International Journal of Digital Earth

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjde20

The role of machine intelligence in photogrammetric 3D modeling – an overview and perspectives

Rongjun Qin & Armin Gruen

To cite this article: Rongjun Qin & Armin Gruen (2020): The role of machine intelligence in photogrammetric 3D modeling – an overview and perspectives, International Journal of Digital Earth, DOI: 10.1080/17538947.2020.1805037

To link to this article: https://doi.org/10.1080/17538947.2020.1805037

Published online: 10 Aug 2020.

INTERNATIONAL JOURNAL OF DIGITAL EARTH
https://doi.org/10.1080/17538947.2020.1805037

REVIEW ARTICLE

OPEN ACCESS

The role of machine intelligence in photogrammetric 3D modeling – an overview and perspectives

Rongjun Qin[a,b,c] and Armin Gruen[d]

[a]Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA; [b]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA; [c]Translational Data Analytics Institute, The Ohio State University, Columbus, OH, USA; [d]Information Architecture, ETH Zürich, Zürich, Switzerland

ABSTRACT
The process of modern photogrammetry converts images and/or LiDAR data into usable 2D/3D/4D products. The photogrammetric industry offers engineering-grade hardware and software components for various applications. While some components of the data processing pipeline work already automatically, there is still substantial manual involvement required in order to obtain reliable and high-quality results. The recent development of machine learning techniques has attracted a great attention in its potential to address complex tasks that traditionally require manual inputs. It is therefore worth revisiting the role and existing efforts of machine learning techniques in the field of photogrammetry, as well as its neighboring field computer vision. This paper provides an overview of the state-of-the-art efforts in machine learning in bringing the automated and 'intelligent' component to photogrammetry, computer vision and (to a lesser degree) to remote sensing. We will primarily cover the relevant efforts following a typical 3D photogrammetric processing pipeline: (1) data acquisition (2) geo-referencing/interest point matching (3) Digital Surface Model generation (4) semantic interpretations, followed by conclusions and our insights.

ARTICLE HISTORY
Received 13 May 2020
Accepted 30 July 2020

KEYWORDS
Photogrammetry; camera calibration; 3D modeling; machine learning; object recognition; semantic interpretation

1. Introduction

Photogrammetry is a focused and widely applicable field, largely backboned by the geospatial industry in 3D measurable model and process generation and mapping. Given the fact that photogrammetry shares the geometric aspects of its neighboring disciplines – computer vision and computer graphics, thus providing an integrated production line from data acquisition, geometric data processing, 2D/3D/4D interpretation, recognition modeling, to data administration and representation (Gruen et al., 2009; Lu et al., 2004; Qin, 2015a; Qin et al., 2016; Shan et al., 2020). Although photogrammetric techniques aim at providing practical solutions for data generation and interpretation, with often manual interactions (e.g. geometric modeling, object identification and monitoring), the endeavors of the field are to develop more automated methods that involve the use of AI (artificial intelligence) techniques to: (1) automate processes that traditionally require heavy manual operation (e.g. building extraction and modeling) (Gruen and Wang, 1998; Lillesand et al., 2014); (2) improve performance of processes in terms of efficiency and robustness (e.g. in point matching).

CONTACT Rongjun Qin qin.324@osu.edu Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, 2036 Neil Avenue, Columbus, OH, USA

Hereby, the term 'Artificial Intelligence' is sometimes regarded as misleading and might often subject to wrong ideas and expectations: for instance Prof. Sebastian Thrun (Co-founder of the On-Line Academy Udacity, former Google Vice-President) has remarked in a recent interview with the German Newspaper DIE ZEIT: '(AI) is not a good choice of words. Existing systems are not intelligent. Essentially, they do pattern recognition in large datasets. They can learn rules and apply them. Missing are emotions, creativity, freedom of opinions, autonomy. A computer cannot handle this'(ZEIT, 2020). On the other hand, 'Machine Learning' is a technical term that is explicitly about computational methods that learn associations/functions from data rather than traditional physical-model based methods. Therefore, in this paper, the term 'Machine Intelligence' largely refers to machine learning techniques with services to the 'intelligence' aspects and needs of applications.

AI methods, in case of successful performance, can potentially have a big impact on the Society at large, the human behavior and everyday activities. As example, let us have a look at the technique of face recognition.

Photogrammetric face measurement is by no means new. Already in Lacmann, 1950, 151 ff. (Buchholtz, 1950), a number of examples can be found related to Medicine and Anthropology. We find there already the structured light technique as measurement method. Of course, the techniques have improved since then in different directions. Nowadays, already these methods are practiced in medical field such as orthodontics and others (Deli et al., 2013; Haleem and Javaid, 2019), as well as anthropology such as skeleton reconstructions (Lussu and Marini, 2020). But the most recent and relevant innovation is the face recognition which uses the 3D information from face measurement plus AI techniques for recognition tasks. These systems (based on structured light techniques) are already built into smartphones (e.g. iPhone X). A leading company in this field is Megvii Technology, China. They use the software Face ++ with the Deep Learning software FrameworkBrain ++, and work on a number of applications, as for instance

- payments in shops ('Smile to Pay')
- replacement of boarding cards on airports
- ATM machine access
- 24-hours supermarket without personnel
- control of sleeping rooms in student dormitories
- control of public toilettes
- criminology
- traffic rules violations (China: 176 Mill. monitoring cameras, until 2020: 400 Mill. new ones)

An interesting application was reported from the train station of Zhengzhou early in 2018 (Stern.de, 2018). Some policemen were equipped with special sunglasses, including the required sensors and connected to tablets with a database of criminals. This way 7 criminals could be recognized.

Under those new scenarios the major question is 'How to keep privacy'? For instance, as of 2016 there are 117 Million Americans in the face recognition database of the FBI (Newman, 2016).

The recent prevalence of machine learning (interchangeable with the term 'artificial intelligence', but more to the point) has shown a great potential in addressing complex tasks with impressive performance, thus attracting attention in the field of photogrammetry and in particular in computer vision (Hinton and Salakhutdinov, 2006; LeCun et al., 2015). Although neither AI itself, nor the involvement of AI in the field is new, while the recent rise of their development have encouraged us to revisit their role in the field of photogrammetry, as well as their already active role in computer vision (Goodfellow et al., 2016; Szegedy et al., 2016). A very recent study in nature neuroscience (Bonnen et al., 2020) indicated that the binocular viewing geometry evidentially shape the human neural representation and therefore there is a great potential to utilize 3D modeling techniques to enhance 'AI'. There are a plethora of existing works that apply machine learning for solving spatially related issues, and the recent top tier computer vision conferences (e.g. CVPR (IEEE Conference on

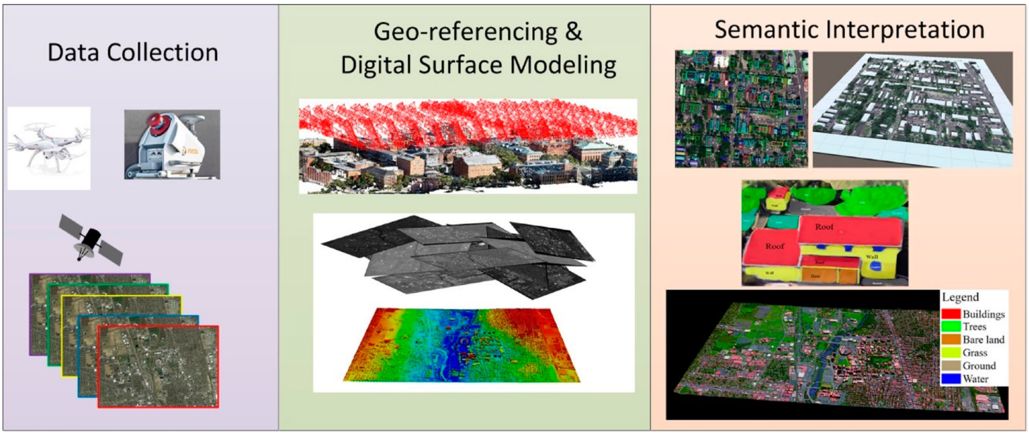

Figure 1. Schematic images of major tasks in 3D photogrammetry and remote sensing.

Computer Vision and Pattern Recognition), ICCV (IEEE international conference on Computer Vision), and ECCV (European Conference ion Computer Vision), etc.) are filled with machine learning (deep learning in particular) based works, with some of them relevant to photogrammetry and remote sensing (Lu et al., 2020; Robinson et al., 2019) (Blaha et al., 2016; Ozcanli et al., 2016; Treible et al., 2018). In this article, we provide a general overview of works that use machine learning and address critical components of the photogrammetric data processing pipeline, including (1) data acquisition; (2) geo-referencing; (3) Digital Surface Model generation; (4) semantic interpretation. Examples are shown in Figure 1.

It should be noted that since some of these components share very similar objectives and topics with their counterparts in computer vision, we introduce works from both fields without explicit distinctions. Due to the large body of existing works, our overview is not able to include each individual work. However, we will cover works that are mostly representative in the sense of being ‘self-learning & automated’ in solving matters related to the aforementioned four components.

This overview paper is organized as follows: Section 2 provides a technical overview on the topics to be covered, including a very brief introduction of a typical photogrammetric processing pipeline, as well as a general introduction of statistical learning and deep learning methods. Section 3 outlines existing efforts in aspects related to the introduced photogrammetric data processing pipelines. Section 4 draws conclusions with our general insights regarding the ‘self-learning’ in photogrammetry.

2. A technical overview

2.1. Photogrammetric data processing pipeline

Despite many other applications, in general, the photogrammetric data processing mainly refers to the spatially related data production of various types, such as orthophotos, digital surface models (DSM), digital terrain models (DTM), 3D polygonal/polyhedral models (all level of details) (Gröger et al., 2007), 4D products (with the time dimension), and of course, applications associated with their immediate level or final forms (Qin et al., 2016). The photogrammetric data processing pipeline, since it can be very often executed fully automatically by using imagery from off-the-shelf cameras, inexpensive sensor platforms like UAVs and Open Source or otherwise affordable software, allows non-experts of various kinds to use it and produce 3D models for their domain applications. On Facebook one can witness a large number of groups (e.g. Photogrammetry Group, https://www.facebook.com/groups/3dphotogrammetry/), with partly several thousand members, who post their works on a daily basis. While most of these models look very attractive visually, usually nothing is said about their fidelity and accuracy.

The pipeline consists of two sets of broadly defined problems, (1) geometric processing, (2) object labeling, topology reconstruction and change detection. Geometric processing (GP) refers to the process of converting raw sensory data all the way to explicit 3D information, e.g. 3D measurements/3D triangle meshes with photo-realistic textures. The second problem set, object labeling, topological reconstruction and change detection (Anders et al., 2020; Cornelis et al., 2008; Diakité et al., 2014; Liebelt and Schmid, 2010; Peng et al., 2019; Qin et al., 2016; Verdie et al., 2015), refer to the processes of identifying the types of objects and their individual components (e.g. planar, cylindrical, polyhedral), modeling geometrical/topological relationships of these objects/components (Cornelis et al., 2008; Foerstner, 1999; Lafarge et al., 2008; Liang et al., 2019), as well as tracking the chronological differences of these objects based on the time-sequence datasets to build a 4D information stack (Anders et al., 2020; Bouziani et al., 2010; Doxani et al., 2010; Goncaluves, 2010; Tian et al., 2010).

These two problem sets are in concert with the low-level and mid/high-level vision problems in computer vision (definition may slightly vary), where topics are even more widely defined (Forsyth and Ponce, 2002). The low-level vision topics originally deals with tasks that stay in the retina level that do not need cognition processes, e.g. edge/interest point extraction and image or point cloud matching. The high-level vision problem usually refers to vision tasks that trigger cognition process, e.g. object type recognition, human activity recognition and parsing. The latter have a direct tie to intelligence and thus machine learning techniques are investigated in this area (Förstner and Wrobel, 2016; Szeliski, 2010).

There are a few different taxonomies on machine learning methods based on different criterions. To include the recently boosted deep neural networks, we differentiate the machine learning methods into statistical learning (or shallow learning methods) (James et al., 2013) and deep learning methods (primarily refers to the deep neural networks) (Goodfellow et al., 2016). The major differences between them are the complexity of the models and the use of manually crafted feature extraction method or implicitly learned representations (features) (LeCun et al., 2015), a schematic figure is shown in Figure 2.

2.2. Statistical learning/shallow classifier

Prior to the recent prevalence of the deep neural networks (Krizhevsky et al., 2012; LeCun et al., 2015), statistical learning plays a major role in the high-level vision problems or object

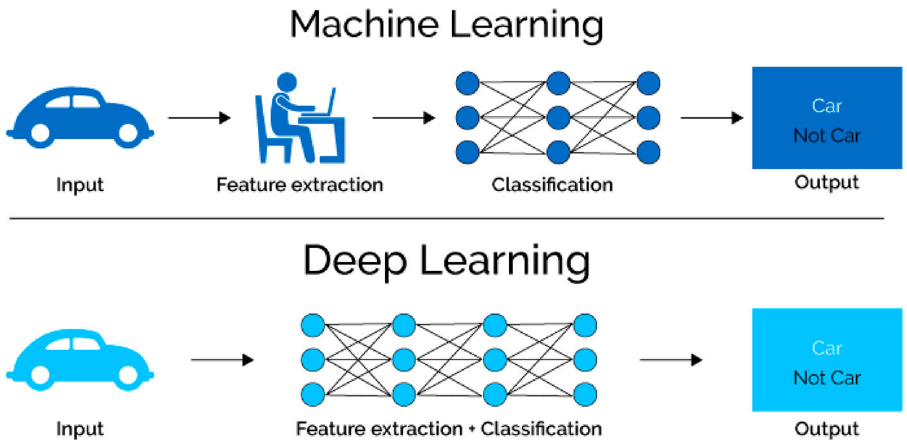

Figure 2. Difference between traditional machine learning (often called statistical learning) and deep learning (xenonstack, 2020).

recognition/image classification in photogrammetry and remote sensing (Cheng et al., 2017; Costa et al., 2017; Gómez et al., 2016; Huang and Zhang, 2013; Qin, 2015b; Wijaya et al., 2015). The statistical learning methods mainly refer to learning methods that use sets of statistical tools for modeling and understanding complex datasets. Very often they are used in parallel to the machine learning and artificial intelligence fields in computer science, broadly incorporating methods such as logistic regression (Menard, 2018), support vector machine (Wang, 2005), decision trees (Friedl and Brodley, 1997) and maximal-likelihood classifiers (Foody et al., 1992).

Given that the number of parameters which the statistical model is manageable is relatively small (normally less than a few hundreds), such methods are also called shallow classifiers (Bianchini and Scarselli, 2014). These methods have played predominant roles in the area of satellite image classification (Ma et al., 2017), in particular in the application of land-cover & land-use change mapping (Justice et al., 2015), having been implemented in multiple professional photogrammetry and remote sensing software packages. Although in practice it is generally accepted that, to produce reasonably high-accuracy land-cover maps, given the increasing amount of data to be processed for smart applications (high-frequency monitoring) (Ahmed et al., 2008; Qin, 2014; Rathinam et al., 2008), there exists a high demand for further improvements in terms of fewer ad-hoc samples (those from the images to be processed) and higher accuracy (Yosinski et al., 2014). Moreover, areas of interest are also shifted to applications of intelligent interpretation of very-high resolution (sub-meter level) satellite images, as well as 3D point cloud data produced by photogrammetry methods and LiDAR sensors (Qin, 2019a; Qin, 2019b; Qin et al., 2015; Qin and Gruen, 2014; Tian, 2013).

Among all the statistical learning methods, Support Vector Machine (SVM) (Wang, 2005) and Random Forest (RF) (Breiman, 2001) are two of the major representative methods used in classifying very-high resolution data (Pal, 2005; Pal and Mather, 2005), and have been favorably applied in many applications in object recognition, retrieval and modeling. Neural Networks (NN) is also one of the comparable methods in classification, and normally refers to those with no more than two hidden layers (Anderson, 1995; De Veaux and Ungar, 1997) (in contrast of the deep learning neural network in the next section). In the field of computer vision, these methods have been widely used in many intelligence-requiring applications including human action recognition, car detection, face recognition (Galantucci et al., 2006), etc. (Oreifej and Liu, 2013; Osuna et al., 1997; Wang et al., 2009).

2.3. Deep learning

The concept of deep learning mainly operates in artificial neural networks (NN) – a hierarchical learning model (using iterative functionals formulated into multiple layers) developed a few decades ago (70s) (De Veaux and Ungar, 1997; Hampshire and Pearlmutter, 1991), where 'deep' refers to the number of layers being larger than a normal NN (less than 3–5 layers) (Goodfellow et al., 2016). Associated with the keywords 'deep' there are potentially a large amount of unknown parameters involved, demanding for very large amounts of training samples (Krizhevsky et al., 2012; Schmidhuber, 2015). The new development of neural networks begins in 2006 with Hinton's work in his idea of deep belief nets – a strategy of adding new layers of information while fixing parameters in previous layers (Hinton and Salakhutdinov, 2006). The use of deep NN later achieved state-of-the-art results in speech recognition (Hinton et al., 2012), traditionally worse than the state-of-the-art statistical classifiers.

The real impact of deep learning in the community started from 2012 on (Krizhevsky et al., 2012), when Hinton's groups won the Image Large-Scale Visual Recognition Challenges (LSVRC) (Berg et al., 2010; Russakovsky et al., 2015), achieving more than 10% improvement over the second best algorithm in the competition. Since then papers in the vision community on deep learning techniques in different applications are getting outdated very quickly and absorb a large part of the recent top tier computer vision conferences. The deep learning in the AI community was summarized by Andrew Beam and many others, as successful contribution to the modern engineering success

(Beam, 2017): (1) larger availability of high-quality labeled datasets for training (e.g. crowd-sourcing dataset), (2) much increased computing power; (3) solver-friendly node activation functions; (4) engineering techniques for robust optimization techniques (i.e. dropout, batch normalization, data augmentation) (Goodfellow et al., 2016). While the high performance does not necessarily indicate a superior solution in all matters, the deep learning models bring along two other major problems: (1) More complex networks need larger-than-ever volumes of training samples, thus raising the need for high-quality training samples in applications where crowd-sourcing label data are not available. (2) Training such a complex system requires trials of parameter settings and NN structure design that brings another dimension of hand-engineering, which so far has not shown less efforts than traditional hand-crafting processing of features. Efforts were suggested to address these problems: (1) Transfer learning allowing reuse of training samples from different domains, or fine-tuning of the existing networks, meaning using a few layers attached to an existing network (or at least part of it) for training with fewer samples (Jia et al., 2014; Larochelle, 2020; Tajbakhsh et al., 2016); (2) More robust regularization, optimization techniques, more efficient learning and NN architectures (Goodfellow et al., 2016).

Attempts in the photogrammetry and remote sensing community started closely following the success of the deep learning in the vision communities in some of the applications (Li et al., 2019; Zhang et al., 2016). However, there is the common issue of lacking samples: this is particularly problematic for the photogrammetry and remote sensing datasets, since the size of the community is relatively small in comparison to the CV community, with much fewer contributors developing standard training samples that are large enough for effective training of deep NN (although there exists some data sets for particular types of sensors (Gerke, 2014)). The photogrammetry and remote sensing datasets often capture the ground scene with similar perspectives (top-view). The great variety of data from different sensor types, like multi-head cameras, multi/hyper spectral images, LiDAR, altimeter and synthetic aperture radar (SAR), etc., leads to even poorer availability of corresponding training samples. Moreover, the variety and scales of objects in a small image patch vary from very small objects to large metropolitan buildings or landscape features, introducing more issues in the deep learning based recognition models (Cheng et al., 2017).

3. Existing efforts in photogrammetry and computer vision

Among the first applications of AI technology in photogrammetry were two projects executed within the Swiss National Research Program 23 (NRP23) ‘Artificial Intelligence and Robotics’. Scientific-technical goals of this nation-wide program were:

- Implementation and refinement of AI methods, especially in robotics
- Understanding of perceptions and learning processes; building bridges between AI and cognitive sciences and psychology

In this context we worked on two problems under the project ‘Design and Analysis of Spatial Image Sequences’: (a) using expert-system technology (‘low level AI’) to automate the sensor placement task in close-range (especially industrial) applications (Mason and Gruen, 1994, 1995; Mason and Këpuska, 1992) and (b) using Neural Network technology to automatically recognize signalized ground control points in aerial images (Këpuska and Mason, 1995; Mason and Gruen, 1994; Mason and Gruen, 1995; Mason and Këpuska, 1992). While project (a) led to quite successful solutions, project (b) did not return the desired results. After spending much time on network training the recognition performance was very instable.

In the meantime, of course, the AI methods have developed further substantially, and it is high time to look at these issues again.

The fields of photogrammetry and its industrial applications have been greatly impacted in the past decades by the development of computer vision, e.g. largely fully automated triangulation

(Fischler and Bolles, 1981; Snavely, 2010) and high performance per-pixel dense matching (Hirschmüller, 2005; Hirschmüller, 2008), etc., which traditionally rely on manual intervention in terms of blunder elimination of tie point observations and measurement of high-quality DSMs (Digital Surface Model). In particular, learning-based high resolution data interpretation, such as object extraction, action recognition and classification are attracting great attention (Forsyth and Ponce, 2002; Liu et al., 2020), along with the development of deep learning (Schmidhuber, 2015). It is expected that these traditionally manually intensive works in photogrammetry and remote sensing may profit considerably from these new techniques. Although AI for geometric processing (from acquiring images to obtaining standard orthophotos, point clouds and DSM products) in engineering practice is not yet widely used, the intention for better and more robust performance using learning-based methods were largely fueled by the CV community. In this section, we provide a non-inclusive and very brief overview on the development of computer learning on topics relevant to a typical photogrammetry and remote sensing processing pipeline (highlighted in bold thereafter).

3.1. Data acquisition

Data acquisition for photogrammetric purposes has changed dramatically in recent years. Satellite images are nowadays available at very high spatial and time resolution, coming with stereo or even triplet overlap and providing for many spectral channels. Aerial cameras are using the 5-head principle, collecting nadir and 4 oblique images simultaneously from every point of view. In close-range photogrammetry the old concept, in times of analogue imagery, was to generate as few images as possible, because image taking and measurement was very time-consuming and costly. Now we take rather too many pictures with the aim to get a very high overlap and redundancy in measurement, and may through away those which are not needed or of insufficient quality.

Aerial photogrammetric image acquisition is pretty standard and normally consists of an image block (or a few images only) with defined overlap to ensure the required theoretical accuracy and mapping resolution. In old times this is done fully manually or with a timekeeper (Eisenbeiß, 2009; Mikhail et al., 2001), often aided by the use of GPS (global positioning system) and potentially IMU (inertial measurement unit) systems (Colomina and Molina, 2014; Qin et al., 2013). The development of automatic camera shutter integrated with the GPS/IMU offers the capability for path planning and intelligent data acquisition, which largely boosts the use of the modern UAV-based (Unmanned Aerial Vehicle) photogrammetry by non-experts for smart civil and environmental applications (Boroujeni et al., 2012; Nex and Remondino, 2014; Pix4D, 2017).

However, in close-range applications like Cultural Heritage, underwater photogrammetry and many others there exist a great variety of different network designs. For occasions where sensors are not available to provide at least approximate exterior orientation parameters of the camera, e.g. indoor mapping with low-cost sensors, or 3D modeling of complex-shaped object with handheld cameras, a coarse model of the scene may be generated and then high-quality acquisition positions can be redefined to set up the waypoints.

Given the high-cost of satellite images, the acquisition is very important, considering the potential impact of high cloud coverage. Algorithms were developed to automatically detect cloud coverage (Champion, 2012; Coakley and Bretherton, 1982) and for steering the camera angle to regions with less clouds.

3.2. Geo-referencing

The most influential development in the past two decades in the CV community that impacts the geo-referencing tasks are probably the robust interest point extraction/matching (i.e. SIFT (Lowe, 2004), SURF (Bay et al., 2006), and many others) and the random sample consensus (RANSAC) techniques (Fischler and Bolles, 1981) for blunder elimination, thus enabling robust fully automated bundle adjustments (Deseilligny and Clery, 2011; Snavely, 2010). While the photogrammetry industry optimizes the robustness and accuracy of the geo-referencing tasks by optimizing individual steps of the workflow, the CV communities are interested in performing free-network adjustment in large, unordered image sets (acquired using unknown cameras and metadata, i.e. from internet photos) (Agarwal et al., 2011; Frahm et al., 2010), where efforts were devoted to the development of fast pair-wise graph matching algorithms (Wu, 2014), as well as intelligent algorithms for grouping images (Filliat, 2007) that might be geographically close (without an initial relative orientation process). Given the fact that most of the high-performance point extractors and matchers are computationally expensive, a learning-based method was also developed to predict the matchability (Hartmann et al., 2014) of image pairs without performing the actual matching to avoid exhaustive search. Moreover, such pose-estimation tasks have been attempted by deep learning methods, in which end-to-end networks directly take image pairs and regress their six orientation parameters (Dharmasiri et al., 2018; Kendall et al., 2015). Comparing to existing approaches, it is free from complex geometric computations and handling of camera parameters and feature matches. Although it has not achieved the state-of-the-art accuracy yet, it might have rooms for improvement.

3.3. Camera calibration

Camera calibration is regarded as a standard and well-established procedure before data acquisition or during the geo-referencing (also regarded as self-calibration). It requires either known object-space points or a cloud of well-distributed tie points, and the underlying principle for calibration is the use of a set of pre-defined calibration models (also called ‘additional parameters’) in bundle adjustment (Fraser, 2013; Gruen and Huang, 2013; Remondino and Fraser, 2006). In recent years there are a few works that consider the use of machine learning models (i.e. neural networks) to learn the non-linear relationship between the distorted and corrected 2D coordinate (Pedra et al., 2013), by taking direct point measurements, as well as ‘black-box’ approaches that take ‘perceptual of scenes’ as cues to directly predict distortion grids and intrinsic parameters of single images (Bogdan et al., 2018; Hold-Geoffroy et al., 2018). These novel attempts have demonstrated the possibility of using data to directly predict corrected/rectified images, which might implicitly take scene contents instead of rigorous geometric relationships. These methods are still far from be practical for actual photogrammetric productions due to lack of uncertainty measures and being scene specific.

3.4. DSM generation by image matching

The development of image-based per-pixel dense matching for DSM generation has been quite significant in the past decades and many techniques were developed both in the photogrammetry (Gruen, 2012) and computer vision community, e.g. multi-image and constraint-based matching (Zhang, 2005), dynamic programming (Veksler, 2005), semi-global matching (Hirschmüller, 2008), patch-based matching (Furukawa and Ponce, 2010), graph-cut (Vicente et al., 2008). Dense matching can be generally categorized based on the number of images used for computation: (1) stereo matching and (2) multi-view image matching. This leads to some fundamental algorithmic differences (Broadhurst et al., 2001; Furukawa et al., 2010; Furukawa and Ponce, 2009; Zhang, 2005): for example, stereo matching is able to utilize the rectified epipolar images for easier algorithm implementation (correspondences lying on the same row), while epipolar rectified images do not generally exist for more than two images. An obvious advantage of multi-view matching is that it is able to take redundant measurements to improve the robustness and accuracy of per-point matching (Zhang and Gruen, 2006). A few algorithms have shown great advances in utilizing the multiple observations such as multi-photo matching (Baltsavias, 1991), patch-based multi-view matching (Furukawa and Ponce, 2010), voxel-based space carving (Broadhurst et al., 2001). The state-of-the-art algorithms formulate the photo-consistency condition (computed from two or more images) within a global energy optimization framework. Points matched through multi-image matching provide the optimal accuracy, while in practice if the occluded pixels (both in multi-image and stereo matching) are not handled carefully, it may pose negative impact on neighboring pixels through the solver. For instance, the object space semi-global matching (SGM) algorithm (Bethmann and Luhmann, 2015) takes the average matching scores across multiple images, and turns the disparity computation procedure in a voxel object space. Since a mechanism determining the occluded pixels before averaging the scores is lacking, the results are not reported better than in the original algorithms. Given that stereo-based matching algorithms are particularly effective, practical implementation sometimes favors a multi-depth fusion approach (multi-stereo algorithms) (Seitz et al., 2006; Wenzel et al., 2013), where stereo matching are performed on permutated and selective pairs and then a depth/DSM fusion step utilizes the redundant information in the object space. This type of methods leaves the information fusion in the object space and can easily extend the state-of-the-art stereo matching algorithms to multi-view scenario. However, a theoretically more powerful concept is that of geometrically constrained multi-view matching. It allows to determine the matching parameters for all images involved plus the object space coordinates of the point in question in one simultaneous solution. By computation of the covariance matrix of all system unknowns one has an excellent tool for quality analysis of the matching process. Depending on the situation different constraints can be formulated: Epipolar-; collinearity-; X,Y-; Z-constraint.

Since dense matching (surface reconstruction) normally refers to the low-to-mid level vision problems, most of the development in the past decades focused on formulating the matching as global energy minimization problem (Boykov et al., 2001). Fairly recent developments (in the past few years) have shifted part of the focus to methods utilizing deep learning techniques in aid of such low-level vision problems. For example, Zbontar and Lecun (Zbontar and LeCun, 2016) used a Siamese network that learns the similarity scores through a two-channel convolutional neural network: the training takes texture patches of positive and negative matches from the benchmark dataset (i.e. KITTI and Middleburry) (Geiger et al., 2012; Scharstein and Szeliski, 2002; Scharstein and Szeliski, 2014). The top performers in these two benchmarks used similarity scores from the trained network, and some of these methods also learn the cost – propagation paths (Seki and Pollefeys, 2017). This shows that the hyper-parameterized deep learning models are able to accommodate the variations of the particular datasets, thus rendering better similarity scores leading to better performances in dense matching (Scharstein and Szeliski, 2014). There are attempts that aim to take the stereo matching problem as a per-pixel regression problems, which takes a stereo pair as an input and outputs a disparity map (Chang and Chen, 2018; Zhang et al., 2019), and they have shown better performances in benchmark datasets. However, the performance of these algorithms in practical applications is still questionable, as most of them were only applied to similar datasets where the training samples came from. Essentially, to make fair comparisons to classical methods, one needs to show that that these learning-based methods consistently outperform across different datasets, while not requiring new samples from those datasets for training. However, this probably involves another field of study related to the transferability of the networks (Celik et al., 2020; Yosinski et al., 2014).

3.5. Semantic interpretation

The interpretation/understanding and 3D modeling of a scene has a direct connection to machine intelligence. As already mentioned in Section 2.2 and 2.3, in the photogrammetry & remote sensing community this largely refers to image content classification (top-view or oblique) and reality/ semi-generic based 3D modeling. With the increasing exposure to deep learning the topic of interpreting aerial/satellite images are gaining great attentions in the Geo and CV communities, evidenced by recent worldwide learning challenges: (1) Kaggle satellite recognition challenge (Planet, 2017); (2) IARPA (Intelligence Advanced Research Projects Activity) functional mapping of the world challenge (IARPA, 2017) and (3) USSOCOM (United States Special Operations Command) urban 3D building detection challenge (USSOCOM, 2017), as well as the annual Earth-vision workshop (Tuia et al., 2015) hosted by both communities.

Land-cover classification of VHR images, where traditionally statistical learning methods were actively practiced, now use the concept of deep learning for boosting performances (Krizhevsky et al., 2012). Although the requirements of a large amount of samples were not addressed for VHR data classification, a compromised solution used by the researchers is to perform a fine-tuning using existing networks (Tajbakhsh et al., 2016), where the last (or the last a few) layers will be retrained with fewer samples on the aerial/satellite dataset. Another popular solution is to use the fully connected convolutional neural networks (Long et al., 2015). This usually still adopts the existing networks trained from ImageNets (Marmanis et al., 2016; Russakovsky et al., 2015; Sherrah, 2016) while adding a deconvolution layer for training using fully classified images as training. There are quite a few works adopting this approach to remote sensing data for building detection and classification (Bittner et al., 2017; Zhong et al., 2016). However, the required training samples are somewhat different from those in traditional landcover classification, as this requires a fully per-pixel labeled data as the input for classification. Although there is in general a lack of training datasets, the ISPRS website provides a benchmark based on the Vaihingen data with fully labeled references for training and testing (Blaha et al., 2016; Gerke, 2014; ISPRS, 2018). The semantic segmentation in CV (Liu et al., 2019a; McCormac et al., 2017; Wang et al., 2020), has also been intensively investigated, with the potential to be used in SLAM and self-driving systems for perception.

As one of the photogrammetric products, the point clouds generated from multi-view stereo/stereo or LiDAR scanning are sometimes considered as raw input for many applications such as classification and change detection (Hebel et al., 2013; Liu et al., 2019b; Teo and Shih, 2013). Although there have been many approaches for point clouds classification, obtaining high-quality classified point clouds in practice (e.g. for projects at city scale using data such as airborne or mobile LiDAR) is still a semi-automated approach that requires operator interactions with algorithms of choice, data specific parameter tuning, training data collection and direct point editing. Classification approaches for point clouds are very much in line with machine learning methods from early statistical classifiers such as SVM or random forest (RF) based point clouds classification (Li et al., 2016; Rau et al., 2014; Zhang et al., 2013), to nowadays deep learning based models (Chen et al., 2020; Qi et al., 2017a; Qi et al., 2017b; Shi and Rajkumar, 2020), in which PointNet (Qi et al., 2017a) was the pioneer work that explores structural information for segmentation with unstructured inputs. Although it is no longer the top-performing approach, it drives many other methods that bring the capacity of approaches to a notable level. Today, the benchmark datasets and open competitions (such as IEEE data fusion contest (Le Saux et al., 2019)) explicitly designed tasks with training data that match the needed training data volume for deep learning models (Hackel et al., 2017; Niemeyer et al., 2014; Tong et al., 2020), in which the traditional methods (shallow classifiers) are somewhat less competent. Nevertheless, it should be noted that in practical applications, the volume of available data for training is still of critical concern, which might be insufficient to drive deep learning models.

4. Conclusions

Due to changes in sensor technology and computing capabilities photogrammetric data acquisition and processing has changed fundamentally during the past four decades. Currently we witness the development of automated and self-learning (‘intelligent’) solutions in the field of computer vision, photogrammetry and remote sensing. This paper provides a very brief overview of the major developments relevant to the fields of photogrammetry and remote sensing. Our description is somewhat general and far from being inclusive as a technical review. Indeed, given such fast developments, it is impossible to cover the contents even with an extended manuscript. The take-away messages from this manuscript are mainly a brief skim on the relevant developments in different areas of photogrammetry and remote sensing. It is seen that as compared to other applications that having gained large success in AI (e.g. speech recognition has been successfully used in many AI devices), vision-based AI is relatively preliminary. Although the great breakthrough in using deep learning methods for object recognition has driven a large amount research investigations, its practical uses are limited by (1) processing 2D/3D signals and associated AI tasks are far more complicated, and (2) there is in general a lack of representative and large enough datasets for various photogrammetry & remote sensing based applications. Thus image-based AI in the field of photogrammetry is mostly still a topic for exploration. In particular, practical AI algorithms in remote sensing (i.e. land-cover classification), are still based on traditional statistical learning methods, with a success rate of 75–90% (e.g. overall accuracy) in good quality images with well-crafted training samples. Transferring the learning samples from other datasets to different applications remains to be challenging, but it is becoming more important in the training-data demanding deep learning.

The essence of deep learning is to utilize complex and non-linear models to approximate processes that are of complex nature. Some of the existing methods taking such a process as a 'black-box' mapping that sometimes ignore the nature of their rigorous physical models, will likely to turn these deep learning models to be very problem- and data- specific with no transparent mechanisms, which eventually weill yield non-trustworthy systems. For example, a recent work (Jin et al., 2020) comparing feature point/descriptor extraction methods in a very data baseline, shows that these 'black-box' models, claimed as 'best' in their respectively tested dataset, are not even as robust and as accurate as manually crafted features (Lowe, 2004) more than a decade ago when bringing them into practice. Therefore, it is worth to rethink, that when it comes to professional practice, a more organic use of these 'intelligent' models, for example, by only placing them to specific components that traditionally do not well, or to devise the 'black-box' models to be more transparent and analyzable. This is helpful to develop more trustworthy and intelligent systems for data processing and interpretation.

Although the AI based methods are not yet widely used in the geo-community, given the exponential growth of the machine learning fields and data science, and the need for developing trustworthy AI approaches, will likely drive useful products to be possibly used in practice in a few years. Apart from that, fully automated general image understanding remains an elusive problem for many years to come.

Acknowledgements

The work is partially supported by the Office of Naval Research (Award No. N000141712928). The authors appreciate the anonymous reviewers for their feedbacks which improved this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work is partially supported by the Office of Naval Research [Award No. N000141712928].

ORCID

Rongjun Qin http://orcid.org/0000-0002-9207-6103

References

Agarwal, S., Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. 2011. "Building Rome in a Day." *Communications of the ACM* 54 (10): 105–112.

Ahmed, A., M. Nagai, C. Tianen, and R. Shibasaki. 2008. "UAV Based Monitoring System and Object Detection Technique Development for a Disaster Area. International Archives of Photogrammetry." *Remote Sensing and Spatial Information Sciences* 37: 373–377.

Anders, K., L. Winiwarter, R. Lindenbergh, J. G. Williams, S. E. Vos, and B. Höfle. 2020. “4D Objects-by-Change: Spatiotemporal Segmentation of Geomorphic Surface Change from LiDAR Time Series.” *ISPRS Journal of Photogrammetry and Remote Sensing* 159: 352–363.

Anderson, J. A. 1995. *An Introduction to Neural Networks*. Cambridge: MIT press.

Baltsavias, E. P. 1991. “Multiphoto Geometrically Constrained Matching.” Ph.D Thesis, Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, ETH, Zürich.

Bay, H., T. Tuytelaars, and L. Van Gool. 2006. “Surf: Speeded Up Robust Features.” *Computer Vision–ECCV* 2006: 404–417.

Beam, A. 2017. “Deep Learning 101 – Part 1: History and Background.” Accessed 03.16.2018. https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html.

Berg, A., J. Deng, and L. Fei-Fei. 2010. “Large Scale Visual Recognition Challenge 2010”.

Bethmann, F., and T. Luhmann. 2015. “Semi-global Matching in Object Space. The International Archives of Photogrammetry.” *Remote Sensing and Spatial Information Sciences* 40 (3): 23.

Bianchini, M., and F. Scarselli. 2014. “On the Complexity of Neural Network Classifiers: A Comparison Between Shallow and Deep Architectures.” *IEEE Transactions on Neural Networks and Learning Systems* 25 (8): 1553–1565.

Bittner, K., S. Cui, and P. Reinartz. 2017. “Building Extraction from Remote Sensing Data Using Fully Convolutional Networks.” The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 42 481.

Blaha, M., C. Vogel, A. Richard, J. D. Wegner, T. Pock, and K. Schindler. 2016. “Large-scale Semantic 3d Reconstruction: An Adaptive Multi-Resolution Model for Multi-Class Volumetric Labeling.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3176–3184. https://ieeexplore.ieee.org/document/7780715.

Bogdan, O., V. Eckstein, F. Rameau, and J.-C. Bazin. 2018. “DeepCalib: a Deep Learning Approach for Automatic Intrinsic Calibration of Wide Field-of-View Cameras.” *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production* 1–10. https://dl.acm.org/doi/10.1145/3278471.3278479.

Bonnen, K., T. B. Czuba, J. A. Whritner, A. Kohn, A. C. Huk, and L. K. Cormack. 2020. “Binocular Viewing Geometry Shapes the Neural Representation of the Dynamic Three-Dimensional Environment.” *Nature Neuroscience* 23 (1): 113–121.

Boroujeni, N. S., S. A. Etemad, and A. Whitehead. 2012. “Computer and Robot Vision (CRV), 2012 Ninth Conference.” Robust Horizon Detection Using Segmentation for UAV Applications, 346–352.

Bouziani, M., K. Goïta, and D.-C. He. 2010. “Automatic Change Detection of Buildings in Urban Environment from Very High Spatial Resolution Images Using Existing Geodatabase and Prior Knowledge.” *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1): 143–153.

Boykov, Y., O. Veksler, and R. Zabih. 2001. “Fast Approximate Energy Minimization via Graph Cuts. Pattern Analysis and Machine Intelligence.” *IEEE Transactions* 23 (11): 1222–1239.

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32.

Broadhurst, A., T. W. Drummond, and R. Cipolla. 2001. “A Probabilistic Framework for Space Carving.” Proceedings of IEEE International Conference on Computer Vision, Vancouver, British Columbia, Canada, 7–14 July, 388–393.

Buchholtz, A. 1950. “Die Photogrammetrie in ihrer Anwendung auf nicht-topographischen Gebieten, Otto Lacmann.” In: La fotogrammetria nella sua applicazione in campi non-topografici. S. Hirzel Verlag, Leipzig (1950), XII+ 220 pag., 240 fig. e· 3 tavole. Prezzo 24. – DM. Elsevier.

Celik, Y., M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya. 2020. “Automated Invasive Ductal Carcinoma Detection Based Using Deep Transfer Learning With Whole-Slide Images.” Pattern Recognition Letters.

Champion, N. 2012. “Automatic Cloud Detection from Multi-Temporal Satellite Images: Towards the Use of Pléiades Time Series.” International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39 B3.

Chang, J.-R., and Y.-S. Chen. 2018. “Pyramid Stereo Matching Network.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5410–5418.

Chen, J., B. Lei, Q. Song, H. Ying, D. Z. Chen, and J. Wu. 2020. “A Hierarchical Graph Network for 3D Object Detection on Point Clouds.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual 392–401. https://openaccess.thecvf.com/content_CVPR_2020/papers/Chen_A_Hierarchical_Graph_Network_for_3D_Object_Detection_on_Point_CVPR_2020_paper.pdf.

Cheng, G., J. Han, and X. Lu. 2017. “Remote Sensing Image Scene Classification: Benchmark and State of the Art.” *Proceedings of the IEEE* 105 (10): 1865–1883.

Coakley, J. A., and F. P. Bretherton. 1982. “Cloud Cover from High-Resolution Scanner Data: Detecting and Allowing for Partially Filled Fields of View.” *Journal of Geophysical Research: Oceans* 87 (C7): 4917–4932.

Colomina, I., and P. Molina. 2014. “Unmanned Aerial Systems for Photogrammetry and Remote Sensing: A Review.” *ISPRS Journal of Photogrammetry and Remote Sensing* 92: 79–97.

Cornelis, N., B. Leibe, K. Cornelis, and L. Van Gool. 2008. “3d Urban Scene Modeling Integrating Recognition and Reconstruction.” *International Journal of Computer Vision* 78 (2–3): 121–141.

Costa, H., G. M. Foody, and D. S. Boyd. 2017. "Using Mixed Objects in the Training of Object-Based Image Classifications." *Remote Sensing of Environment* 190: 188–197.

Deli, R., L. M. Galantucci, A. Laino, R. D'Alessio, E. Di Gioia, C. Savastano, F. Lavecchia, and G. Percoco. 2013. "Three-dimensional Methodology for Photogrammetric Acquisition of the Soft Tissues of the Face: A New Clinical-Instrumental Protocol." *Progress in Orthodontics* 14 (1): 32.

Deseilligny, M. P., and I. Clery. 2011. "Apero, an Open Source Bundle Adjusment Software for Automatic Calibration and Orientation of Set of Images." Proceedings of ISPRS International Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures, Trento, Italy, 2–4 March, 269–276.

De Veaux, R. D., and L. H. Ungar. 1997. "A Brief Introduction To Neural Networks." Unpublished: http://www. cis. upenn. edu/∼ ungar/papers/nnet-intro. ps.

Dharmasiri, T., A. Spek, and T. Drummond. 2018. "Eng: End-to-end Neural Geometry for Robust Depth and Pose Estimation Using CNNs." Asian Conference on Computer Vision, Perth 625–642. https://arxiv.org/abs/1807.05705.

Diakité, A. A., G. Damiand, and D. Van Maercke. 2014. "Topological Reconstruction of Complex 3D Buildings and Automatic Extraction of Levels of Detail." *In: Eurographics Workshop on Urban Data Modelling and Visualisation*, 25–30.

Doxani, G., K. Karantzalos, and M. Tsakiri-Strati. 2010. "Automatic Change Detection in Urban Areas Under a Scale-Space, Object-Oriented Classification Framework." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII (Part4/C7), 7 p. (on CDROM).

Eisenbeiß, H. 2009. "UAV Photogrammetry." Ph.D Thesis, Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, Zürich.

Filliat, D. 2007. "A Visual Bag of Words Method for Interactive Qualitative Localization and Mapping." Robotics and Automation, 2007 IEEE International Conference, Rome, Italy.

Fischler, M. A., and R. C. Bolles. 1981. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Communications of the ACM* 24 (6): 381–395.

Foerstner, W. 1999. "3D-city Models: Automatic and Semi- Automatic Acquisition Methods." In *Proceedings of Photogrammetric Week' 99*, edited by Dieter Fritsch, 291–304. Heidelberg: Wichmann.

Foody, G. M., N. Campbell, N. Trodd, and T. Wood. 1992. "Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum-Likelihood Classification." *Photogrammetric Engineering and Remote Sensing* 58 (9): 1335–1341.

Förstner, W., and B. P. Wrobel. 2016. *Photogrammetric Computer Vision*, 816. New York: Springer International Publishing.

Forsyth, D. A., and J. Ponce. 2002. *Computer Vision: A Modern Approach*. New York: Prentice Hall Professional Technical Reference.

Frahm, J.-M., P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, and S. Lazebnik. 2010. "Building Rome on a Cloudless Day." European Conference on Computer Vision, Heraklion, Greece.

Fraser, C. S. 2013. "Automatic Camera Calibration in Close Range Photogrammetry." *Photogrammetric Engineering & Remote Sensing* 79 (4): 381–388.

Friedl, M. A., and C. E. Brodley. 1997. "Decision Tree Classification of Land Cover from Remotely Sensed Data." *Remote Sensing of Environment* 61 (3): 399–409.

Furukawa, Y., B. Curless, S. M. Seitz, and R. Szeliski. 2010. "Towards Internet-Scale Multi-View Stereo." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference, San Francisco, California.

Furukawa, Y., and J. Ponce. 2009. "Accurate Camera Calibration from Multi-View Stereo and Bundle Adjustment." *International Journal of Computer Vision* 84 (3): 257–268.

Furukawa, Y., and J. Ponce. 2010. "Accurate, Dense, and Robust Multiview Stereopsis. Pattern Analysis and Machine Intelligence." *IEEE Transactions* 32 (8): 1362–1376.

Galantucci, L., R. Ferrandes, and G. Percoco. 2006. "Digital Photogrammetry for Facial Recognition." *Journal of Computing and Information Science in Engineering* 6 (4): 390–396.

Geiger, A., P. Lenz, and R. Urtasun. 2012. "Are We Ready For Autonomous Driving? The Kitti Vision Benchmark Suite." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference, Providence, Rhode Island.

Gerke, M. 2014. Use of the Stair Vision Library within the ISPRS 2D Semantic Labeling Benchmark (Vaihingen). Web publication/site, ResearcheGate. doi:10.13140/2.1.5015.9683.

Gómez, C., J. C. White, and M. A. Wulder. 2016. "Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 116: 55–72.

Goncaluves, J. 2010. "3D Laser-Based Scene Change Detection for Basic Technical Characteristics/Design Information Verification." *ESARDA Bulletin* 45: 66–72.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep Learning*. Cambridge: MIT press Cambridge.

Gröger, G., T. Kolbe, and A. Czerwinski. 2007. Candidate OpenGIS® CityGML Implementation Specification (City Geography Markup Language). 119.

Gruen, Armin. 2012. "Development and Status of Image Matching in Photogrammetry." *The Photogrammetric Record* 27 (137): 36–57.

Gruen, A., M. Behnisch, and N. Kohler. 2009. "Perspectives in the Reality-Based Generation, n D Modelling, and Operation of Buildings and Building Stocks." *Building Research & Information* 37 (5–6): 503–519.

Gruen, A., and T. S. Huang. 2013. *Calibration and Orientation of Cameras in Computer Vision*. Berlin: Springer Science & Business Media.

Gruen, A., and X. Wang. 1998. "CC-Modeler: A Topology Generator for 3-D City Models." *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (5): 286–295.

Hackel, T., N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys. 2017. "Semantic3d. Net: A New Large-Scale Point Cloud Classification Benchmark." arXiv preprint arXiv:1704.03847.

Haleem, A., and M. Javaid. 2019. "3D Scanning Applications in Medical Field: A Literature-Based Review." *Clinical Epidemiology and Global Health* 7 (2): 199–210.

Hampshire, J. B., and B. Pearlmutter. 1991. "Equivalence Proofs for Multi-Layer Perceptron Classifiers and the Bayesian Discriminant Function." *Proceedings of the 1990 Summer School* 159–172. Elsevier. doi:10.1016/B978-1-4832-1448-1.50023-8.

Hartmann, W., M. Havlena, and K. Schindler. 2014. "Predicting Matchability." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA.

Hebel, M., M. Arens, and U. Stilla. 2013. "Change Detection in Urban Areas by Object-Based Analysis and On-the-Fly Comparison of Multi-View ALS Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 86: 52–64.

Hinton, G., L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. N. Sainath. 2012. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups." *IEEE Signal Processing Magazine* 29 (6): 82–97.

Hinton, G. E., and R. R. Salakhutdinov. 2006. "Reducing the Dimensionality of Data with Neural Networks." *Science* 313 (5786): 504–507.

Hirschmüller, H. 2005. "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information." IEEE computer Society Conference on Computer Vision and Pattern Recognition, 807–814.

Hirschmüller, H. 2008. "Stereo Processing by Semiglobal Matching and Mutual Information." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2): 328–341.

Hold-Geoffroy, Y., K. Sunkavalli, J. Eisenmann, M. Fisher, E. Gambaretto, S. Hadap, and J.-F. Lalonde. 2018. "A Perceptual Measure for Deep Single Image Camera Calibration." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA.

Huang, X., and L. Zhang. 2013. "An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 51 (1): 257–272.

IARPA. 2017. "Functional Map of the World Challenge." Accessed 03.16.2018. https://community.topcoder.com/longcontest/stats/?module=ViewOverview&rd=16996.

ISPRS. 2018. "ISPRS Semantic Labeling Benchmark Dataset." Last date accessed 18 April 2018. http://www2.isprs.org/commissions/comm3/wg4/3d-semantic-labeling.html.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning*. Orlando: Springer.

Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. 2014. "Caffe: Convolutional Architecture for Fast Feature Embedding." Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, USA.

Jin, Y., D. Mishkin, A. Mishchuk, J. Matas, P. Fua, K. M. Yi, and E. Trulls. 2020. "Image Matching across Wide Baselines: From Paper to Practice." arXiv preprint arXiv:2003.01587.

Justice, C., G. Gutman, and K. P. Vadrevu. 2015. "NASA Land Cover and Land Use Change (LCLUC): An Interdisciplinary Research Program." *Journal of Environmental Management* (148): 4–9.

Kendall, A., M. Grimes, and R. Cipolla. 2015. "Posenet: A Convolutional Network for Real-Time 6-dof Camera Relocalization." Proceedings of the IEEE International Conference on Computer Vision, Araucano Park, Las Condes, Chile.

Këpuska, V., and S. O. Mason. 1995. "A Neural Network Approach to Signalized Point Recognition in Aerial Photographs." *Photogrammetric Engineering & Remote Sensing* 61 (7): 917–925.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*, 1097–1105. location: Lake Tahoe, Nevada, USA.

Lafarge, F., X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny. 2008. "Automatic Building Extraction from DEMs Using an Object Approach and Application to the 3D-City Modeling." *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (3): 365–381.

Larochelle, H. 2020. "Few-Shot Learning. Computer Vision: A Reference Guide." 1–4.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–444.

Le Saux, B., N. Yokoya, R. Hansch, M. Brown, and G. Hager. 2019. "2019 Data Fusion Contest [Technical Committees]." *IEEE Geoscience and Remote Sensing Magazine* 7 (1): 103–105.

Li, S., W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson. 2019. "Deep Learning for Hyperspectral Image Classification: An Overview." *IEEE Transactions on Geoscience and Remote Sensing* 57 (9): 6690–6709.

Li, Z., L. Zhang, X. Tong, B. Du, Y. Wang, L. Zhang, Z. Zhang, H. Liu, J. Mei, and X. Xing. 2016. "A Three-Step Approach for TLS Point Cloud Classification." *IEEE Transactions on Geoscience and Remote Sensing* 54 (9): 5412–5424.

Liang, M., B. Yang, Y. Chen, R. Hu, and R. Urtasun. 2019. "Multi-task Multi-Sensor Fusion for 3d Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA.

Liebelt, J., and C. Schmid. 2010. "Multi-view Object Class Detection with a 3d Geometric Model." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference, San Francisco, California.

Lillesand, T., R. W. Kiefer, and J. Chipman. 2014. *Remote Sensing and Image Interpretation*. New York: John Wiley.

Liu, C., L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. L. Yuille, and L. Fei-Fei. 2019a. "Auto-deeplab: Hierarchical Neural Architecture Search for Semantic Image Segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA.

Liu, L., W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen. 2020. "Deep Learning for Generic Object Detection: A Survey." *International Journal of Computer Vision* 128 (2): 261–318.

Liu, W., J. Sun, W. Li, T. Hu, and P. Wang. 2019b. "Deep Learning on Point Clouds and Its Application: A Survey." *Sensors* 19 (19): 4188.

Long, J., E. Shelhamer, and T. Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Araucano Park, Las Condes, Chile.

Lowe, D. G. 2004. "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision* 60 (2): 91–110.

Lu, X., Z. Li, Z. Cui, M. R. Oswald, M. Pollefeys, and R. Qin. 2020. "Geometry-Aware Satellite-to-Ground Image Synthesis for Urban Areas." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Online.

Lu, D., P. Mausel, E. Brondizio, and E. Moran. 2004. "Change Detection Techniques." *International Journal of Remote Sensing* 25 (12): 2365–2401.

Lussu, P., and E. Marini. 2020. "Ultra Close-range Digital Photogrammetry in Skeletal Anthropology: A Systematic Review." *PloS one* 15 (4): e0230948.

Ma, L., M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu. 2017. "A Review of Supervised Object-based Land-Cover Image Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 277–293.

Marmanis, D., J. D. Wegner, S. Galliani, K. Schindler, M. Datcu, and U. Stilla. 2016. "Semantic Segmentation of Aerial Images with an Ensemble of CNNs. ISPRS Annals of the Photogrammetry", Remote Sensing and Spatial Information Sciences 3 473.

Mason, S. O., and A. Gruen. 1994. "Expert System Based Design of Sensor Configurations for Vision-based Inspection." NRP 23 Symposium on Artificial Intelligence and Robotics, 35–55.

Mason, S. O., and A. Gruen. 1995. "Automating the Sensor Placement Task for Accurate Dimensional Inspection." *Computer Vision and Image Understanding* 61 (3): 454–467.

Mason, S. O., and V. Këpuska. 1992. "CONSENS: An Expert System for Photogrammetric Network Design." *Allgemeine Vermessungs Nachrichten*, 384–393.

McCormac, J., A. Handa, A. Davison, and S. Leutenegger. 2017. "Semantic Fusion: Dense 3D Semantic Mapping with Convolutional Neural Networks." Robotics and Automation (ICRA), 2017 IEEE international Conference, Marina Bay Sands Singapore, Singapore.

Menard, S. 2018. *Applied Logistic Regression Analysis*. New York: SAGE.

Mikhail, E. M., J. S. Bethel, and J. C. McGlone. 2001. *Introduction to Modern Photogrammetry*. New York: Wiley.

Newman, L. H. 2016. "Ops Have a Database of 117M Faces." You're Probably in It. Accessed 04.17.2018. https://www.wired.com/2016/10/cops-database-117m-faces-youre-probably/.

Nex, F., and F. Remondino. 2014. "UAV for 3D Mapping Applications: A Review." *Applied Geomatics* 6 (1): 1–15.

Niemeyer, J., F. Rottensteiner, and U. Soergel. 2014. "Contextual Classification of Lidar Data and Building Object Detection in Urban Areas." *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 152–165.

Oreifej, O., and Z. Liu. 2013. "Hon4d: Histogram of Oriented 4d Normals for Activity Recognition from Depth Sequences." Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference, Portland, USA.

Osuna, E., R. Freund, and F. Girosit. 1997. "Training Support Vector Machines: An Application to Face Detection." Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference, 130–136.

Ozcanli, O., Y. Dong, J. Mundy, H. Webb, R. Hammoud, and V. Tom. 2016. "Automatic Geolocation Correction of Satellite Imagery." *International Journal of Computer Vision* 116 (3): 263–277.

Pal, M. 2005. "Random Forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing* 26 (1): 217–222.

Pal, M., and P. Mather. 2005. "Support Vector Machines for Classification in Remote Sensing." *International Journal of Remote Sensing* 26 (5): 1007–1011.

Pedra, A. V. B. M., M. Mendonça, M. A. F. Finocchio, L. V. R. de Arruda, and J. E. C. Castanho. 2013. "Camera Calibration Using Detection and Neural Networks." *IFAC Proceedings Volumes* 46 (7): 245–250.

Peng, D., Y. Zhang, and H. Guan. 2019. "End-to-end Change Detection for High Resolution Satellite Images Using Improved Unet++." *Remote Sensing* 11 (11): 1382.

Pix4D. 2017. Pix4D, http://pix4d.com/. Last date accessed June 09 2017.

Planet. 2017. "Understanding the Amazon from the Space." Accessed 03.16.2018. https://www.kaggle.com/c/planet-understanding-the-amazon-from-space.

Qi, C. R., H. Su, K. Mo, and L. J. Guibas. 2017a. "Pointnet: Deep Learning on Point Sets for 3d Classification and Segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA.

Qi, C. R., L. Yi, H. Su, and L. J. Guibas. 2017b. "Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space." *In: Advances in Neural Information Processing Systems* 5099–5108. location: Long Beach, USA.

Qin, R. 2014. "An Object-Based Hierarchical Method for Change Detection Using Unmanned Aerial Vehicle Images." *Remote Sensing* 6 (9): 7911–7932.

Qin, R. 2015a. "3D Change Detection in an Urban Environment with Multi-Temporal Data." Ph.D Thesis, Swiss Federal Institute of Technology, ETH, Zürich.

Qin, R. 2015b. "A Mean Shift Vector-Based Shape Feature for Classification of High Spatial Resolution Remotely Sensed Imagery." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (5): 1974–1985. doi:10.1109/JSTARS.2014.2357832.

Qin, R. 2019a. "A Critical Analysis of Satellite Stereo Pairs for Digital Surface Model Generation and a Matching Quality Prediction Model." *ISPRS Journal of Photogrammetry and Remote Sensing* 154: 139–150.

Qin, R. 2019b. An Operational Pipeline for Generating Digital Surface Models from Multi-Stereo Satellite Images for Remote Sensing Applications." IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 5960–5963

Qin, R., and A. Gruen. 2014. "3D Change Detection at Street Level Using Mobile Laser Scanning Point Clouds and Terrestrial Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 90: 23–35.

Qin, R., A. Grün, and X. Huang. 2013. "UAV Project – Building a Reality-Based 3D Model." *Coordinates* 9: 18–26.

Qin, R., X. Huang, A. Gruen, and G. Schmitt. 2015. "Object-Based 3-D Building Change Detection on Multitemporal Stereo Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (8): 2125–2137. 10.1109/JSTARS.2015.2424275.

Qin, R., J. Tian, and P. Reinartz. 2016. "3D Change Detection – Approaches and Applications." *ISPRS Journal of Photogrammetry and Remote Sensing* 122: 41–56.

Rathinam, S., Z. W. Kim, and R. Sengupta. 2008. "Vision-based Monitoring of Locally Linear Structures Using an Unmanned Aerial Vehicle." *Journal of Infrastructure Systems* 14 (1): 52–63.

Rau, J.-Y., J.-P. Jhan, and Y.-C. Hsu. 2014. "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): 1304–1319.

Remondino, F., and C. Fraser. 2006. "Digital Camera Calibration Methods: Considerations and Comparisons. International Archives of Photogrammetry." *Remote Sensing and Spatial Information Sciences* 36 (5): 266–272.

Robinson, C., L. Hou, K. Malkin, R. Soobitsky, J. Czawlytko, B. Dilkina, and N. Jojic. 2019. "Large Scale High-Resolution Land Cover Mapping with Multi-Resolution Data." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein. 2015. "Imagenet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision* 115 (3): 211–252.

Scharstein, D., and R. Szeliski. 2002. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms." *International Journal of Computer Vision* 47 (1–3): 7–42.

Scharstein, D., and R. Szeliski. 2014. "Middlebury Stereo Vision Page." Accessed 3 December, 2014. http://vision.middlebury.edu/stereo/.

Schmidhuber, J. 2015. "Deep Learning in Neural Networks: An Overview." *Neural Networks* 61: 85–117.

Seitz, S. M., B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. 2006. "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference, New York, USA.

Seki, A., and M. Pollefeys. 2017. "SGM-nets: Semi-Global Matching with Neural Networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA.

Shan, J., Z. Hu, P. Tao, L. Wang, S. Zhang, and S. Ji. 2020. "Toward a Unified Theoretical Framework for Photogrammetry." *Geo-spatial Information Science* 23 (1): 75–86.

Sherrah, J. 2016. "Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery." arXiv preprint arXiv:1606.02585.

Shi, W., and R. Rajkumar. 2020. "Point-GNN: Graph Neural Network for 3d Object Detection in a Point Cloud." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online.

Snavely, N. 2010. "Bundler: Structure from Motion (SFM) for Unordered Image Collections." Available online: phototour. cs. washington. edu/bundler/.

Stern.de. 2018. "Gesichtserkennung durch Sonnenbrille: Chinas Polizei auf futuristischer Verbrecherjagd." In: Stern.de of 11. February 2018.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2016. "Rethinking the Inception Architecture for Computer Vision." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA.

Szeliski, R. 2010. “Computer Vision: Algorithms and Applications.” Springer Science & Business Media.

Tajbakhsh, N., J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. 2016. “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?” *IEEE Transactions on Medical Imaging* 35 (5): 1299–1312.

Teo, T.-A., and T.-Y. Shih. 2013. “Lidar-based Change Detection and Change-Type Determination in Urban Areas.” *International Journal of Remote Sensing* 34 (3): 968–981.

Tian, J. 2013. “3D Change Detection from High and Very High Resolution Satellite Stereo Imagery.” Ph.D Thesis, Institute for Geoinformatics and Remote Sensing (IGF) University of Osnabrück.

Tian, J., H. Chaabouni-Chouayakh, P. Reinartz, T. Krauß, and P. d’Angelo. 2010. “Automatic 3D Change Detection Based on Optical Satellite Stereo Imagery. International Archives of Photogrammetry.” *Remote Sensing and Spatial Information Sciences* 38 (Part 7B): 586–591.

Tong, G., Y. Li, D. Chen, Q. Sun, W. Cao, and G. Xiang. 2020. “CSPC-Dataset: New LiDAR Point Cloud Dataset and Benchmark for Large-Scale Scene Semantic Segmentation.” *IEEE Access* 8: 87695–87718.

Treible, W., S. Sorensen, A. D. Gilliam, C. Kambhamettu, and J. L. Mundy. 2018. “Learning Dense Stereo Matching for Digital Surface Models from Satellite Imagery.” arXiv preprint arXiv:1811.03535.

Tuia, D., J. D. Wegner, K. Schindler, J. Zerubia, and G. Moser. 2015. “Report on the IEEE GRSS/ISPRS Workshop EarthVision@ CVPR 2015 (Boston, MA).” IEEE-INST Electrical Electronics Engineers INC 445 Hoes Lane, Piscataway, NJ 08855-4141 USA.

USSOCOM. 2017. “Urban 3D Challenge.” Accessed 03.16.2018. https://wwwtc.wpengine.com/urban3d.

Veksler, O. 2005. “Stereo Correspondence by Dynamic Programming on a Tree.” IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California, USA, June 20–26, 384–390.

Verdie, Y., F. Lafarge, and P. Alliez. 2015. "Lod Generation for Urban Scenes." *ACM Transactions on Graphics* 34 (3): 15.

Vicente, S., V. Kolmogorov, and C. Rother. 2008. “Graph Cut Based Image Segmentation with Connectivity Priors.” Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, 24–26 June, 1–8.

Wang, L. 2005. *Support Vector Machines: Theory and Applications*, 431. New York: Springer.

Wang, X., T. X. Han, and S. Yan. 2009. “An HOG-LBP Human Detector with Partial Occlusion Handling.” Proceedings of IEEE International Conference on Computer Vision, Miami, Florida, USA, 20–25 June, 32–39.

Wang, X., S. Liu, H. Ma, and M.-H. Yang. 2020. “Weakly-Supervised Semantic Segmentation by Iterative Affinity Learning.” *International Journal of Computer Vision* 128 (2020): 1736–1749.

Wenzel, K., M. Rothermel, and D. Fritsch. 2013. “SURE–The IFP Software for Dense Image Matching.” Photogrammetric Week 13, 59–70.

Wijaya, A., R. S. Budiharto, A. Tosiani, D. Murdiyarso, and L. Verchot. 2015. “Assessment of Large Scale Land Cover Change Classifications and Drivers of Deforestation in Indonesia. The International Archives of Photogrammetry.” *Remote Sensing and Spatial Information Sciences* 40 (7): 557.

Wu, C. 2014. “VisualSFM: A Visual Structure from Motion System.” Last date accessed 26 Feb 2014. http://ccwu.me/vsfm/.

xenonstack. 2020. “Automatic Log Analysis using Deep Learning and AI.” Accessed 31 July 2020. https://www.xenonstack.com/blog/log-analytics-deep-machine-learning/.

Yosinski, J., J. Clune, Y. Bengio, and H. Lipson. 2014. “How Transferable are Features in Deep Neural Networks?” In: *Advances in Neural Information Processing Systems*, Montreal, Quebec, Canada, 3320–3328.

Zbontar, J., and Y. LeCun. 2016. “Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches.” *Journal of Machine Learning Research* 17 (1-32): 2.

Zeit, D. 2020. “There Are No Obstacles in the Air.” Newspaper Interview in German, DIE ZEIT, No. 16, 8 April, 35–36.

Zhang, L. 2005. “Automatic Digital Surface Model (DSM) Generation from Linear Array Images.” Ph. D Thesis, Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, Zürich.

Zhang, L., and A. Gruen. 2006. “Multi-image Matching for DSM Generation from IKONOS Imagery.” *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (3): 195–211.

Zhang, J., X. Lin, and X. Ning. 2013. “SVM-based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas.” *Remote Sensing* 5 (8): 3749–3775.

Zhang, F., V. Prisacariu, R. Yang, and P. H. Torr. 2019. “Ga-net: Guided Aggregation Net for End-to-end Stereo Matching.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA.

Zhang, L., L. Zhang, and B. Du. 2016. “Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art.” *IEEE Geoscience and Remote Sensing Magazine* 4 (2): 22–40.

Zhong, Z., J. Li, W. Cui, and H. Jiang. 2016. “Fully Convolutional Networks for Building and Road Extraction: Preliminary Results.” Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, Beijing, China.