MULTISPECTRAL AND MOBILE MAPPING ISPRS WG III/5 DATA SET: FIRST ANALYSIS OF THE DATASET IMPACT

M. Previtali *, M. Garramone, M. Scaioni

1 Department of Architecture, Built Environment and Construction Engineering (DABC), Politecnico di Milano, via Ponzio 31
Milano, Italy - (mattia.previtali, manuel.garramone, marco.scaioni)@polimi.it

Commission III, WG III/5

KEY WORDS: ISPRS WG III/5 Data Set, Laser intensity, LiDAR, Mobile Laser Scanning, Multispectral

ABSTRACT:

Light Detection and Ranging (LiDAR) technology is playing a major role in different applications. Indeed, the possibility of exploiting either 3D geometric features and radiometric information makes LiDAR suitable for a wide range of practical domains. LiDAR proved also being quite flexible in terms of platforms where it can be implemented for the acquisition, spanning from airborne sensors up to car-based and hand-held instruments. Due to the rapid expansion of research concerning LiDAR intensity, the ISPRS WG III/5 launched in 2017 an initiative aimed at providing free access to LiDAR data acquired by modern multispectral ALS sensors as well as MLS data. The two datasets, MLS Data Set 1 – (“Sun Prairie”), and Multispectral LiDAR Data Set 2 – (“Tobermory”), were provided by Teledyne Optech Company (Canada) and were made freely available to researchers upon request. This paper is presenting the first results of this initiative in terms of applications, application domains and topics tackled by applicants. The relevance of this data set is also evaluated through a bibliometric analysis considering both Scopus and Web of Science indexed databases to analyse the main directions where the scientific research, the technical development and the application interest is moving to.

1. INTRODUCTION

Light Detection and Ranging (LiDAR) technology nowadays plays a relevant role into different application fields ranging from Cultural Heritage and archaeology (Shanoer and Abed, 2018), forestry and vegetation (You et al., 2017), geosciences (Jaboyedoff et al., 2012), topographic mapping (Shan and Toth, 2009), and infrastructures and asset management (Neupane and Gharaiibe, 2019; Jung et al, 2019). In addition, it has proved to be an important tool supporting the development and implementation of policies and strategies finalized to improve the quality of life, especially in urban areas, and to cope with the global societal challenges (climate change, reduction of carbon emission, social inclusion, sustainability – see, e.g., Speak et al., 2019; Gülçin et al., 2021).

Indeed, the potential of LiDAR technology is not only connected to the possibility of reconstructing digital 3-D representations by sampling at high acquisition rate and spatial resolution the topographic surface, the ocean seabed and other objects, but also in the possibility of using laser intensity return data as a key element for segmentation and classification purposes (Yan et al., 2012; Morsy et al., 2017; Hänsch and Hellwich, 2020).

As reported in Scioni et al. (2018), in the last decade several researchers investigated the exploitation of LiDAR intensity in many applications, often in conjunction with geometrical features. Indeed, object extraction using LiDAR data, acquired from different laser scanning platforms (airborne – ALS –, terrestrial – TLS –, mobile – MLS –, Unmanned Aerial Sistems/Vehicles – UAS/UAV, portable and hand-held) in urban environment and forest application is an active research field. Several scholars are using either new tools for data extraction (i.e., Artificial Intelligence – AI), new sensors characterized by both higher spatial and spectral resolution and new types of information (i.e., geometrical features in conjunction with laser intensity and RGB images). In particular, the potential of exploiting LiDAR intensity in classification tasks has motivated the development of multispectral ALS sensors that may be used to acquire information at different wavelengths, enabling this way applications such as land cover classification, shallow water bathymetry, and forestry analysis, among others (see Scioni et al., 2018).

Another successful application field in the recent years has been related to the use of MLS data for both mapping and 3D city modelling. In this domain, research is in general less focused on the acquisition of multispectral information, while it is more dedicated to the integration of point clouds and RGB data for photorealistic virtual/augmented-reality applications and classification purposes. Due to the rapid expansion of the above-mentioned research fields, the availability of a reference data set is of major importance to provide some free data to carry out research and for comparing different approaches on the same and/or similar problems. In addition, the availability of a reference data set can be used to highlight the most relevant topics and application fields associated to LiDAR data analysis and processing. Even if some ISPRS initiatives and Working Groups have been already providing LiDAR benchmarking data sets for specific tasks, e.g., semantic segmentation of 3D point clouds (Kölle et al., 2021), data classification (Niemeyer et al., 2014), the aim of the ISPRS WG III/5 ‘Information Extraction from LiDAR data’ initiative was to provide a data set acquired by modern multispectral ALS sensors as well as MLS sensors with increased spectral and radiometric resolution. The data set was launched in 2017, see Section 2. The aim of this paper is to present the first results achieved by different scholars using this data set in terms of published papers, application fields and main topics tackled. In Sections 3 and 4 we report some results of a survey among the applicants that was carried out to identify the

* Corresponding author
main requirements from scholars and to plan future activities connected with the data set (e.g., provide ground truth, etc.). In addition, in order to analyse those directions where the scientific research, the technical progress and the application interest is moving to, in Section 5 we reported those findings from a bibliometric analysis related to the application of laser intensity and multispectral LiDAR data. We focused on these two topics because we retained their maturity not completely achieved yet, while MLS may be considered as a more consolidated research field.

The remaining of the paper is organized as follows: Section 2 provides an overview of ISPRS WG III/5 Data Set; Section 3 describes the methodology used for the analysis of the Data Set impact; Section 4 presents the results of the analyses; and eventually Section 6 presents some final discussion.

2. ISPRS WG III/5 DATA SET

In 2017 the ISPRS WG III/5, in collaboration with Teledyne Optech, released a new data set on “Information Extraction from LiDAR Intensity Data: Multi-Spectral and Mobile LiDAR data” (Scaioni et al., 2018). The data set can be requested by the research community via the WG website (www2.isprs.org/commissions/comm3/wg5.html) by signing a form accepting the use of the data for research purpose only. Scholars are given access to the sensor data and are encouraged to share their results with the WG and the community. In particular, at the moment two data sets are provided within this initiative:

1. MLS Data Set 1 – (“Sun Prairie”); and
2. Multispectral LiDAR Data Set 2 – (“Tobermory”).

In the following subsections, both data sets are briefly described.

2.1. MLS Data Set 1

MLS Data Set 1 covers an urban environment located in Sun Prairie (Wisconsin, US). The data set was acquired by using the dual-head sensor MLS system Optech Lynx SG moving along a couple of roads in Sun Prairie city for a total distance of approx. 2 km.

A measurement rate of 1,200 kHz (600 kHz per sensor) and a scanning rate of 250 Hz per each sensor (total of 500 lines/second) have been used during data acquisition. Technical specifications of Optech Lynx SG are presented in Table 1. The data set contains a typical urban environment with several object classes: buildings, roads, trees, cars, pedestrians, poles, etc. This makes the data set suitable for urban classification, 3D city modelling and building detection. Severe occlusions due to cars and trees are present in building facades.

The sample Data Set 1 consists of three strips collected using the two sensors (S1 and S2), stored in 6 LAS files for a total size of 4.5 GB.

2.2. Multispectral LiDAR Data Set 2

Multispectral LiDAR Data Set 2 covers a natural coastal area located in Tobermory (Ontario, Canada). Optech Titan multispectral ALS sensor was used for data acquisition on April 2015 at a flying height of approx. 460 m above ground and a speed of 140 knots. The Optech Titan sensor has 3 channels (1550 nm, 1064 nm and 532 nm), each simultaneously collecting data that are recorded either as discrete values and as full-waveform data. The sensor FoV is 40°. In Table 2 the main technical properties of Optech Titan are reported.

| Parameter | Specification |
|-----------|--------------|
| Number of lidar sensors | 1 |
| Laser: LiDAR | 945 nm |
| Maximum range | 300 m |
| Absolute accuracy | 5 mm, 0.1 |
| Laser measurement rate | 75-600 kHz programmable |
| Measurements per laser pulse | up to 4 simultaneous |
| Scan frequency | up to 300 lines/sec programmable |
| Power requirements | 12 VDC, 30 A max. |
| Operating temperature | -10°C to +40°C (extended range available) |
| Storage temperature | -40°C to +60°C |
| Relative humidity | 0-95% non-condensing |
| Laser classification | ECID/CDM class 1 safe |
| Vehicle | Fully adaptable to any vehicle |

Table 1. Technical specifications of the dual-head sensor Mobile Laser Scanning (MLS) system Optech Lynx SG adopted for the acquisition of MLS Data Set 1 (source: www.teledyneoptech.com).

Table 2. Technical specifications of the Optech Titan Airborne Laser Scanning (ALS) system adopted for the acquisition of Multispectral LiDAR Data Set 2 (source: www.teledyneoptech.com).

Independent normalized LiDAR intensity images (see Fig. 1) can be generated for each channel allowing high density topographic mapping, shallow water bathymetry, vegetation mapping or even 3D land classification.

Multispectral LiDAR Data Set 2 contains a couple of small harbours, rock coastline and water depths making it suitable for bathymetry. The area covered different vegetated areas and an alternation of zones characterized by both moderate-density urbanized areas and sparse houses. The Data Set consists of 11 strips over an area of approximately 10 km x 2 km at a point density of approx. 12 points/m². It is provided upon request into three main archives (one per each spectral channel) stored in LAS files for a total size of 26.4 GB.
3. METHODOLOGY

In this section we would like to analyse the impact of the WG III/5 Data Sets considering the development of applications per each of them (Data Sets 1 and 2) and the number of published scientific papers. On the other hand, due to the still limited number of scholars who have already completed their research on the basis of our Data Sets, we investigated those research fields where LiDAR intensity and multispectral LiDAR have been applied. This second investigation has been accomplished by means of a bibliometric analysis based on the identification of some relevant keywords and their search in two popular citation databases (Scopus and Web of Science).

In the following subsections the methodology adopted for the analysis of the applications developed by the applicants to the WG III/5 Data Sets (Subsect. 3.1) and the extended bibliometric analysis (Subsect. 3.2) are reported.

3.1 Analysis of the developed applications

In a first step we analysed the applications developed on the basis of each Data Set. The following items have been considered:

1. Geographic distribution of the applicants per each Data Set;
2. Requested Data Sets;
3. Academic position of the applicants;
4. Main application field in which the Data Sets were used (e.g., forestry, city modelling, etc.); and
5. Specific data application (e.g., road inventory, extraction of buildings, etc.)

In the second step of the research a questionnaire was sent to all the applicants to identify:

1. Usefulness of the Data Set in the applicants’ research and which specific feature was considered as more relevant and innovative in the Data Set;
2. Main issues with the Data Set (e.g., data accessibility, handling data volume, etc.);
3. Scientific production; and

4. Requirements and suggestions to improve the Data Set.

3.2 Bibliometric Analysis

The bibliometric analysis consisted in looking for publications focused on some specific topics, in order to carry out a comprehensive science mapping analysis. Different tools have been developed to this purpose (e.g., Gagolewski, 2011; van Eck and Waltman, 2014; Aria and Cuccurullo, 2017).

The first step of the bibliometric analysis, completed in April 2021, concerned the definition of the research keywords (Fig. 2). The research was conducted both in Scopus (2021) and Web of Science (WoS – Clarivate Analytics, 2021) in order to compare the results from different databases. A set of 388 documents from Scopus and 309 from Web of Science was found.

The filtering phase by categories and subject areas allowed to exclude results not related to the aim of the research (e.g., arts, humanities, medicine, etc.). At the end of this step, the final number of references was 370 (Scopus) and 285 (WoS). The bibliometric analyses were carried out, using the R package Bibliometrix (Aria and Cuccurullo, 2017), version 3.0. Findings will be presented in Subsection 4.3.

4. ANALYSIS OF THE RESULTS

4.1 Results from ISPRS WG III/5 Data Sets applications

Starting from the launch of the initiative in 2017, 28 applicants requested the ISPRS WG III/5 Data Sets.

Applicants have mainly requested both Data Sets (see Fig. 3) even if there is a specific tendency of higher interest in the Multispectral LiDAR Data Set 2 (“Tobermory”), demonstrating the relevance of the multispectral LiDAR topic in the scientific production.

Figure 1. The three wavelengths (represented by green, yellow and red vertical bars in the upper figure) of Optech Titan compared with spectral signature of different objects and normalized LiDAR intensity images for the three channels (at the bottom) – source: www.teledyneoptech.com.

Figure 2. Research keywords used in Scopus (“Article Title, Abstract, Keywords” fields) and in Web of Science (“Topic” field).

Figure 3. Requested Data Sets.
The majority of the applicants belongs to the Middle East and Far East institutions (see Fig. 4) with a specific predominance of People's Republic of China affiliations (46% of the applications).

![GEOGRAPHIC DISTRIBUTION](image1)

**Figure 4.** Geographic distribution of the applicants of both WG III/5 Data Sets.

The majority of the applicants (86%) are young researchers (MSc and PhD students and research assistants) as it can be observed in Figure 5.

![ACADEMIC POSITION](image2)

**Figure 5.** Academic position of the applicants.

Concerning the applicant’s field of research (see Fig. 6) it is worth to notice that researchers dealing with DTM extraction and forestry are mainly requesting the Multispectral LiDAR Data Set 2. On the other hand, applicants focusing on mobile mapping requested only the MLS Data Set 1 or requested both. Researchers dealing with city mapping are generally requesting both Data Sets highlighting the suitability of both of them to cover this topic.

![APPLICANT FIELD OF RESEARCH](image3)

**Figure 6.** Research field of the applicants.

The last aspect that has been analysed consisted in the specific research topic afforded on the basis of WG III/5 Data Sets (see Fig. 7). It can be noticed that few studies are addressing issues connected with radiometric calibration of the instrument and data filtering of the LiDAR data source. Research activities are mainly involving classification issues either of the entire point cloud or by identifying specific elements (e.g., buildings, roads, or urban objects as light poles and power lines). Mapping purposes and DTM/DSM generation are secondary topics.

![SPECIFIC RESEARCH USING THE WG III/5 DATASET](image4)

**Figure 7.** Specific research topic investigated on the basis of the ISPRS WG III/5 Data Sets.

### 4.2 Results of the questionnaire

The questionnaire was aimed at collecting feedbacks from the applicants either in terms of comments and suggestions on the Data Sets (structure of the data, clearness of the provided material, etc.) and to check the consistency of the scientific production with respect to the research purposes declared during the application step.
The collected feedbacks are showing that the majority of the applicants are mainly exploiting the multispectral “Tobermory” data as most innovative and element in the dataset. This is mainly connected with vegetation and forestry study as well as DTM/DSM production. Authors are not identifying specific issues with the management of the dataset in terms of data volume and size. One of the concerns is addressing the proper way of understanding the management of the multispectral LiDAR data, that currently are delivered in three separate files and three separate point clouds. The results obtained with the requested ISPRS WG III/5 data Sets were published mainly in scientific journals and conference proceedings. However, two MSc dissertations made use of the provided Data Set. Currently the dataset is provided as not-classified data. The main requirement and suggestion concern the availability of ground-truth data for classification purposes both for the “Sun Prairie” and the “Tobermory” Data Sets.

4.3 Results from the bibliometric analysis

The annual scientific production from 1981 to 2020, represented in Figure 8, shows an increasing interest in the analysed topics. The annual growth rate is 14.78% (Scopus) and 17.30% (WoS). The annual growth rate excludes the year 2021 because it is not yet completed at the time of writing. Since the first articles were published in the Eighties, a significant increase is registered from 2007 and a second peak can be seen in 2016, when a number of approx. 70 references were published, considering the sum of both databases. In order to better understand the main themes covered by these documents, a further analysis was carried out by filtering the results through five sub-keywords:

- Segmentation;
- Classification;
- Feature extraction;
- Vegetation; and
- Forestry.

The results were plotted on Venn diagrams, as shown below in Figure 9 for Scopus and Web of Science databases, respectively. These graphs allow to identify if a link between sub-keywords exists and its strength. Diagrams demonstrate that there is a strong connection between the five sub-topics. In particular, Web of Science database contains references with a maximum of three sub-keywords at the same time, while Scopus database provides a small number of articles with all the sub-keywords.

Figure 8. Published scientific literature according to those keywords’ combinations in Figure 2 (1981-2020).

Figure 9. Venn diagram (Scopus on the left, Web of Science on the right). Charts realized on meta-chart.com.
Figure 10 shows the social structure analysis, considering the countries’ collaboration network (number of nodes of 25 and minimum edge of 2). The two social structures obtained are quite similar. In fact, a central cluster is present, led by China, the USA, and Canada, and three smaller clusters complete the networks.

![Collaboration network by countries](image)

**Figure 10.** Collaboration network by countries (Scopus at the top, WoS at the bottom).

### 5. CONCLUSIONS

The paper would like to give an overview on the exploitation of two Data Sets provided by the ISPRS WG III/5 on ‘Information Extraction from LiDAR Intensity Data’. One data set consists of airborne multispectral LiDAR data, while the other concerns Mobile Laser Scanning (MLS) data. On one side, feedbacks received by the applicants to both Data Sets have been analysed with the aim of identifying the main research interests, the application fields, and the geographical distribution. Requirements from scholars have been also to plan the future activities for a further enhancement of the Data Sets themselves.

Since the number of outputs was quite limited, we extended the analysis to cover the overall scientific production on these topics. Even if the highest growth rate in the published scientific literature was showed in the period 2014-2015 the topic is still relevant and most probably would continue in the upcoming years (the scientific production has recently achieved a ‘plateau’ in the latest years). As proved by either the bibliometric research and the analysis of the applicants feedbacks, the most popular research topics connected with LiDAR intensity address the classification and the segmentation tasks with a specific focus on forestry applications. The strong connection among feature extraction, segmentation and classification is proved by the bibliometric analysis showing the high interdependency existing among those topics. Recently advances in Artificial Intelligence and its various branches, i.e., machine learning, deep learning, reinforcement learning, etc., are revolutionizing the way classification, segmentation and feature extraction are addressed. The bibliometric research carried out in this paper is not taking into consideration specific technologies. However, the relevance and the number of scientific production involving Artificial Intelligence and point cloud is significantly increasing in the last few years (see also Zhang et al., 2019). In this context the long-wave generated by Artificial Intelligence may keep the attention on LiDAR data classification and segmentation higher in the next years.

In this scenario the relevance of the ISPRS WG III/5 dataset for researchers dealing with LiDAR classification seems high also in the next years. Indeed, the multispectral nature of the data, the variety of object classes included in the surveyed areas and the different typology of sensors used for the acquisition, represent important characteristics of the dataset. However, the impact of the dataset can be increased strengthening connections and synergies with other similar initiatives by enlarging the number and the variety of available free data. In addition, as highlighted by applicants the availability of a ground-truth seems a key requirement to increase the usefulness of the dataset.

**Acknowledgements**

Acknowledgements goes to Teledyne Optech Company (Canada) for providing both ISPRS WG III/5 Data Sets data sets.

**REFERENCES**

Aria, M., Cucurullo, C., 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. doi: 10.1016/j.joi.2017.08.007.

Clarivate Analytics, 2021. Web of Science. https://apps.webofknowledge.com (last access on 25 April 2021).

Elsevier, 2021. Scopus. https://www.scopus.com/search/form.uri?display=basic#basic (last access on 25 April 2021).

Gagolewski, M., 2011. Bibliometric impact assessment with R and the CITAN package. *Journal of Informetrics*, 5(4), 678-692. doi: 10.1016/j.joi.2011.06.006.

Güçin, D., van den Bosch, C. C. K., 2021. Assessment of Above-Ground Carbon Storage by Urban Trees Using LiDAR Data: The Case of a University Campus. *Fores*, 12(1), 62.

Hänsch, R., Hellwich, O., 2020. Fusion of Multispectral LiDAR, Hyperspectral, and RGB Data for Urban Land Cover Classification. *IEEE Geoscience and Remote Sensing Letters*.

Jaboyedoﬀ, M., Oppikofer, T., Abellán, A., Derron, M. H., Loye, A., Metzger, R., Pedrazzini, A., 2012. Use of LiDAR in landslide investigations: A review. *Natural Hazards*, 61, 1–24.

Jung, J., Che, E., Olsen, M. J., Parrish, C., 2019. Efﬁcient and robust lane marking extraction from mobile lidar point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 1–18.

Kölle, M., Laupheimer, D., Schmohl, S., Haala, N., Rottensteiner, F., Wegner, J. D., Ledoux, H., 2021. H3D: Benchmark on Semantic Segmentation of High-Resolution 3D Point Clouds and textured Meshes from UAV LiDAR and Multi-View-Stereo. *arXiv preprint arXiv:2102.05346*.
Morsy, S., Shaker, A., El-Rabbany, A., 2017. Multispectral LiDAR data for land cover classification of urban areas. Sensors, 17(5), 958.

Neupane, S. R., Gharaibeh, N. G., 2019. A heuristics-based method for obtaining road surface type information from mobile lidar for use in network-level infrastructure management. Measurement, 131, 664-670.

Niemeyer, J., Rottensteiner, F., Sörgel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 87, January 2014, pp 152-165

Scaioni, M., Höfle, B., Baungarten Kersting, A.P., Barazzetti, L., Previtali, M., Wujanz, D., 2018. Methods for Information Extraction from Lidar Intensity Data and Multispectral Lidar Technology. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-3, 1503-1510. Doi:10.5194/isprs-archives-XLII-3-1503-2018.

Shan, J., Toth, C.K., 2009. Topographic Laser Scanning and Ranging, Principles and Processing, Taylor and Francis Group, Boca Raton, FL-USA.

Shanoer, M.M., Abed, F.M., 2018. Evaluate 3D laser point clouds registration for cultural heritage documentation. The Egyptian Journal of Remote Sensing and Space Science, 21(3), 295-304. doi: 10.1016/j.ejrs.2017.11.007.

Speak, A., Escobedo, F. J., Russo, A., Zerbe, S., 2020. Total urban tree carbon storage and waste management emissions estimated using a combination of LiDAR, field measurements and an end-of-life wood approach. Journal of Cleaner Production, 256, 120420.

Yan, W. Y., Shaker, A., Habib, A., Kersting, A. P., 2012. Improving classification accuracy of airborne LiDAR intensity data by geometric calibration and radiometric correction. ISPRS Journal of Photogrammetry and Remote Sensing, 67, 35-44.

You, H., Wang, T., Skidmore, A. K., Xing, Y. 2017. Quantifying the effects of normalisation of airborne LiDAR intensity on coniferous forest leaf area index estimations. Remote sensing, 9(2), 163. Doi: 10.3390/rs9020163

van Eck, N.J., Waltman, L., 2014. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. Journal of Informetrics, 8(4), 802-823. doi: 10.1016/j.joi.2014.07.006.

Zhang, J., Zhao, X., Chen, Z., Lu, Z., 2019. A review of deep learning-based semantic segmentation for point cloud. IEEE Access, 7, 179118-179133.