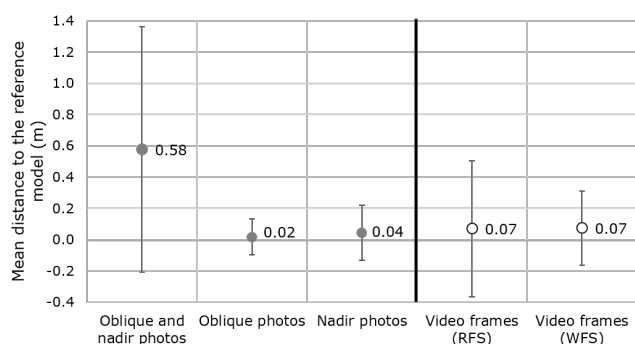
**Table 4.** Computed mean point cloud densities (points per cell or m<sup>2</sup>) and proportion of empty cells (%) from horizontal and vertically projected point cloud density raster maps. Bold numbers are relevant values, in every study case and perspective.

Study case	Perspective	Reference 3DPC	3-D point cloud (PC) source			
			Nadir photos	Oblique photos	Video (RFS)	Video (WFS)
Tainan	Vertical	779 0 %	408 <b>0</b> %	<b>519</b> 4 %	438 <b>0</b> %	135 2 %
	Horizontal	3129 0 %	1649 3 %	<b>1967</b> <b>2</b> %	1816 3 %	573 3 %
Pescara del Tronto	Vertical		167 <b>0</b> %		<b>667</b> 4 %	255 8 %
	Horizontal		499 4 %		<b>1752</b> <b>1</b> %	667 3 %

depictability in photo-based 3DPCs (Fig. 10). RFS-video-based 3DPCs distributions were consistently aligned with the distributions extracted from the reference 3DPC for all 3-D features analyzed. The distributions of the other 3DPCs showed deviations in at least one feature. The relation of these features to damage characteristics was confirmed by the inclusion and analysis of a non-damaged segment extracted from the reference 3DPC, which clearly presented deviations in all 3-D features.

3.3 Application analysis

Debris volume change was analyzed for Tainan using multi-temporal photo-based 3DPCs, and as expected a negative trend was observed in accordance with the cleanup operations (Fig. 11). However, the comparison with the estimation made with the reference 3DPC for the first date (7 February 2016) revealed large overestimations in these series. The debris volume estimation made with the 3DPC generated from the default dataset (i.e., the combined oblique and nadir photos) was around 8000 m<sup>3</sup> more than the estimation done with the reference 3DPC. This is linked to the displacement effects observed in the 3-D external accuracy



**Figure 8.** External accuracy assessment of video- and photo-based 3DPCs of Tainan. Comparison of the mean distance between the analyzed and the reference 3-D point clouds.

assessment. In contrast, volume estimations made with the oblique-photo-based 3DPCs and video RFS dataset for the same date presented more accurate results. The estimation with the oblique-based 3DPC was only 486 m<sup>3</sup> above the reference (1.6 % of error), while the most accurate estimation was obtained with the RFS-video-based 3DPC, with an error of only 136 m<sup>3</sup> (0.4 %).

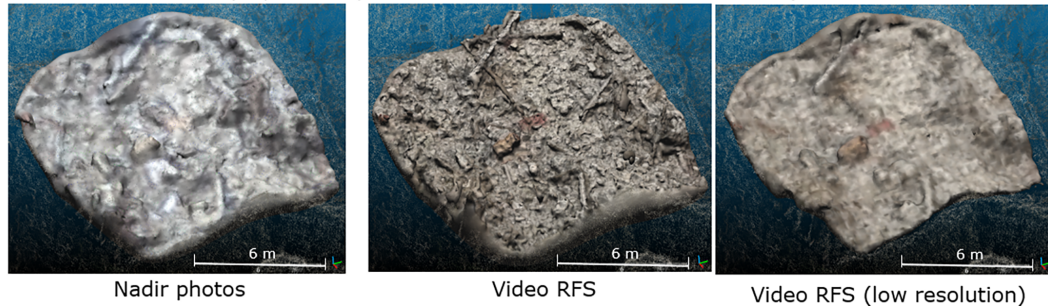
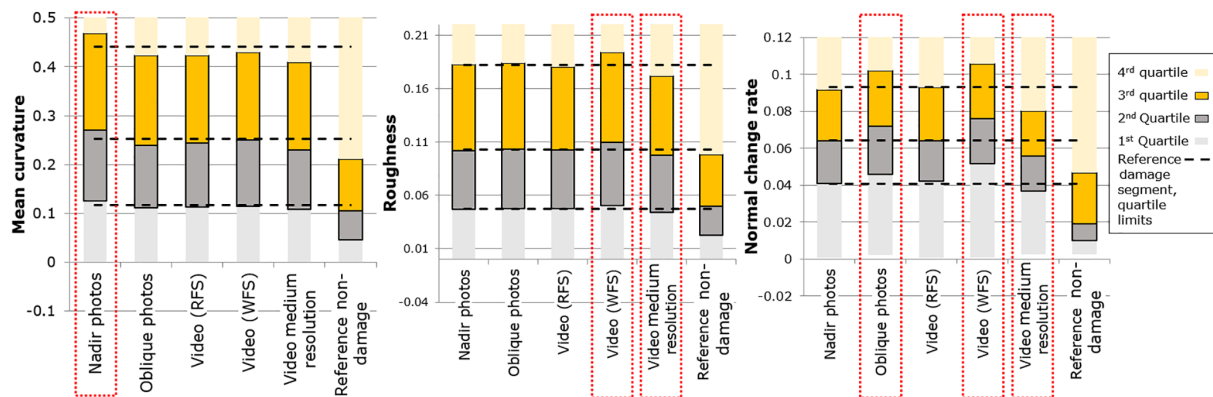
## 4 Discussions

In the 2-D quality analysis lower IQI values for the video datasets were obtained in comparison to photos in the case of Tainan, while they were similar in the case of Pescara del Tronto. Low video IQI values were caused by low resolution and smoke presence in Tainan video frames, whereas in the case of Pescara del Tronto these were comparable since at-nadir photos were affected by a large GSD and the presence of motion-blur effects, generated during the acquisition of nadir photos by the high altitude and low shutter speed, respectively (see Sect. 2.1). The WFS dataset improved IQI values and reduced variance in comparison to the RFS video dataset, since the WFS approach filtered frames affected by external factors (e.g., low-texture zones, smoke presence, and moving objects), clearly indicated by low IQI values in IQI series (Fig. 12). Despite direct 2-D video quality being lower than that of photos, similar accuracies were obtained from the depictability of 2-D damage-related analysis. In this analysis, external factors were also responsible for false-positive errors and low accuracy in the classification of 2-D damage-related features, rather than the video data quality characteristics. Moreover, the analyzed video frames were chosen from the RFS dataset; therefore, higher accuracies could be obtained using frames from the WFS dataset. Low resolution was the only video characteristic that affected 2-D quality; other data-related quality artifacts, such as motion-blur effects, were instead identified in nadir photos. Therefore, sufficient 2-D quality and accurate depictability of damage-

related features of video data demonstrated that they can perform 2-D-based SDA as accurately as photos.

Concerning the 3-D quality assessment, comparable qualities were obtained for video- and photo-based 3DPCs. Photo-based 3DPCs of Tainan generated with the combined oblique and nadir dataset and its degraded-resolution datasets (Table 2: datasets 1 to 5) produced largely displaced 3DPCs that presented the lowest quality (Fig. 13). A main possible cause is the large difference in the spatial and spectral characteristics of the combined datasets which affected the automatic tie point selection during the initial 3-D modeling step. A less probable cause is a misregistration error due to the lack of GCPs for orienting oblique photos and their low  $z$  variability (see Sect. 2.2 and 2.1). A solution would be a refinement procedure; however, this implies a time-demanding operation that hinders photo data usability. In contrast, 3DPCs generated independently from nadir and oblique photos showed higher qualities. Three-dimensional qualities of video-based 3DPCs were also high and comparable to the ones of oblique- or nadir-photo-based 3DPCs. In some quality measurements, video-based 3DPCs even outperformed the quality of the 3DPCs generated from oblique or nadir photos. For example, the Tainan RFS-video-based 3DPC was denser and presented a lower percentage of empty cells; occluded areas were uniquely identified in the RFS-video-based 3DPC (Fig. 7). The better performance of video data in 3DPC generation is associated with their small GSD, which compensated for their lower resolution, their complete scope of the scene in the case of Tainan, and the motion blur that affected the nadir-photo-based 3DPCs in the case of Pescara del Tronto. Similarly, in the depictability of 3-D damage-related features analysis, 3-D damage-related features were in many cases better identifiable in the RFS-video-based 3DPC than in the nadir- and oblique-photo-based 3DPCs, both visually and using the CNN model. The influence of video artifacts and quality characteristics was not evident either in this quality analysis. Most effects were produced by external factors, such as the presence of smoke and changes in the scene in the case of Tainan, and hilly topography which affected completeness in the case of Pescara del Tronto. However, the influence of external factors was still marginal compared to the favorable small GSD and coverage of video frames. In relation to the WFS approach, 3DPC noise mainly related to smoke in Tainan was reduced by filtering the affected frames; however it did not improve, and even reduced, the quality obtained using the RFS-video-based 3DPC. The limited number of video frames in the WFS dataset resulted in low 3DPC precision, completeness, and depictability of 3-D damage-related features. Noise in 3DPCs was also reduced when using the down-sampled Tainan video and photo datasets. Proximate or even higher-quality 3-D parameters of RFS-video-based 3DPCs, together with their proper depictability of 3-D damage-related features, indicated that video frames can also be used for accurate 3-D-based SDA.



**Pescara del Tronto** (depictability of debris features in 3-D mesh models)**Figure 9.** Comparison of the depictability of debris features in photo- and video-based mesh models of Pescara del Tronto. Lower depictability in nadir-photo-based mesh due to their high GSD and the presence of motion-blur effects.**Figure 10.** Tainan segment-level distributions of 3-D damage-related features. Dotted rectangles highlight discordant distributions from the reference damage segment and thus low depictability of 3-D damage-related features.

The application analysis led to results consistent with the 3-D quality assessment, since also here the displacement effect of the combined oblique- and nadir-photo-based 3DPC produced a large overestimation in comparison with the reference. Video data in spite of their low resolution and the influence of all the mentioned external factors achieved an accurate estimation, comparable to the one using the oblique-photo-based 3DPC. Thereby, it was demonstrated that the RFS-video-based 3DPC was able to perform debris volume estimation application with a level of accuracy the same as or better than with the photo-based 3DPC.

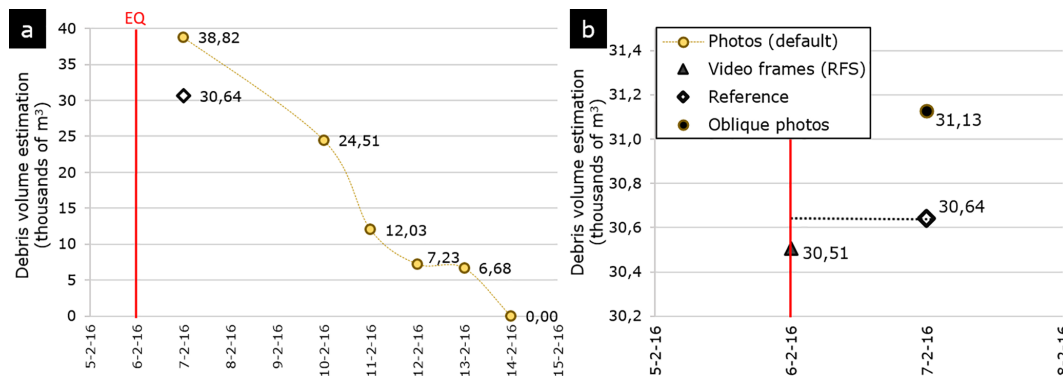
Results obtained in this research approach the ones observed in previous studies. Accuracies achieved for the depictability of 2-D damage-related feature analysis are similar to the ones obtained in Vetrivel et al. (2017), mainly for the case of Pescara del Tronto. Likewise, a similar external accuracy of 5 cm was obtained by Alsadik et al. (2015) for video-based 3DPCs, 2 cm more than the one obtained in this research. Here the method used to select the highest-quality video frames (i.e., WFS) produced low-density point clouds, in contrast to Xu et al. (2016), who claimed that the reduction of motion-blur frames generated denser 3DPCs. The authors, however, did not consider redundancy, the reduction of which

implies the elimination of a large number of frames. Dealing with frame redundancy remains a challenging issue; until now empirical values have been used to define a proper frame selection rate, though this is data- and case-specific (Alsadik et al., 2015; Clift and Clark, 2016). A video characteristic that was not considered but that influenced the accuracy of the produced 3DPC is the lack of proper geo-positional information. All video-based 3DPCs were registered in relation to the reference 3DPC; however, the lack of reference information or adequate GCPs will affect usability of this kind of data. Similar results are expected with the use of other commercial softwares (e.g., Agisoft, Photoscan) or open-source packages (e.g., ColMap, MicMac).

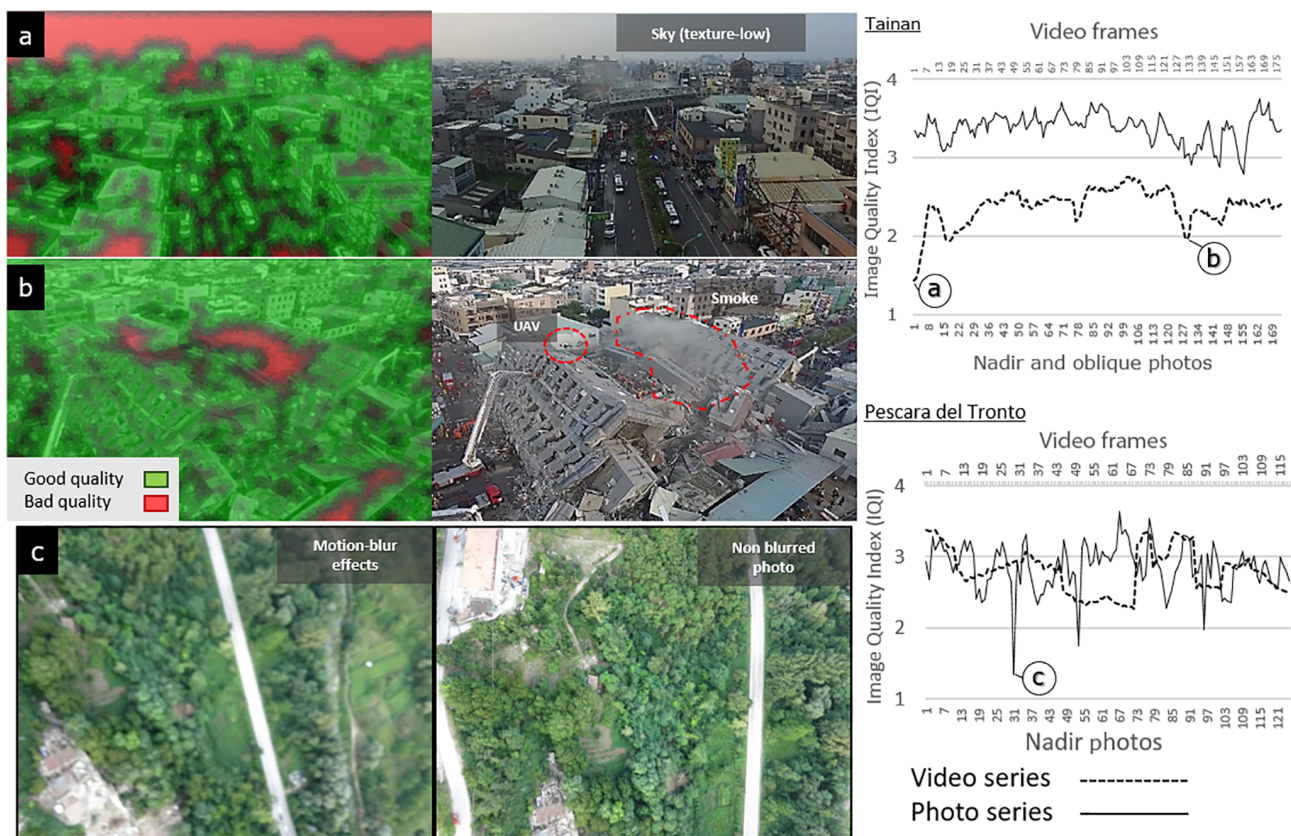
## 5 Conclusions

The usability of video data for SDA was investigated based on the analysis of their 2-D and 3-D quality and application in post-disaster activities. The 2-D quality analysis indicated possible sources of error expressed by low IQIs for the video datasets, attributed to low resolution and external factors (e.g., smoke presence, low-texture zones). However, the analysis of the depictability of 2-D damage-related fea-





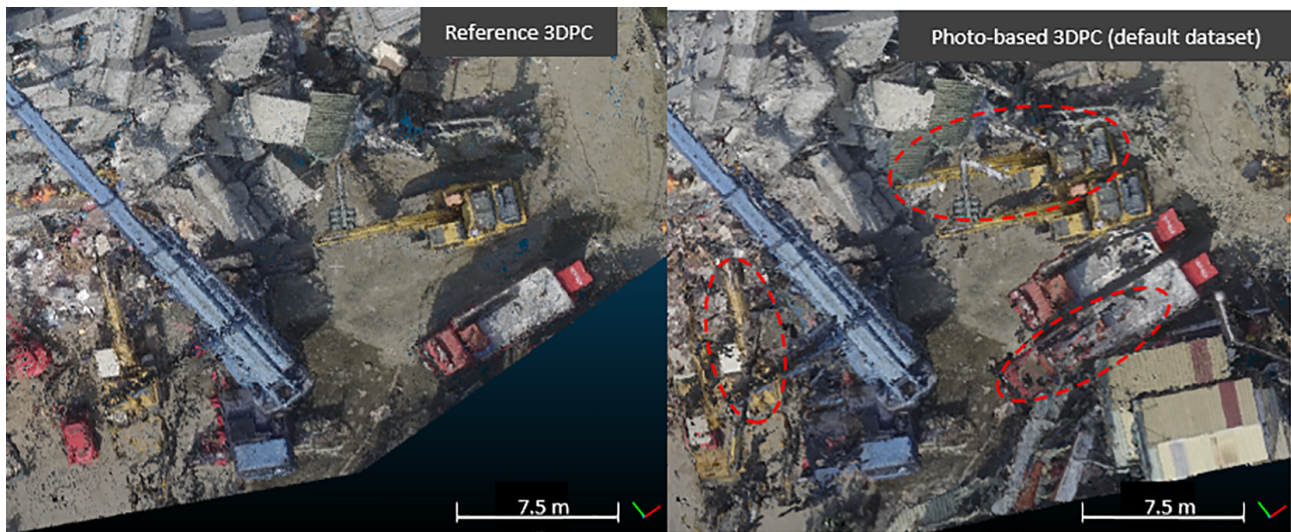
**Figure 11.** (a) Debris volume change analysis using the Tainan multi-temporal photo-based 3DPC and reference model. (b) Volume estimations for video-based 3DPC (random frame selection), oblique-photo-based 3DPC, and reference model. Red line: earthquake (EQ) date.



**Figure 12.** Image quality index (IQI) series for Pescara del Tronto and Tainan. (a) The sky as a low-texture area is considered a likely source of 3-D modeling error in IQI maps (red area). (b) Also smoke and moving objects (UAV). (c) In Pescara del Tronto motion-blur effects result in IQI values.

tures demonstrated that this kind of data can be used for an accurate 2-D-based SDA. Video-based 3DPCs in turn exhibited a geometrical quality comparable to 3DPCs generated with photos acquired with still cameras. The 3DPCs generated with the RFS video dataset were in certain cases more precise and complete than those from nadir or oblique photos. External accuracy identified slightly less accurate 3DPCs

generated with the RFS video dataset; however, based on the depictability of 3-D damage-related feature analysis, it is sufficient for 3-D-based SDA. Similarly, the RFS-video-based 3DPC was able to estimate debris volume more accurately than models based on oblique photos, confirming the good usability of video data. Data resolution was the main quality parameter affecting 2-D and 3-D video quality; however, it



**Figure 13.** Displacement effect in 3DPC generated with the default photo dataset (combined nadir and oblique photos), due to a misregistration or differences in spectral and spatial characteristics between oblique and nadir datasets. Red dashed circles: noise in 3DPCs.

has been seen that a small GSD (i.e., low altitude) complemented with a complete scope of the scene can compensate for low-video-resolution effects. Consequences of other data artifact and quality characteristics, such as redundancy and motion-blur effects were not identified; in contrast, external factors such as smoke presence, terrain texture, or changes in the scene had more of an influence on video data quality and usability. The reduction of redundant video frames can also result in low-quality 3DPCs, as observed with WFS approach. The use of a more elaborated approach might result in more accurate 3DPCs; however, the variability of data and external characteristics limits the development of an optimal frame selection approach that can deal with all conditions. A relevant quality parameter not considered in this research is the availability of accurate GCPs for 3DPC geo-registration. The absence of proper GCPs, mainly for oblique images or video frames, may generally affect the 3-D quality, as the displacement observed in the 3DPC generated using the combined oblique–nadir photos. Methods for direct geo-registration, such as described by Turner et al. (2014), could be applied and tested with video data. Also, as video data were limited to 2.1 MP resolution, better performance is expected for current 4 K ultra-high-definition (UHD) videos of 8.8 MP. Both aspects could be investigated by a synthetic experiment as a continuation of this research, and this may also contribute in the application of video data for real-time damage estimations with deep learning methods.

**Data availability.** Data used in this research are third-party property. For request contact Jyun-Ping Jhan (jyun-pingjhan@geomatrics.ncku.edu.tw) (Tainan, Taiwan) and Filiberto Chiabrando (filiberto.chiabrando@polito.it) (Pescara del Tronto).

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** The work was supported by FP7 project INACHUS (Technological and Methodological Solutions for Integrated Wide Area Situation Awareness and Survivor Localisation to Support Search and Rescue Teams), grant number 607522. The data used in this research were acquired and kindly provided by Jyun-Ping Jhan, Institute of Photogrammetry, University of Stuttgart; Jiann-Yeou Rau, National Cheng Kung University, Tainan (Tainan UAV photos and GCPs); and Filiberto Chiabrando Politecnico di Torino, Italy (Pescara del Tronto UAV photos and GCPs). The CNN damage classification model was in turn provided by Anand Vetrivel, who also contributed to this research with insight and expertise.

Edited by: Filippo Catani

Reviewed by: two anonymous referees

## References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S.: Slic superpixels, Tech. rep., 2010.
- Ahmed, M. T., Dailey, M. N., Landabaso, J. L., and Herrero, N.: Robust Key Frame Extraction for 3D Reconstruction from Video Streams, in: VISAPP (1), edited by: Richard, P. and Braz, J., 231–236, INSTICC Press, <http://dblp.uni-trier.de/db/conf/visapp/visapp2010-1.html#AhmedDLH10>, 2010.
- Alsadik, B., Gerke, M., and Vosselman, G.: Efficient use of video for 3D modelling of cultural heritage objects, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W4, 1–8, <https://doi.org/10.5194/isprsannals-II-3-W4-1-2015>, 2015.

- Clift, L. G. and Clark, A. F.: Video frame extraction for 3D reconstruction, in: 2016 8th Computer Science and Electronic Engineering (CEECE), 152–157, <https://doi.org/10.1109/CEECE.2016.7835905>, 2016.
- CloudCompare: CloudCompare 3D point cloud and mesh processing software Open Source Project, <http://www.cloudcompare.org/>, 2017. (last access: 29 May 2018)
- Dong, L. and Shan, J.: A comprehensive review of earthquake-induced building damage detection with remote sensing techniques, *ISPRS Journal of Photogrammetry and Remote Sensing*, 84, 85–99, <https://doi.org/10.1016/j.isprsjprs.2013.06.011>, <http://www.sciencedirect.com/science/article/pii/S0924271613001627>, 2013.
- 3DFlow: 3DF Zephyr, Version 3.0, computersoftware, available at: <http://www.3dflow.net/>, 2017.
- 3D Flow: 3DF Zephyr 3.0 user manual, available at: <http://www.3dflow.net/technology/documents/3df-zephyr-tutorials/computersoftwaremanual>, 2017.
- Eugster, H. and Nebiker, S.: Geo-registration of video sequences captured from Mini UAVs: Approaches and accuracy assessment, in: The 5th International Symposium on Mobile Mapping Technology, Padua, Italy. Retrieved March, vol. 12, 2012, 2007.
- Fernandez Galarreta, J., Kerle, N., and Gerke, M.: UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning, *Nat. Hazards Earth Syst. Sci.*, 15, 1087–1101, <https://doi.org/10.5194/nhess-15-1087-2015>, 2015.
- FSD: Drones in Humanitarian Action. A guide to the use of airborne systems in humanitarian crises, Tech. rep., Swiss Foundation for Mine Action (FSD), Geneva, <http://drones.fsd.ch/>, 2016. (last access: 29 May 2018)
- Gerke, M. and Kerle, N.: Automatic structural seismic damage assessment with airborne oblique Pictometry© imagery, *Photogr. Eng. Remote Sens.*, 77, 885–898, 2011.
- Girardeau-Montaut, D.: Completeness/Coverage of point cloud, <http://www.cloudcompare.org/forum/viewtopic.php?t=1433#p5553>, 2017. (last access: 29 May 2018)
- Jarzabek-Rychard, M. and Karpina, M.: Quality analysis on 3D building models reconstructed from UAV imagery, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B1, 1121–1126, <https://doi.org/10.5194/isprs-archives-XLI-B1-1121-2016>, 2016.
- Kerle, N. and Stekelenburg, R.: Advanced structural disaster damage assessment based on aerial oblique video imagery and integrated auxiliary data sources, in: *ISPRS 2004: proceedings of the XXth ISPRS congress: Geoimagery bridging continents*, 12–23, Citeseer, 2004.
- Kersten, T. P. and Lindstaedt, M.: Automatic 3D object reconstruction from multiple images for architectural, cultural heritage and archaeological applications using open-source software and web services, *Photogrammetrie-Fernerkundung-Geoinformation*, 2012, 727–740, 2012.
- Khoshelham, K., Elberink, S. O., and Xu, S.: Segment-based classification of damaged building roofs in aerial laser scanning data, *IEEE Geosci. Remote Sens. Lett.*, 10, 1258–1262, 2013.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in: *Advances in Neural Information Processing Systems 25*, edited by: Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., 1097–1105, Curran Associates, Inc., <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>, 2012. (last access: 29 May 2018)
- Mitomi, H., Matsuoka, M., and Yamazaki, F.: Application of automated damage detection of buildings due to earthquakes by panchromatic television images, in: *The 7th US National Conference on Earthquake Engineering*, CD-ROM, p. 10, 2002.
- Nex, F. and Remondino, F.: UAV for 3D mapping applications: a review, *Appl. Geomat.*, 6, 1–15, <https://doi.org/10.1007/s12518-013-0120-x>, 2014.
- Ogawa, N. and Yamazaki, F.: Photo-interpretation of building damage due to earthquakes using aerial photographs, in: *Proceedings of the 12th World Conference on Earthquake Engineering*, 1906, 2000.
- Oude Elberink, S. and Vosselman, G.: Quality analysis on 3D building models reconstructed from airborne laser scanning data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 157–165, <https://doi.org/10.1016/j.isprsjprs.2010.09.009>, 2011.
- Ozisik, D. and Kerle, N.: Post-earthquake damage assessment using satellite and airborne data in the case of the 1999 Kocaeli earthquake, Turkey, in: *Proc. of the XXth ISPRS congress: Geoimagery bridging continents*, 686–691, 2004.
- Pix4D: Pix4D Mapper: Drone Mapping Software for Desktop + Cloud + Mobile, <https://pix4d.com/>, version 3.1.18, 2017.
- Saito, K., Spence, R., Booth, E., Madabhushi, M., Eguchi, R., and Gill, S.: Damage assessment of Port au Prince using Pictometry, in: *8th International Conference on Remote Sensing for Disaster Response*, Vol. 30, p. 2010, Tokyo, 2010.
- Samadzadegan, F. and Rastiveisi, H.: 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, 415–420, 2008.
- Singh, S. P. Jain, K., and Ravibabu Mandla, V.: 3D scene reconstruction from video camera for virtual 3d city modeling, *Am. J. Eng. Res.*, 3, 140–148, 2014.
- Soudarissanane, S., Lindenbergh, R., and Gorte, B.: Reducing the error in terrestrial laser scanning by optimizing the measurement set-up, in: *XXI ISPRS Congress, Commission I-VIII*, 3–11 July 2008, Beijing, China, International Society for Photogrammetry and Remote Sensing, 2008.
- Tu, J., Sui, H., Feng, W., and Song, Z.: Automatic Building Damage Detection Method Using High-Resolution Remote Sensing Images and 3D GIS Model, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43–50, 2016.
- Turner, D., Lucieer, A., and Wallace, L.: Direct georeferencing of ultrahigh-resolution UAV imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 52, 2738–2745, 2014.
- Vetrivel, A., Gerke, M., Kerle, N., and Vosselman, G.: 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*, <https://doi.org/10.1016/j.isprsjprs.2015.03.016>, 2015.
- Vetrivel, A., Duarte, D., Nex, F., Gerke, M., Kerle, N., and Vosselman, G.: Potential of multi-temporal oblique airborne imagery for structural damage assessment, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3, 355–362, <https://doi.org/10.5194/isprs-annals-III-3-355-2016>, 2016a.

- Vetrivel, A., Gerke, M., Kerle, N., and Vosselman, G.: Identification of structurally damaged areas in airborne oblique images using a Visual-Bag-of-Words approach, *Remote Sens.*, <https://doi.org/10.3390/rs8030231>, 2016b.
- Vetrivel, A., Gerke, M., Kerle, N., Nex, F., and Vosselman, G.: 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, *ISPRS Journal of Photogrammetry and Remote Sensing*, <https://doi.org/10.1016/j.isprsjprs.2017.03.001>, 2017.
- Weinmann, M., Jutzi, B., Hinz, S., and Mallet, C.: Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers, *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286–304, 2015.
- Xu, Z., Wu, T. H., Shen, Y., and Wu, L.: Three dimensional reconstruction of large cultural heritage objects based on UAV video and TLS data, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 985–988, <https://doi.org/10.5194/isprs-archives-XLI-B5-985-2016>, 2016.