Processing of Extremely High Resolution LiDAR and RGB Data: Outcome of the 2015 IEEE GRSS Data Fusion Contest—Part B: 3-D Contest

A.-V. Vo, L. Truong-Hong, D. F. Laefer, D. Tiede, S. d'Oleire-Oltmanns, A. Baraldi, M. Shimoni, *Member, IEEE*, G. Moser, *Senior Member, IEEE*, and D. Tuia, *Senior Member, IEEE*

***Abstract*—In this paper, we report the outcomes of the 2015 data fusion contest organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society. As for previous years, the IADF TC organized a data fusion contest aiming at fostering new ideas and solutions for multisource studies. The 2015 edition of the contest proposed a multiresolution and multisensorial challenge involving extremely high resolution RGB images (with a ground sample distance of 5 cm) and a 3-D light detection and ranging point cloud (with a point cloud density of approximatively 65 pts/m^2). The competition was framed in two parallel tracks, considering 2-D and 3-D products, respectively. In this Part B, we report the results obtained by the winners of the 3-D contest, which explored challenging tasks of road extraction and ISO containers identification, respectively. The 2-D part of the contest and a detailed presentation of the dataset are discussed in Part A.**

***Index Terms*—Image analysis and data fusion (IADF), light detection and ranging (LiDAR), multiresolution-data fusion, multisource-data fusion, multimodal-data fusion, object identification, road detection, very high resolution (VHR) data.**

I. Introduction to Part B

THREE-DIMENSIONAL high-spatial-resolution data have become a fundamental part in a growing number of applications such as urban planning, cartographic mapping, environmental impact assessment, cultural heritage protection, transportation management, and civilian and military emergency responses [2]. Among the data sources available, a light detection and ranging (LiDAR) sensor offers a fast and effective way to acquire 3-D data [3].

In this framework, this paper is the second of a two-part manuscript presenting and critically discussing the scientific outcomes of the 2015 edition of the Data Fusion Contest organized by the IADF TC of the IEEE-GRSS.[1] The 2015 Contest released to the international community of remote sensing a topical and complex image dataset involving 3-D information, multiresolution/multisensor imagery, and extremely high spatial resolutions. The dataset was composed of an RGB orthophoto and of a LiDAR point cloud acquired over an urban and harbor area in Zeebruges, Belgium (see [1, Sec. II]).

Given the relevance of this dataset for the modeling and extraction of both 2-D and 3-D thematic results, the Contest was framed as two independent and parallel competitions. The 2-D Contest was focused on multisource fusion for the generation of 2-D processing results at extremely high spatial resolution: the interested reader can find the presentation and discussion of the results in [1]. The 3-D Contest explored the synergistic use of 3-D point cloud and 2-D RGB data for 3-D analysis at extremely high spatial resolution. Its results are discussed in detail in this paper. In both cases, participating teams submitted original open-topic manuscripts proposing scientifically relevant contributions to the fields of 2-D/3-D extremely high resolution image analysis. Even though LiDAR [4] and VHR RGB [5] data were considered in the past contests, for the first time, a 3-D competition considering their joint use is proposed to the community.

The LiDAR system (which is made up of the LiDAR sensor, a GPS receiver, and an inertial measurement unit) emits intense focused beams of light and measures the time it takes for the reflections to be detected by the sensor. This information is used to compute ranges, or distances, to objects. The 3-D coordinates (i.e., x, y, z or latitude, longitude, and elevation) of the target objects are computed from the time difference between the laser pulse being emitted and returned, the angle at which the pulse was emitted, and the absolute location of the sensor on or above the surface of the Earth [6]. LiDAR instruments can rapidly measure the Earth's surface, at sampling rates greater than 150 kHz. The resulting product is a densely spaced network

Manuscript received January 01, 2016; revised April 15, 2016 and June 09, 2016; accepted June 13, 2016. The work of D. Tuia was supported by the Swiss National Science Foundation under Grant PP00P2-150593. The work presented in Section III was supported by the European Commission under European Research Council Grant ERC-2012-StG _ 20111012 "RETURN—Rethinking Tunnelling in Urban Neighbourhoods" Project 30786. *(Corresponding author: Devis Tuia)*.

A.-V. Vo, L. Truong-Hong, and D. F. Laefer are with University College Dublin, Dublin 4, Ireland (e-mail: anh-vu.vo@ucdconnect.ie; linh.truonghong@ucd.ie; debra.laefer@ucd.ie).

D. Tiede and S. d'Oleire-Oltmanns are with the Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria (e-mail: dirk.tiede@sbg.ac.at; Sebastian.dOleire-Oltmanns@sbg.ac.at).

A. Baraldi is with the Department of Agricultural and Food Sciences, University of Naples Federico II, 80138 Napoli, Italy, and also with the Department of Geoinformatics—Z_GIS, University of Salzburg, 5020 Salzburg, Austria (e-mail: andrea6311@gmail.com).

M. Shimoni is with the Signal and Image Centre, Department of Electrical Engineering, Royal Military Academy (SIC-RMA), B-1000 Brussels, Belgium (e-mail: mshimoni@elec.rma.ac.be).

G. Moser is with the University of Genoa, Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture (DITEN), University of Genoa, 16126 Genova, Italy (e-mail: gabriele.moser@unige.it).

D. Tuia is with the Department of Geography, University of Zurich, 8006 Zürich, Switzerland (e-mail: devis.tuia@geo.uzh.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2016.2581843

[1] http://www.grss-ieee.org/community/technical-committees/data-fusion/

of highly accurate georeferenced elevation points, often called a point cloud, that can be used to generate 3-D representations of the Earth's surface and its features. Typically, LiDAR-derived elevations have absolute accuracies of about 10–20 cm [7], [8].

LiDAR, as a remote-sensing (RS) technique, has several advantages. Chief among them are high-resolution and high-accuracy horizontal and vertical spatial point cloud data, large coverage areas, the ability of users to resample areas quickly and efficiently, and the extraction of a 3-D presentation of the scene [9], [10]. The intensity of LiDAR points can be used as additional useful information for the segmentation of features in the scene [11], [12]. Moreover, the shadow effects are alleviated in LiDAR data [13]. However, due to the nature of the point cloud data (irregular distribution), the laser measurements do not always allow a precise reconstruction of the 3-D shape of the target (e.g., building edges) [13], [14]. LiDAR point clouds have been used successfully for urban object recognition and reconstruction [15], [16], but the task can be strongly eased by using LiDAR in conjunction with spectral information [13], [17], especially when the point density of the LiDAR data is low.

To overcome these shortcomings, integrated approaches fusing LiDAR data and optical images are increasingly used for the extraction and the classification of objects in urban scenes [18], [19], such as trees [20]–[22], buildings [14], [23]–[25], roads [13], [26]–[31], vehicles [17], [32], [33], and other small objects [33], [34]. Although the combined use of different data sources is theoretically better than using a single source, there are still drawbacks. It is costly or sometimes even impossible to obtain different types of data for many applications. The difficulties of processing are increasing, and when combining different data or using object cues from multiple sources, a proper fusion methodology is needed to achieve an accurate outcome [13]. Nevertheless, these multisource fusion methods have strict requirements for data acquisition and registration [14] (e.g., for pixel-based fusion, subpixel accuracy is required). For this last aspect, several approaches provide frameworks for automated registration of 2-D images onto 3-D range scans. Most of the available methods are based on extracting and matching features (e.g., points, lines, edges, rectangles, or rectangular parallelepipeds) [35]–[37]. Others describe solutions that combine 2-D-to-3-D registration with multiview geometry algorithms obtained from the parameters of the camera [38], [39].

A major topic in the LiDAR-RGB fusion literature is the detection of roads and objects (i.e., buildings, containers, marine vessels, and vehicles). The integration of 2-D optical and 3-D LiDAR datasets provides photorealistic textured impression that facilitates the detection and the extraction of large-scale objects from the scene. The literature mainly covers the topic of feature-based fusion for building extraction [23], building surface description [29], [40], detection of roof planes and boundaries [24], structure monitoring [25], and urban building modeling [41], but it also addresses the extraction and the identification of small- to medium-scale objects [32]–[34], [42].

Automatic extraction of roads in complex urban scenes from remotely sensed data is very difficult to perform using single RS source [13]. In passive imagery, the occlusion of the road surface by vertical objects creates artifacts such as shadows, radiometric inhomogeneity, and mixed spectra that complicate road detection. The properties of airborne LiDAR imagery make it a better data source for road extraction in urban scenes. Free-of-shadow effects, relatively narrow scanning angle (typically 20–40° [26]), laser reflectance, and elevation information allow good separation of roads from other urban objects [12], [27]. However, due to the aforementioned lack of spectral information and irregular distribution of LiDAR points, more effort is needed to extract accurate break lines or features, such as road curbs and sidewalks [13]. Given the pros and the cons of LiDAR and aerial imagery, it has been suggested that these data be fused to improve the degree of automation and the robustness of automatic road detection [13], [26]–[28].

The 2015 Data Fusion Contest involved two datasets acquired simultaneously by passive and active sensors. Both datasets were acquired on March 13, 2011, using an airborne platform flying at an altitude of 300 m over the harbor area of Zeebruges, Belgium (51 : 33°N, 3 : 20°E). The Department of Communication, Information, Systems and Sensors of the Belgian Royal Military Academy provided the dataset and evaluated its accuracy, while service provider acquired and processed the data. The passive data are 5-cm-resolution RGB orthophotos acquired in the visible wavelength range. The active source is a LiDAR system that acquires the data using scan rate, angle, and frequency of 125 kHz, 20°, and 49 Hz, respectively. For obtaining a digital surface model with a point spacing of 10 cm, the area of interest was scanned several times in different directions with a high-density point cloud rate of 65 pts/m^2. The scanning mode was "last, first and intermediate." More details on the dataset can be found in Part A of this paper [1].

In this paper, we present the works of the winning teams of the 3-D contest and provide a general discussion of the results: first for the 3-D contest and then overall for the 2015 IEEE GRSS Data Fusion Contest. We invite the readers interested in the 2-D contest, as well as in the detailed presentation of the datasets to refer to the sister publication, i.e., the Part A manuscript [1]. For the 3-D track, the papers awarded were:

1) First place: "Aerial laser scanning and imagery data fusion for road detection in city scale" by A.-V. Vo, L. Truong-Hong, and D. F. Laefer from University College Dublin (Ireland) [43].
2) Second place: "Geospatial 2-D and 3-D object-based classification and 3-D reconstruction of International Standards Organization (ISO) containers depicted in a LiDAR dataset and aerial imagery of a harbor" by D. Tiede, S. d'Oleire-Oltmanns, and A. Baraldi from the University of Salzburg (Austria) and the University of Naples Federico II (Italy) [44].

In this paper, the overall set of submissions is presented in Section II. Then, the approaches proposed by the winning and runner-up teams of the 3-D contest are presented in Sections III and IV, respectively. A discussion of these approaches is reported in Section V. Finally, a general discussion on the Data Fusion Contest 2015 concludes the paper in Section VI.

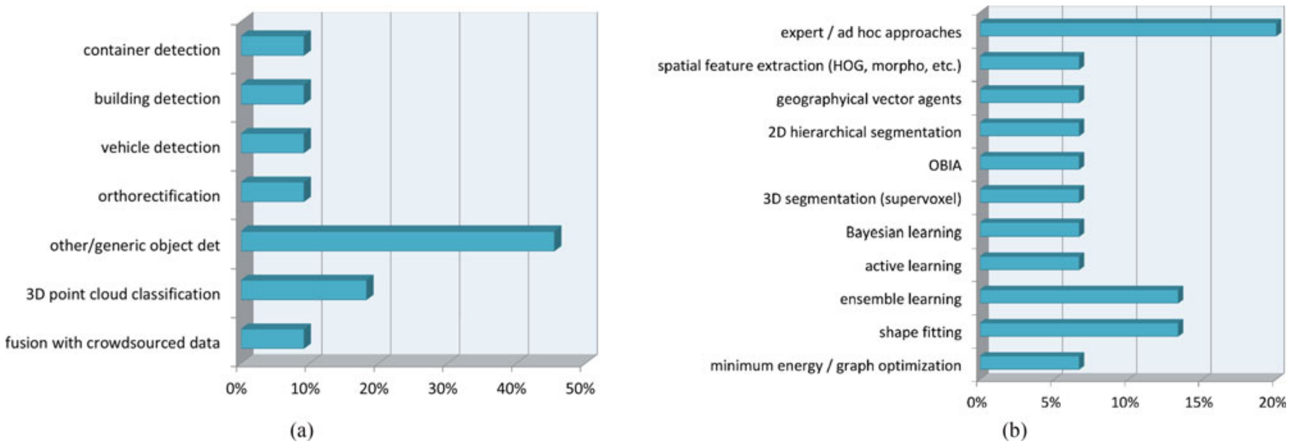

Fig. 1. Summary of the ten submissions to the 3-D contest by topics (a) and approaches considered (b).

II. Discussion of the 3-D Contest: The Submissions

A total of ten submissions were received for the 3-D contest. The participants mainly addressed object detection as the topic of their research. In total, 80% of the submissions (see Fig. 1) used the datasets to extract or improve the detection of different objects in the urban and the harbor environments either targeting specific types of objects (containers, buildings, or vehicles) or addressing general-purpose object detection tasks. Problems of 3-D point cloud data classification, orthorectification, and fusion with crowdsourced data were also considered by the participants. In particular, fusion with data from crowdsourcing has been investigated only recently in RS, and a submission on this topic further confirms that new challenges were explored in addition to those that are more traditional for the IADF community, a scenario similar to what was also remarked for the 2-D Contest [1].

Regardless of the prevailing focus on detection problems, and as may be observed in Fig. 1, the methodological approaches were highly heterogeneous. Most proposed processing schemes were complex and time consuming and integrated various learning and optimization procedures. Several methods made use of expert systems or *ad hoc* approaches that were customized to the expected properties of the targets to be detected. This was quite an expected result due to the usual need to incorporate prior information in the identification of ground objects at extremely high spatial resolution. Segmentation techniques were also prominent among the submitted manuscripts. Traditional (e.g., classical region growing) or advanced hierarchical 2-D segmentation methods were considered. Contributions using 3-D segmentation algorithms, which operate on 3-D voxel data rather than 2-D pixel data, are quite popular in other areas (e.g., computer-aided tomography or ecography), but are less frequently seen in aerial RS, were also received.

Several learning approaches were integrated into the proposed target detection schemes. From a methodological perspective, they encompassed Bayesian, ensemble, and active learning, as well as Markovian minimum energy methods with suitable graph-theoretic formulations (e.g., graph cuts). From the viewpoint of applications, these learning contributions in the submissions to the 3-D Contest were mostly customized to case-specific classification subproblems involved in the target detection pipelines. In the two following sections, the approaches proposed by the first and second ranked teams are presented.

III. Aerial Laser Scanning and Imagery Data Fusion for Road Detection in City Scale

Automatic road detection from RS data is useful for many real-world problems such as autonomous navigation. Similar to road detection from imagery data [45], existing methods for road extraction from laser scanning data can be categorized as either top-down (i.e., model-based) or bottom-up (i.e., data driven). Top-down methods rely on prior knowledge (i.e., about shape, color, and intensity) of the objects to be detected. The approach is often more robust than the bottom-up counterpart. However, its lack of flexibility due to the reliance on the prior knowledge is a significant drawback. Authors in [46], [47] provide examples of top-down road detection methods. The former projected a spline model atop an existing map to point cloud data to produce a 3-D road map. Each road segment was then fitted by a 2-D active contour. The contours were attached to road curbs, which were indicated by sudden height changes in the point data. The latter study utilized a Hough transformation to detect roads where road segments were assumed to have thin straight ribbon shapes.

Bottom-up solutions exploit the homogeneity (e.g., color, intensity, height) between adjacent data points to segment the data into homogeneous continuous patches. Over- and under-segmentation are the common challenges to this solution. As an example of the bottom-up approach, [48] grouped adjacent points incrementally based on the difference in the height and laser reflectance between a point and its neighbors. Other exemples of bottom-up efforts to address this problem can be found in [49]–[51]. In that research, point data are filtered according to their laser reflectance and their height with respect to a digital terrain model before undergoing a connected component labeling process for a group of points.

While most methods in both categories successfully exploit laser reflectance, absolute height, and height variance as a means

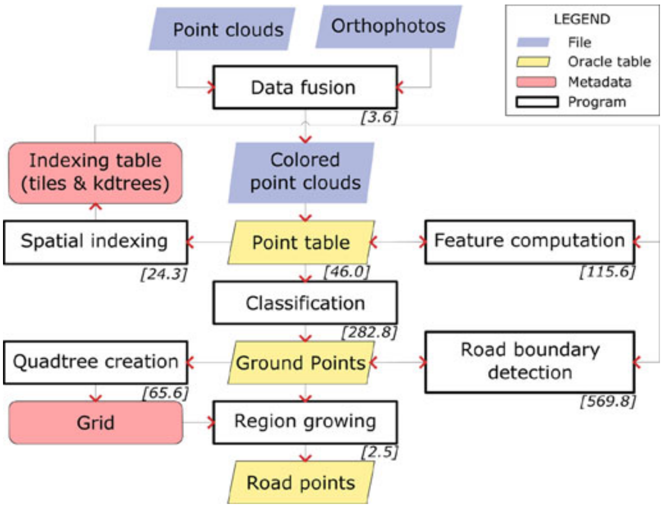

Fig. 2. General workflow of the road detection process; bracketed numbers indicate processing times (in minutes) on an Intel Xeon CPU E5-2665 0 @ 2.4 GHz with 32-GB RAM; the processed dataset contains 40 904 208 points and the number of ground points is 26 251 729 points

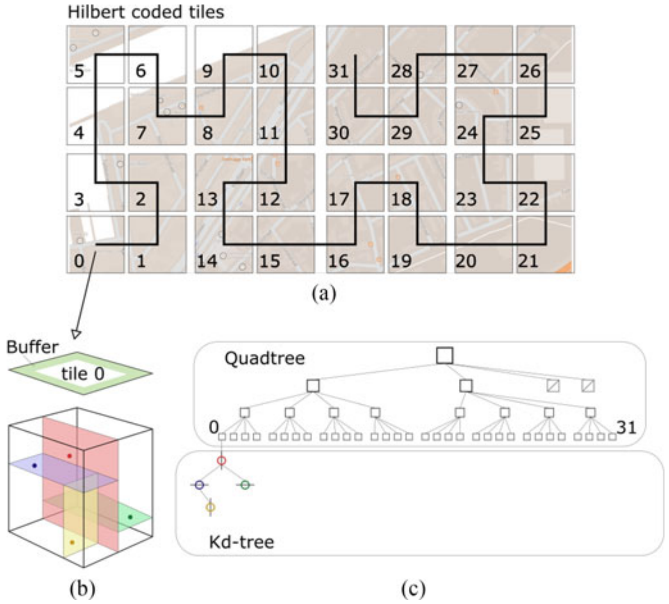

Fig. 3. Hybrid 2-D quadtree–3-D kd-tree indexing. (a) Top indexing level: Hilbert coded regular tiles. (b) Bottom indexing level: local kd-tree of tile 0 in the Euclidean 3-D space. (c) Hybrid index represented as a hierarchical structure.

for distinguishing road versus nonroad data, when airborne LiDAR data are considered, the relatively limited data density makes the accurate delineation a very complex task. Given the LiDAR point data provided by the Data Fusion Contest 2015 together with high-resolution imagery data, the issue of road extraction was revisited with the two following objectives in mind: 1) improve geometric analysis on LiDAR point data to better exploit the laser data density; and 2) investigate added benefits of imagery data when being used in conjunction with LiDAR data.

The awarded approach for the 3-D track introduced three main components:

1) an end-to-end point cloud processing workflow;
2) a new algorithm for extracting points on road boundaries from aerial laser scanning (ALS) data fused with aerial imagery data;
3) an innovative data management strategy which is essential for large-scale high-resolution data analyses.

A. General Data Processing Workflow

In this section, we briefly introduce the main components of the proposed workflow and the data management issues. Each component is then detailed in one of the following sections (see III-B to III-C).

1) Main Ingredients of the Workflow: The proposed road extraction workflow is illustrated in Fig. 2. The ALS point clouds and the orthophotos are fused together to form colored point clouds based on the spatial relationship between the LiDAR points and the orthophotos' pixels. If a 2-D footprint of a point is enclosed inside a pixel, the fusion assigns the pixel's color to the data point. The colored point clouds are then loaded into an Oracle database and indexed using a hybrid quadtree/kd-tree indexing scheme. Next, features such as normal vector, local surface roughness, and hue/saturation/lightness (HSL) are computed. Then, a supervised classification (see Section III-B) classifies points in the following three classes:

1) ground;
2) building;
3) unassigned.

Road curbs and other obstacles bounding road regions are then detected before running a quadtree-based region growing algorithm (see Section III-C); this algorithm connects an initial seed road point to other ground points, if an obstacle-free path exists that connects the seed to the new points.

2) Data Management: For the purpose of the study, tiles 5 and 6 of the given dataset were selected (see [1, Fig. 2]), consisting of approximately 40 million LiDAR points. The large nature of the data precludes storing the entire point cloud within the main memory for most conventional computers. So, an out-of-core data management solution is required. Spatial indexing is necessary since data retrieval based on spatial conditions is frequent and computationally demanding during the process. In this study, a hybrid quadtree/kd-tree structure (see Fig. 3) was implemented atop an Oracle database storing LiDAR points in order to enable fast all-nearest-neighbor (ANN) computation on the points. The hierarchical structure of the hybrid index is shown in Fig. 3. At the top level, point data were partitioned into multiple 125 m × 125 m tiles using a Hilbert code implementation [see Fig. 3(a)]. The Hilbert implementation maps each point in its spatial domain to a numeric code (i.e., Hilbert code) indicating a unique, finite, rectangular cell containing that point. The mapping efficiently facilitated by the bit-interleaving process [52] is used for splitting the data, as well as rapidly identifying portions of data related to spatial queries during the workflow. A Morton code or any other spatial locational code [53] can be used as an alternative to the Hilbert code. The tile size was selected based on the amount of memory dedicated for the index (e.g., approximately 120 megabytes/tile).

Under each tile, a 3-D kd-tree was built [see Fig. 3(b)]. The kd-tree includes points in a buffer around the tile, plus the points within this tile itself to avoid discontinuity around the tile boundaries [the green region in Fig. 3(b)]. The buffer must be sufficiently large to provide adequate neighboring points for the Feature Computation and the Road Boundary Detection steps (see Fig. 2). The neighborhood sizes in those two steps, in turn, depend on the point data density. The sizes should be sufficiently large to obtain adequate points for the robust neighbor-based computations and small enough to expose the local characteristics expected from the computations. In this research, a buffer size of 1 m is selected. The kd-trees were made reusable by being serialized and stored as binary large objects in an indexing table. Each kd-tree was retrieved and deserialized back to the main memory, once a search within its extent was invoked. This hybrid indexing adapts well to the spatial distribution of ALS data (i.e., dominantly horizontal (2-D) at the global level and fully 3-D at the local level). Within a tile, ANN queries are fast because its associated 3-D index resides in the main memory. This implementation is not yet generic, nor optimal, but did sufficiently support all spatial queries performed in this study.

TABLE I
COMBINATIONS OF POINT FEATURES FOR CLASSIFICATION

	H	NV	SR	II	LI	HV	Comments
FV0	●	●	●				Most important features
FV1				●	●	●	Exclude core features
FV2	●	●	●	●	●	●	**Best combination**
FV3	●	●	●		●	●	Influence of II
FV4	●	●	●	●		●	Influence of LI
FV5	●	●	●	●	●		Influence of HV
FV6	●	❍	●	●	●	●	(n_x, n_y, n_z) vs. (θ, ϕ) NV
FV7	●	●	●	✧	●	●	RGB vs. HSL color

● feature included for classification
❍ switch normal vector from spherical to Descartes coordinate system
✧ switch color from HSL to RGB space
height (H), normal vector (NV), surface roughness (SR), image intensity (II), laser intensity (LI), height variation (HV), point density (PD)

B. Point Cloud Classification

To reduce the computational efforts of the method, a point classification is performed on the entire point cloud before applying a more demanding road extraction process only to the ground points (see Section III-C). A supervised approach is employed to classify the point cloud into one of three groups: building, ground, or unassigned. Even though the class "building" is not exploited in the later part of the workflow, there is a possibility of relating buildings in the road detection process, since roads and buildings have a strong proximity to each other in an urban design context.

The classification process involves three main steps: first, point features are computed; then, the best-performing feature vector among several selected vectors is chosen, and finally, the entire dataset is classified with the best-performing classifier. The Weka toolkit was utilized to create classification models in this study [54].[2]

After being fused with orthophotos, the point cloud possesses four raw (i.e., from sensor) attributes: intensity and the three RGB color values. In this research, raw intensity values were used, as insufficient information was available under the contest for a correction process. Even though the raw laser intensity delivered from a laser scanner has a certain relationship to scanned surface characteristics (which is expected for the classification), the quantity is influenced by other factors including ranges and angles of incidence [55]. Notably, the registration between 2-D ortho-rectified images and 3-D laser points is imperfect, particularly on vertical surfaces (e.g., building façades) and under overhang structures (e.g., walls underneath overhanging eaves). In such cases, points (e.g., on walls) can be mistakenly assigned colors from objects directly above them (e.g., eaves). Nevertheless, since the objects of interest in this study are mainly horizontal (i.e., roads and pavements), the registration's imperfection has minor impact on the final results. In addition to the attributes derived directly from sensor data, there are several features that can be derived from the point coordinates and the initial attributes that can be beneficial for point classification. In this study, the following features were investigated:

1) height: z value of the point;
2) image intensity, described in two color spaces (RGB and HSV);
3) laser intensity;
4) height variation: maximum variation of z values within a spherical neighborhood N of the given point;
5) surface roughness: indirectly represented by the quadratic mean of orthogonal distances of all points in N to a plane P fitting to all points in N;
6) normal vector of P, represented as (n_x, n_y, n_z) in a Cartesian coordinate system or (θ, ϕ) in a radial coordinate system. An iterative principal component analysis [56] was implemented with a weighting factor inversely proportional to the point-to-plane distance to improve plane fitting.

To analyze the influence of the above features, several combinations (termed feature vectors) were investigated (FV0 to FV5 in Table I). FV6 and FV7 compared the differences caused by the various ways of representing normal vectors and colors. Performance of each feature vector was evaluated by a training-and-evaluating process. Sixteen different regions selected from the original data covering approximately 16% of the study area were manually labeled. Two-thirds of the labeled data were used to build a J48 decision tree classifier with a Weka machine learning toolkit [54]. J48 is Weka's name for its improved implementation of the widely used C4.5 (revision 8) decision tree by Quinlan [57]. The C4.5 algorithm was recognized as one of the top ten most influential data-mining algorithms in 2006 [58]. The classification model is a decision tree generated by a divide-and-conquer strategy with an ultimate pruning step to avoid overfitting. Classifier accuracy was estimated against the remaining labeled data. The classification performance of each feature vector is plotted in Fig. 4 with

[2] http://www.cs.waikato.ac.nz/~ml/weka/index.html

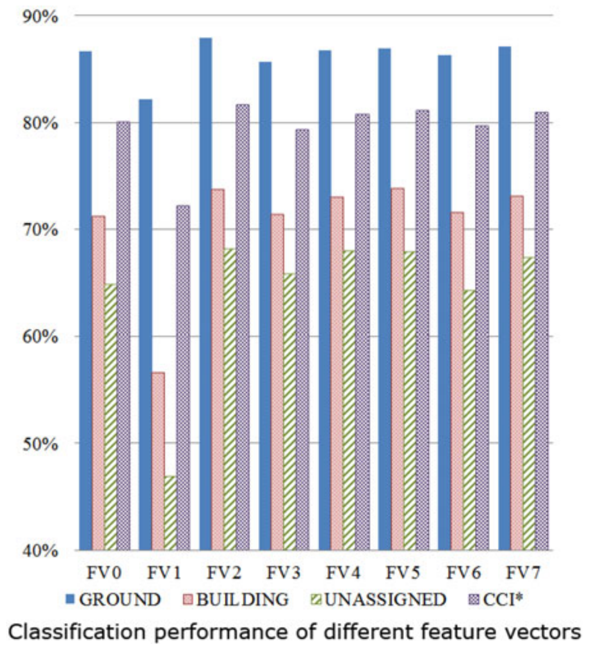

Fig. 4. Performance of the feature vectors in Table I on the held out validation set.

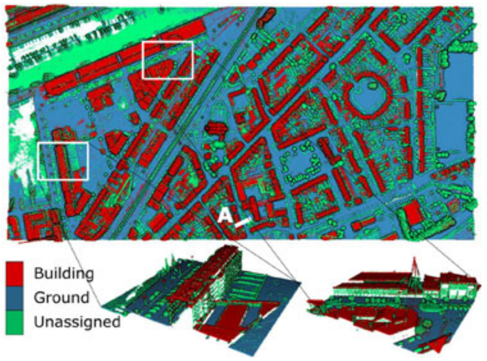

Fig. 5. Classification result. The "A" cross section is visualized in Fig. 6(a).

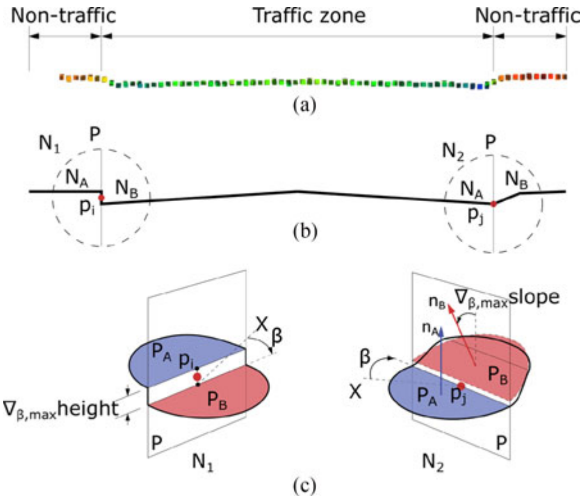

Fig. 6. Directional height and slope variation computation. (a) Real road cross section from LiDAR point cloud. (b) Local neighborhood around road curbs. (c) Directional height and slope variation.

four measures: $F1$ score for ground (blue), building (red), and unassigned points (green), and the number of correctly classified instances (CCI) including all three classes (purple). The $F1$ scores computed for each class are the harmonic means of precision and recall while evaluated against the testing sets, $F1 = 2(\text{precision} \times \text{recall})/(\text{precision} + \text{recall})$.

Based on Fig. 4, the most important features for point classification were height, normal vector, and local surface roughness. The CCI was 80.1% when the three core features were used (FV0) but dropped to 72.3% when excluded (FV1). The absence of image intensity, laser intensity, and height variation reduced the CCI by 2.3%, 0.8%, and 0.9%, respectively (FV3, 4, and 5). Representing normal vectors in a radial form (θ, ϕ) was better for classification than Cartesian coordinates (n_x, n_y, n_z) (FV2 versus FV6) due to the trivial lengths of normal vectors, which can be discarded when the vectors are represented in a spherical system. In Cartesian coordinates, the lengths are blended into all three variables (n_x, n_y, n_z), which complicates the problem without improving classification levels. The HSL color provided better results than the RGB color space (81.7% in FV7 versus 79.7% in FV2) as previously noted by Sithole [59]. While color is treated in its entirety in this research, there is a possibility of investigating each single color component (H/S/V or R/G/B) for point classification (e.g., [60], [61]).

Overall, the classification rates were significantly higher for the class "ground" than for the other two classes (blue versus red and green columns in Fig. 4), mainly because of feature consistency. The best-performing feature vector was identified as the group, which included height, normal vector (θ, ϕ), roughness, HSL color, intensity, and height variation. To exploit all the manually labeled data, all data are then used to build the final classifier. Such a final classifier is applied to the entire dataset (see Fig. 5). Points classified as "ground" were further processed for road extraction, as detailed in the next section.

C. Road Extraction

Laser intensity has been used successfully for distinguishing road surfaces from other materials (e.g., [47], [49], and [50]). Asphalt appears within a very distinctive range in the laser intensity spectrum. However, in this study, the roads were made of various materials, which made intensity less discriminative. Thus, a new method was needed. The proposed method has two main steps: 1) the identification of road curbs and obstacles bounding roads based on the spatial distribution of the point data; and (2) the extraction of road points using a quadtree-based region growing algorithm that considers intensity and color simultaneously.

1) Detection of Road Curbs and Obstacles: Road curbs and obstacles were defined as objects preventing vehicle progression due to height or slope variation within a finite spatial extent (e.g., a 1-m-radius circle). Fig. 6(a) shows an example along a road cross section located at position A in 5. Small features (e.g., curbs) are visible due to the high data density. The method

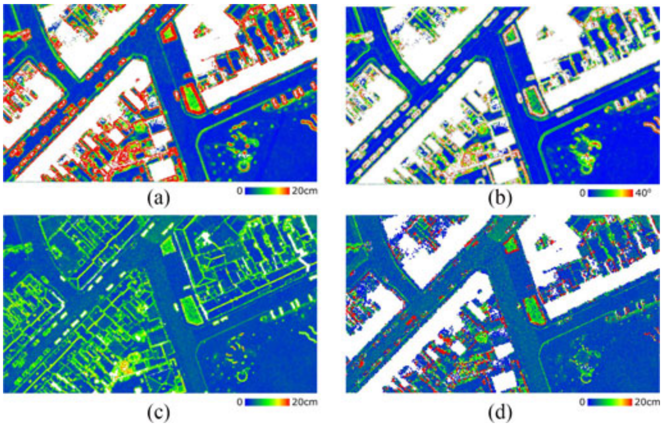

Fig. 7. Directional slope and height variation versus conventional height variation and residual. (a) $\nabla_{\beta,\max}$ height. (b) $\nabla_{\beta,\max}$ slope. (c) Residual. (d) Height variation.

computes two features: directional slope variation, $\nabla\text{slope}(p_i)$, and directional height variation, $\nabla\text{height}(p_i)$. Both are computed for every single point p_i among the ground points.

First, the neighboring points N within the spherical neighborhood of p_i are partitioned into two point groups, NA and NB, by a vertical plane P containing p_i and making an angle of β with the x-axis [see Fig. 6(b)]. PA and PB are defined as the best fit planes to the points in NA and NB. With a given value of β, $\nabla\text{slope}(p_i)$ is defined as the angle between PA and PB, whereas $\nabla\text{height}(p_i)$ is the height difference between the vertical projections of p_i on PA and PB. Finally, the maximum directional slope and height variations are determined with respect to β, $\nabla_{\max}\text{slope}(p_i)$ and $\nabla_{\max}\text{height}(p_i)$ [see Fig. 6(c)].

Fig. 7(a) and (b) present the results of $\nabla_{\max}$slope and $\nabla_{\max}$height for a segment of ground points. Road boundaries are clearly distinguishable and are defined better than using the residual value—see Section III-A—[see Fig. 7(c)] or the nondirectional height variation approaches [see Fig. 7(d)]. Points having $\nabla_{\max}\text{slope} > 8^\circ$ and $\nabla_{\max}\text{height} > 5$ cm were considered as obstacles and were used as inputs to the region growing road extraction algorithm presented in the next section. Thresholds were selected empirically.

2) Quadtree-Based Region Growing: A seeded region growing algorithm was combined with a rasterization for performance enhancement [62] to extract road points (see Fig. 8). Region growing requires an initial seed, which is a pixel (i.e., node in the quadtree) recognized with high certainty as being a portion of the road network to be detected. In this study, the selection of the initial seed is performed manually. An example is given at the plus mark in Fig. 8(b). Around the seed, a buffer approximating a required clearance for one vehicle is constructed (e.g., a 0.75-m-radius circle). Every pixel within the buffer, as well as all points enclosed in the pixel, is labeled as road, if the buffer is obstacle-free. The newly detected road pixels are then set as new seeds for the next iterations, if they satisfy additional intensity and color conditions (i.e., intensity < 550 and hue $\in (0.18, 0.3)$). Similarly to the directional slope and height variations criteria, the color and intensity thresholds were empirically defined. Since each pixel often contains more than a single point, intensity and color values of a pixel are set as the maximum intensity and the average HSL of all points contained in the pixel. These criteria help distinguish roads from grass, even though they are insufficient by themselves for direct road extraction.

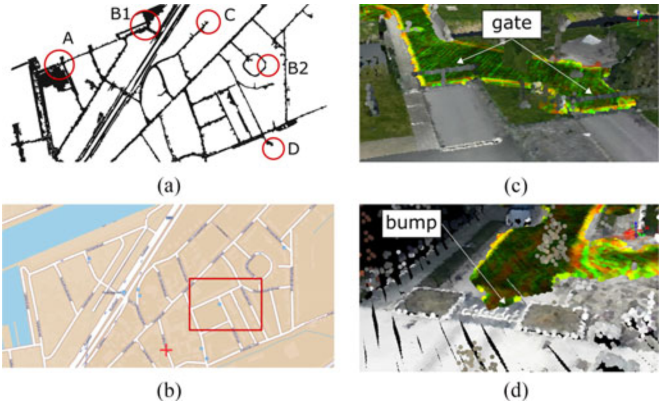

Fig. 8. Road extraction results. (a) Extracted road network. (b) Reference road map (Google Maps). (c) Barriers at D. (d) Speed bump at C.

D. Discussions and Concluding Remarks

The proposed approach was able to successfully identify most of major road segments, except those blocked by speed bumps or other obstacles (e.g., location C&D in Fig. 8). Notably, the approach is not fully automatic and requires human knowledge in the form of threshold selection. Arguably, the thresholds for slope and height variations can be derived from road design codes taking into account data noise and other artifacts caused by data acquisition and processing. Nevertheless, a complete consideration of the issue is complex. Given the time limits of the contest, this was circumvented by the researchers through the adoption of a trial-and-error approach. To assist in this, a graphical user interface was created with sliders to preview the detected roads of specific regions under various threshold. Once an initial set of results is considered as satisfactory, further regions can be tested or the thresholds can be used to populate the entire dataset. Fig. 9 shows the sensitivity of the result to the height and slope variation thresholds separately. When more relaxed criteria were selected (i.e., higher values for the thresholds), more spurious points were detected (see Fig. 9). In contrast, stricter criteria (i.e., low thresholds) lead to a higher miss rate. For the time being, controlling the balance between the two states is left to human intervention. A similar approach was employed for intensity and color-threshold selections. Most false alarms were attributable to large parking lots (e.g., location A in Fig. 8), while most of missed road segments were caused by vehicle congestion at both ends of the segments. A possible solution for the latter issue is to filter transient objects including moving vehicles before conducting the processing. Such a filtering would be straightforward if point data were acquired from multiple overlapping flight strips such as the case of Dublin city's data [63]. Moving objects can be detected based on their inconsistent appearance in the scans of the same scene given

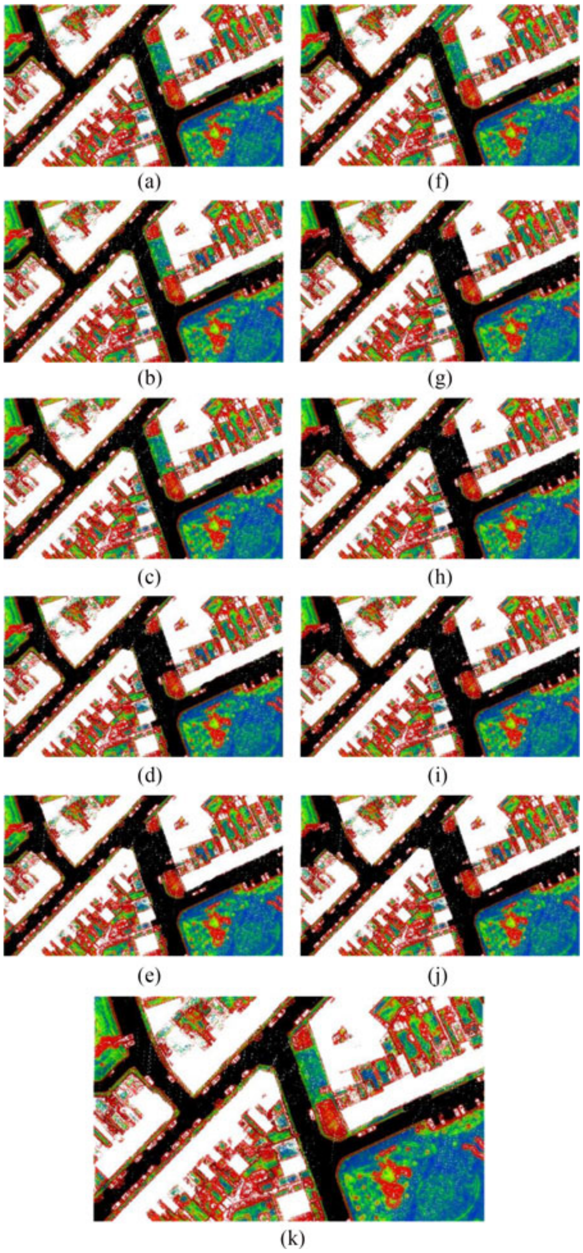

Fig. 9. Influences of slope and height variation thresholds on road detection results; black points are those detected as road. (a) $\bigtriangledown$height $< \infty$; $\bigtriangleup$slope $< 6^\circ$. (b) $\bigtriangledown$height $< \infty$; $\bigtriangleup$slope $< 7^\circ$. (c) $\bigtriangledown$height $< \infty$; $\bigtriangleup$slope $< 8^\circ$. (d) $\bigtriangledown$height $< \infty$; $\bigtriangleup$slope $< 9^\circ$. (e) $\bigtriangledown$height $< \infty$; $\bigtriangleup$slope $< 10^\circ$. (f) $\bigtriangledown$height $<$ 3 cm; $\bigtriangleup$slope $< \infty$. (g) $\bigtriangledown$height $<$ 4 cm; $\bigtriangleup$slope $< \infty$. (h) $\bigtriangledown$height $<$ 5 cm; $\bigtriangleup$slope $< \infty$. (i) $\bigtriangledown$height $<$ 6 cm; $\bigtriangleup$slope $< \infty$. (j) $\bigtriangledown$height $<$ 7 cm; $\bigtriangleup$slope $< \infty$. (k) $\bigtriangledown$height $<$ 5 cm; $\bigtriangleup$slope $< 8^\circ$.

that the analysis can exclude other causes of the absence (e.g., occlusion) [64].

Evaluation against a manual delineation showed a precision of 66.1%, recall of 91.9%, and an $F1$ score of 76.9%. While the LiDAR point cloud provided the highly accurate dense 3-D data enabling detection of fine features such as road curbs or barriers, color from the orthophotos increased the point cloud classification accuracy by 2.3%, equivalent to adding approximately 920 000 points. Orthophoto-based color and laser intensity also helped excluding grassy areas [e.g., B1 and B2 locations in Fig. 8(a)]. Nevertheless, the usage of color would be impeded, if input photos contained significant shadows. In such cases, the additional benefit of using colors is not applicable, and the data processing would have to rely only on geometric processing.

While minimizing computational cost is not the aim of this study, the processing time for each single module in the chain is presented in Fig. 2 (with the total of 18.5 h). Even though the classification step improves the performance, it is not compulsory, and further optimization alleviating the costs is possible. The proposed approach is more widely applicable than many other studies in this field, as it does not require a digital elevation model or a 2-D road map as input. The initial seeding point selection could be automated, as well as the definition of thresholds for directional slope and height variation. The results showed all locations accessible from the initial seed point. On one hand, both data acquisition and processing are mostly automated; therefore, the maps generated by the proposed approach are more likely to be up-to-date. On the other hand, such maps directly imply geometry constraints. Multiple maps can, therefore, be generated from the same set of data to fit different vehicle capacity or user requirements, which may be highly useful for navigation.

IV. Automated Hierarchical 2-D and 3-D Object-Based Recognition and Reconstruction of ISO Containers in a Harbor Scene

A. *Introduction*

The inventory of rapidly changing logistics infrastructures, such as depots and harbors, is crucial to their efficiency. For example, an optimized exploitation of an intermodal storage volume for shipping containers requires an inventory where positional, geometric, and identification attributes of individual containers are known in real time. In the 3-D track of the 2015 GRSS Data Fusion Contest, this work tackled the problem of freight container localization and classification at a harbor by an automated near real-time computer vision (CV) system, in agreement with the ISO 668—Series 1 freight containers documentation adopted as a source of *a priori* (3-D) scene-domain knowledge [65].

B. *Methods*

1) Selected 2-D and 3-D Sensory Datasets: Focusing on the harbor area visible in Tiles #1-2-3-4, refer to [1, Fig. 2], the input datasets selected for use were the uncalibrated three-band true-color RGB aerial (2-D) orthophoto featuring very high spatial resolution, below 10 cm, and the dense 3-D LiDAR point cloud described in [1, Sec. II]. The available DSM was not used, because it appeared to lack nonstationary surface-elevated objects, such as cranes and freight containers, in disagreement with the LiDAR point cloud. In addition, a slight tilting effect was observed to affect the RGB orthophoto in comparison with the LiDAR point cloud, across image locations where above-ground objects, such as container stacks, were depicted. Such a tilting effect could be caused by an image orthorectification process employing as input the aforementioned DSM, where nonstationary above-ground objects were absent. In practice, the target CV system was required to cope with the observed tilting effect when 2-D and 3-D datasets were spatially overlapped.

2) In-House DSM Generation From the LiDAR Point Cloud: To reveal above-ground scene elements, such as containers and cranes, the 3-D LiDAR point cloud was integrated as a raster (2-D gridded) point cloud, where the DSM pixel size was the same of the input orthophoto (5 cm) and the DSM pixel value was the LiDAR highest elevation value z occurring per pixel.

C. Main Workflow

Human panchromatic vision is nearly as effective as color vision in the provision of a complete scene-from-image representation, from local syntax of individual objects to global gist and layout of objects in space, including semantic interpretations and even emotions [66], [67]. This fact means that spatial information dominates color information in the spatiotemporal 4-D real world-through-time domain, described by humans in user-speak [65], as well as in a (2-D) VHR image domain, to be described in technospeak [66], irrespective of data dimensionality reduction from 4-D to 2-D [67], [68]. It agrees with the increasing popularity of the object-based image analysis (OBIA) paradigm [69]–[71], proposed as a viable alternative to traditional 1-D image analysis, where intervector topological (neighborhood) relationships are lost when a 2-D gridded vector set is mapped onto a 1-D vector stream [72]. To develop a CV system capable of 2-D spatial reasoning in a VHR image domain for 3-D scene reconstruction, a hybrid inference system architecture was selected. According to Marr, the linchpin of success of any data interpretation system is architecture and knowledge/information representation, rather than algorithms and implementation [73]. In hybrid inference, deductive and inductive inferences are combined to take advantage of each other and overcome their shortcomings [67], [70], [71], [74]–[76]. On one hand, inductive (bottom-up, learning-from-data, statistical model-based) algorithms, capable of learning from either unsupervised or supervised data, are inherently ill-posed and require *a priori* knowledge in addition to data to become better posed for numerical solution (see [77, p. 39]). In the RS, common practice inductive algorithms are semiautomatic and site-specific [74], [75], [78]. On the other hand, expert systems (deductive, top-down, physical model-based, prior knowledge-based inference systems) are automated, since they rely on *a priori* knowledge available in addition to data, but lack flexibility and scalability [74], [75], [77], [78]. An original hybrid CV system architecture was selected to:

1) comply with the OBIA paradigm [69]–[72];
2) start from a first stage of automated deductive inference, to provide a second stage of inductive learning-from-data algorithms with initial conditions without user interaction;
3) employ feedback loops, to enforce a “stratified” (driven-by-knowledge, class-conditional) approach to unconditional sensory data interpretation [67], [68], [78].

This approach is equivalent to a focus-of-visual-attention mechanism [67], [77] and to the popular divide-and-conquer problem-solving approach [77]. Sketched in Fig. 10, the implemented CV system consisted of three main modules:

1) a first stage of application-independent automated preprocessing for uncalibrated RGB image harmonization, enhancement, and preliminary classification;
2) a second stage of high-level classification with spatial reasoning in a heterogeneous 2-D and 3-D data space;
3) 3-D reconstruction (synthesis) of individual ISO containers.

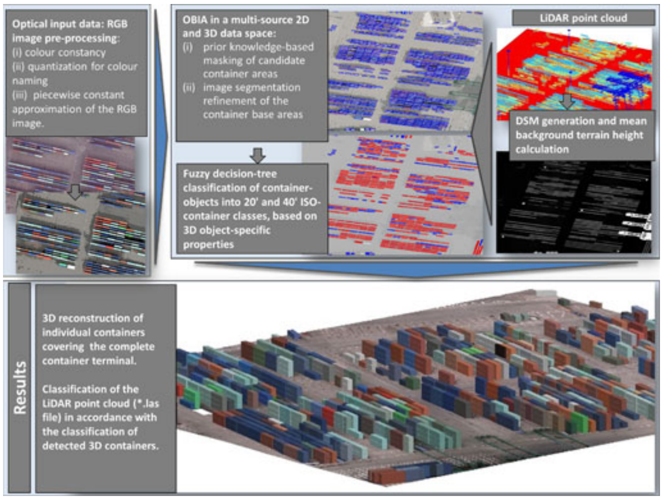
Optical input data: RGB image pre-processing:
(i) colour constancy
(ii) quantization for colour naming
(iii) piecewise constant approximation of the RGB image.
OBIA in a multi-source 2D and 3D data space:
(i) prior knowledge-based masking of candidate container areas
(ii) image segmentation refinement of the container base areas
Fuzzy decision-tree classification of container-objects into 20' and 40' ISO-container classes, based on 3D object-specific properties
LiDAR point cloud
DSM generation and mean background terrain height calculation
Results
3D reconstruction of individual containers covering the complete container terminal.
Classification of the LiDAR point cloud (*.las file) in accordance with the classification of detected 3D containers.

Fig. 10. Adopted workflow. (1) RGB image preprocessing (upper left) and LiDAR data preprocessing (top right). (2) Integration of two object-based image and point cloud analyses (center top). (3) Reconstruction (synthesis) of tangible 3-D container objects (bottom).

D. Automated RGB Image Preprocessing First Stage

1) RGB Color Constancy for Uncalibrated RGB Image Harmonization: Color constancy is a perceptual property of the human vision system ensuring that the perceived colors of objects in a (3-D) scene remain relatively constant under varying illumination conditions [79]. By augmenting image harmonization and interoperability without relying on radiometric calibration parameters, it was considered a viable alternative to the radiometric calibration (Cal) considered mandatory by the Quality Assurance Framework for Earth Observation (QA4EO) guidelines [80]. The QA4EO’s Cal principle requires dimensionless digital numbers to be transformed into a physical variable, provided with a radiometric unit of measure, by means of radiometric Cal metadata parameters [59]. Physical variables can be analyzed by both physical and statistical models, therefore hybrid models, too [55]. On the contrary, quantitative variables provided with no physical unit of measure can be investigated by statistical models exclusively [55]. An original automated (self-organizing) statistical model-based algorithm for multispectral (MS) color constancy, including RGB color constancy as a special case, was designed and implemented in-house (unpublished, patent pending).

2) Forward RGB Image Analysis by Prior Knowledge-Based Vector Quantization (VQ) and Inverse RGB Image Synthesis for VQ Quality Assessment: Widely investigated by the CV community [79], a finite and discrete dictionary of prior RGB color names is equivalent to a static (nonadaptive to data) RGB cube polyhedralization, where polyhedra can be arbitrary, either convex or not, either connected or not. In the seminal work by Griffin [81], the hypothesis was proved that the best partition of a monitor-typical RGB data cube into color categories for pragmatic purposes coincides with human basic colors (BCs) (see [81, p. 76]). Central to this consideration is Berlin and Kay's landmark study of color words in 20 human languages, where they claimed that the BC terms of any given language are always drawn from a universal inventory of eleven color names: black, white, gray, red, orange, yellow, green, blue, purple, pink, and brown [82]. These perceptual BC categories are expected to be "universal," i.e., users can apply the same universal color representation independently of the image-understanding problem at hand [79]. Equivalent to color naming in natural languages [82], prior knowledge-based color space discretization is the deductive automatic counterpart of inductive learning-from-data VQ algorithms (not to be confused with unsupervised data clustering algorithms [77]). In machine learning, the class of predictive VQ optimization problems requires to minimize a known VQ error function, typically a root-mean-square error (RMSE), where the number and location of VQ bins are the system free parameters [77]. For example, in the popular k-means VQ algorithm, the number of VQ levels, k, must be user-defined based on heuristics [83]. When they adopt a Euclidean metric distance in their minimization criterion and they reached convergence, inductive VQ algorithms accomplish a Voronoi tessellation of the input vector space, which is a special case of convex polyhedralization [84]. In contrast with inductive VQ algorithms capable of convex hyperpolyhedralizations, prior spectral knowledge-based decision trees can be designed to partition an input data hyperspace into hyperpolyhedra of any possible shape and size, either convex or concave, either connected or not. Unfortunately, when a data space dimensionality is superior to three, a prior partition of hyperpolyhedra is difficult to think of and currently impossible to visualize. This is the case of the Satellite Image Automatic Mapper (SIAM), an expert software system for MS color naming presented to the RS community in recent years [68]. By definition, expert systems require neither training datasets nor user-defined parameters to run, i.e., they are fully automated. Inspired by the SIAM expert system [68], a novel RGB Image Automatic Mapper (RGBIAM) was designed and implemented (unpublished, patent pending). RGBIAM is an expert software system for RGB cube partitioning into an *a priori* dictionary of RGB color names. By analogy with SIAM, since no total number k of VQ bins can be considered "best" (universal) in general, the implemented RGBIAM supports two coexisting VQ levels, fine and coarse, corresponding to 50 and 12 color names, respectively, provided with interdictionary parent–child relationships (see Fig. 11). Whereas the physical model-based SIAM requires as input a radiometrically calibrated MS image [80], the first-principle model-based RGBIAM requires as input an uncalibrated RGB image, in either true- or false-colors, preprocessed by a color constancy algorithm, to guarantee data harmonization and interoperability across images and sensors.

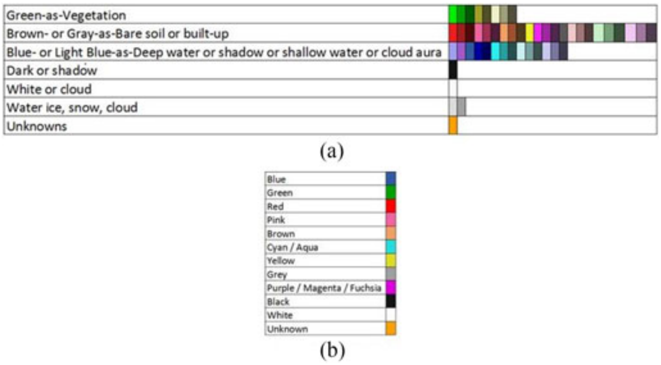

Fig. 11. RGBIAM is an expert system for a monitor-typical RGB color-space discretization into two prior quantization levels, consisting of (a) 50 color bins and (b) 12 color bins, linked by interlegend child–parent relationships. They are fixed *a priori* to be community-agreed upon.

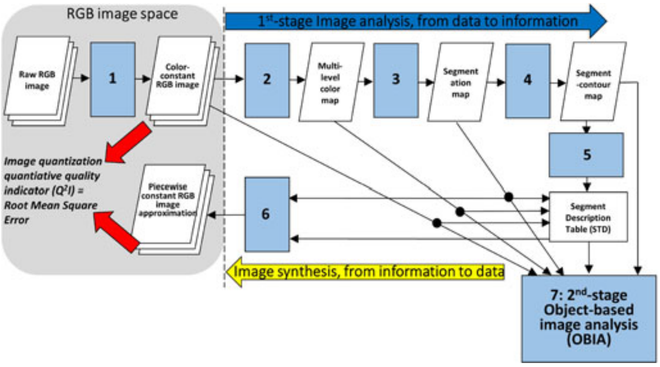
RGB image space
1st-stage Image analysis, from data to information
Raw RGB image
1
Color-constant RGB image
2
Multi-level color map
3
Segmentation map
4
Segment-contour map
5
Image quantization quantiative quality indicator (Q²I) = Root Mean Square Error
Piecewise constant RGB image approximation
6
Segment Description Table (STD)
Image synthesis, from information to data
7: 2nd-stage Object-based image analysis (OBIA)

Fig. 12. First-stage RGBIAM QNQ transform, consisting of six information processing blocks identified as 1 to 6, followed by a high-level object-based image analysis (OBIA) second stage, shown as block 7. Blocks 1–5 cope with direct image analysis. 1: Self-organizing statistical algorithm for color constancy. 2: Deductive VQ stage for prior knowledge-based RGB cube polyhedralization. 3: Well-posed two-pass connected-component multilevel image labeling. 4: Well-posed extraction of image-object contours. 5: Well-posed STD allocation and initialization. Block 6: inverse image synthesis, specifically, superpixel (piecewise)-constant RGB image approximation.

The automated RGBIAM pipeline for a quantitative-to-nominal-to-quantitative (QNQ) transform of a monitor-typical true- or false-color RGB image is shown in Fig. 12. It consists of: 1) a forward RGBIAM's Q-to-N variable transform. It maps an RGB image onto two multilevel color maps, whose legends consist of 50 and 12 color names (see Fig. 11). 2) An inverse RGBIAM's N-to-Q variable transform. It provides an RGB color VQ error estimation, specifically, an RMSE image estimation, in compliance with the QA4EO's requirements on validation [80]. To this end, each of the two RGBIAM's multilevel color maps was deterministically partitioned into an image segmentation map by a well-known two-pass algorithm for connected-component multilevel image labeling [75]. The RGBIAM's planar segments identified in the 2-D color map domain, consisting of connected pixels featuring the same color name, are traditionally known as texels (texture elements), textons [85], tokens [73], or superpixels in the recent CV literature [86]. In other words, RGBIAM works as a texel detector at the Marr's

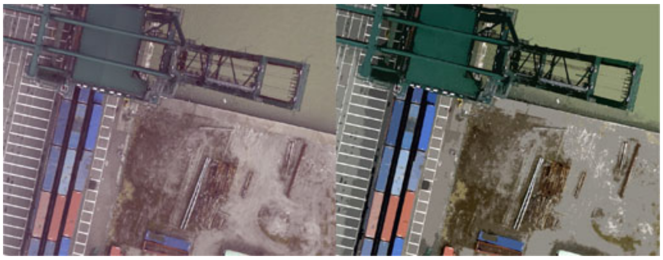

Fig. 13. (Left) Subset of the original uncalibrated RGB image, before color constancy. (Right) Same subset, after automatic RGB image preprocessing, consisting of image enhancement by statistical color constancy, RGBIAM's image mapping into 50 color names and RGBIAM's segment-constant edge-preserving image reconstruction. No image histogram stretching is applied for visualization purposes.

raw primal sketch in low-level (preattentional) vision. Next, for each RGB color quantization level, fine or coarse, a segment description table (SDT) was generated as a tabular representation of the texel information [67]. In an SDT estimated in one image pass, each texel was described by its positional (e.g., minimum enclosing rectangle), photometric (e.g., mean MS value), and geometric attributes (e.g., area). Finally, based on each pair of one SDT and one segmentation map, a texel-constant edge-preserving approximation of the input RGB image (mean value of the RGB image per texel object) was automatically generated in linear time (see Fig. 13). The comparison of the input RGB image with the output reconstructed RGB image allowed estimation of an RMSE image as a community-agreed quantitative quality indicator (Q^2I) in VQ problems [77], [80].

E. Second-Stage Classification With Spatial Reasoning in a Heterogeneous 2-D and 3-D Data Space

Traditionally mimicked by fuzzy logic [87], symbolic human reasoning is grounded on the transformation of a quantitative (numeric) variable, such as ever-varying sensations, into a qualitative (categorical, nominal) variable consisting of fuzzy sets, such as discrete and stable percepts [67]. The second-stage classification of our hybrid CV system was input with five geospatial variables, either quantitative or categorical [88]:

1) A quantitative 3-D LiDAR point cloud.
2) A quantitative LiDAR data-derived DSM.
3) A quantitative piecewise-constant edge-preserving simplification of the original RGB image (see Fig. 13).
4) Two categorical and semisymbolic RGBIAM preclassification maps into color names of the input RGB image (see Fig. 11).
5) Two categorical subsymbolic RGB image segmentation maps, consisting of planar objects automatically extracted from the two multilevel preclassification maps, provided with intermap parent–child relationships.

To infer output variables of higher information quality from the combination of numeric and categorical geospatial variables, the second stage of our hybrid CV system adopted two strategies:

1) A "stratified" approach to quantitative variable analysis, where geospatial numeric variables are conditioned by geospatial categorical variables in compliance with the principle of statistical stratification [67], [68], [78].
2) An OBIA approach to spatial symbolic reasoning on geospatial categorical variables, consisting of discrete geometric objects, either 2-D or 3-D, by means of physical model-based syntactic decision trees [75].

Unlike many traditional data fusion techniques, where geospatial numeric (quantitative) variables are stacked before 1-D (context-insensitive topology-independent) image analysis, typically by means of supervised data classification techniques, the proposed CV system investigates geospatial numeric variables (e.g., height values provided by the LiDAR point cloud) conditioned by geospatial categorical (nominal) variables in the (2-D) image domain (e.g., RGB texels), in compliance with the principle of statistical stratification and the OBIA paradigm. The implemented classification second stage of our hybrid CV system consisted of five subsystems, coded in the Cognition Network Language, within the eCognition Developer software environment (Trimble Geospatial).

1) Convergence-of-Evidence Criterion for Automated Background Terrain Extraction From the DSM and the RGB Image: An automated (well-conditioned) eCognition multiresolution image segmentation algorithm [89] was run to extract planar objects in the DSM image featuring within-object nearly constant DSM values. Merging adjacent 2-D objects whose height differences was below 1 m resulted in, among others, one very large planar object, corresponding to the dominating background terrain across the surface area. Next, color names of background surface types, such as asphalt, bare soil, or water, were visually selected, combined by a logical OR operator and overlapped with the DSM-derived background mask. Finally, a foreground binary mask was generated as the inverse of the background binary mask. Foreground planar objects were candidates for 3-D ISO container detection.

2) Candidate 3-D Object Selection Based on Converging 2-D and 3-D Data-Derived Information: Masked by the foreground binary mask detected at step 1, the RGBIAM's planar objects (texels) detected in the RGB image domain were considered as input geospatial information primitives. In the orthophoto domain, the top view of a 3-D ISO container looked like a single foreground image object provided with its RGBIAM's color name. To assign an object-specific height value z to each foreground planar object, the tilting effect observed in the orthorectified RGB image (discussed in Section IV-B) had to be accounted for. To this end, the height value was estimated as the 90% quantile of the LiDAR point cloud's z elevation values whose (x, y) coordinates fell on the target planar object. Foreground image objects with an estimated height higher than a physical model-based maximum height of 26 m, corresponding to ten stacked ISO containers (considered as the possible maximum stacking height [90]), were removed from the set of 3-D container-candidate objects, such as image objects related to cranes in the scene domain. Finally, a spatial decision rule exploiting interobject spatial relationships was applied to mask out small-size nonelevated 3-D objects, whose planar projection was below the minimum ISO container area and that were isolated, i.e., surrounded by background areas exclusively. The result was a binary mask of container-candidate planar areas, where containers could be stacked up to ten layers.

3) Driven-by-Knowledge Refined Segmentation of the RGB Image: Masked by the binary candidate-container image areas detected in step 2, the edge-preserving smoothed RGB image (see Fig. 13) was input into a well-posed multiresolution eCognition segmentation algorithm [89], whose free-parameter "planar shape compactness" was selected in accordance with prior physical knowledge of the ISO container's length and width [65]. Unlike the first RGBIAM's image partition, based on a nonadaptive-to-data spectral knowledge base and applied image-wide, this second adaptive-to-data image segmentation algorithm was provided with physical constrains and run on a masked image subset (container area only), to make it less prone to inherent segmentation errors and faster to compute. This stage accounted for the LiDAR data-derived DSM image indirectly through the input binary mask, rather than directly by stacking it with the input RGB image, to avoid the aforementioned tilting effect.

4) Three-Dimensional ISO Container Recognition and Classification: Classes of ISO containers in the scene domain were described in user-speak by the following shape and size properties [65]:

1) ISO class 1 of 20' container: rectangular area of 15 m^2, rectangular shape, height 2.6 m.
2) ISO class 2 of 40' container: rectangular area of 30 m^2, rectangular shape, height 2.6 m.

These size values were projected onto the image domain in technospeak, based on a sensor-specific transformation function [67]. The planar projection of a 3-D rectangular object belonging to the ISO container classes 1 and 2 was found to match an eCognition's image object-specific rectangular fit index of 0.75 in range [0, 1]. Input image objects, detected at the previous step 3, were selected based on their fuzzy rectangular shape membership value. Surviving image objects whose height was divided by the standard height of an ISO container, equal to 2.6 m, provided an estimate of the number of stacked ISO containers per image object. Finally, vector container objects were assigned to ISO container classes 1 or 2 based on their length/width relation. In addition, for visualization and 3-D reconstruction purposes, an eCognition function was run per container object to simplify and orthogonalize vector object boundaries in agreement with the main direction identified as the angle featuring the largest sum of object edges per container object. Classified and 3-D container objects were exported as polygon vectors in a standard GIS-ready file format (e.g., *.shp).

5) Semantic Labeling of the 3-D LiDAR Point Cloud: Geospatial locations of the container objects classified in step 4 were spatially intersected with the 3-D LiDAR point cloud, to provide semantic labels to the LiDAR's z values whose (x, y) spatial coordinates fell on the planar projection of a 3-D container.

F. Three-Dimensional Reconstruction of ISO Containers

The synthesis of tangible 3-D container objects took place in a GIS commercial software product (ArcScene, ESRI). The RGB orthophoto was draped over the DSM. Vector 3-D container objects were extruded according to their relative height. Stacked containers were extruded several times, based on the estimated number of stacked containers. Stacked containers were visualized in the same RGB color value estimated for the container on top (see Fig. 14).

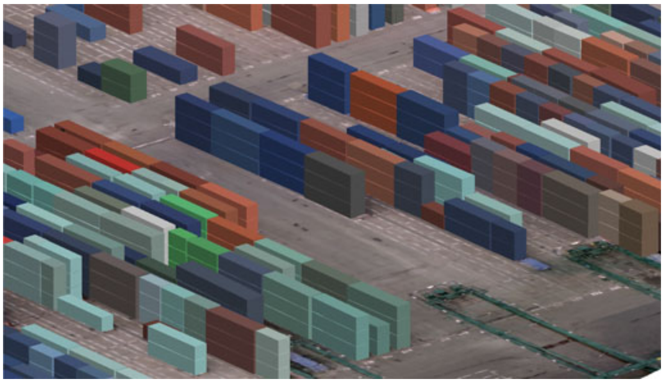

Fig. 14. Subset of the 3-D reconstructed container terminal. Each container is a tangible 3-D object featuring positional, colorimetric, geometric, and identification attributes. Stacked containers were visualized in the same color of the container on top, according to the per-object mean RGB value extracted from the RGB orthophoto.

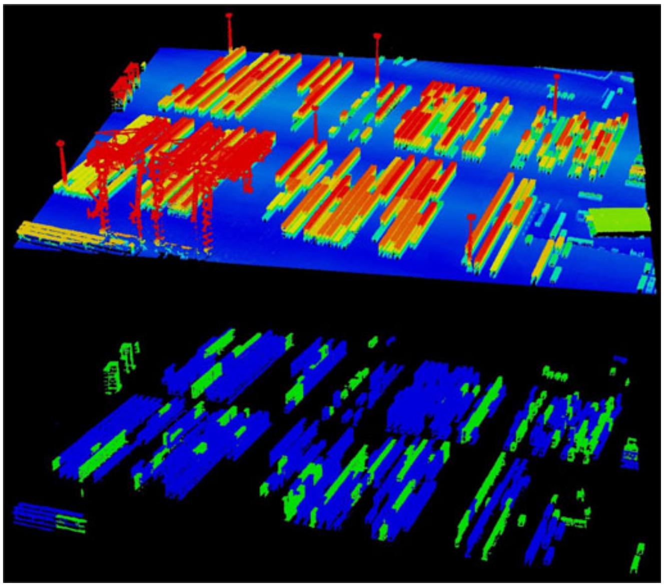

Fig. 15. Above: original LiDAR point cloud, colored from blue to red according to increasing point elevation values. Below: detected and classified LiDAR container point, where green = 20 container class 1, blue = 40 container class 2 [scene extent: (275 × 325 m)].

G. Results and Discussion

The implemented hybrid CV system ran automatically, because prior knowledge initialized inductive 2-D and 3-D data analysis without user interaction, and near real-time, because of its linear complexity with the data size. In a standard laptop computer, computation time was 1 min for the RGBIAM preclassification and less than 5 min for the second-stage OBIA

TABLE II
ACCURACY ASSESSMENT OF THE EXTRACTED CONTAINER TYPES

Container type	Automatic Assessment (2-D)	Visual Assessment (2-D)	Matches	Producer's Accuracy (%)	User's Accuracy (%)
40'	425	426	415	97.42	97.65
20'	265	226	217	96.02	81.89
Aggregated	690	652	632	96.93	91.59

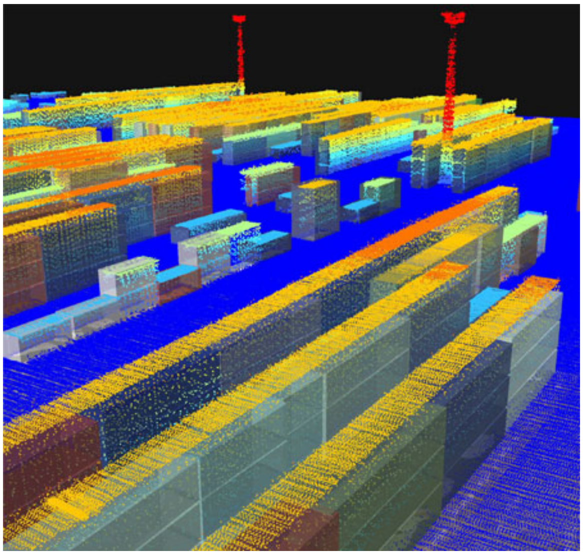

Fig. 16. Overlay of the 3-D container reconstruction with the original LiDAR point cloud, colored from blue over orange to red according to increasing point elevation values. It qualitatively shows the estimated height per container stack looked accurate.

classification. It detected 1659 containers distributed in 690 container stacks, including 118 single containers, 201 stacks of 2, 361 stacks of 3, 2 stacks of 4, and 8 stacks of 6 containers (see Figs. 14 and 15). In comparison with a reference "container truth," acquired by an independent photointerpreter from the RGB image, detected ISO containers revealed a high overall accuracy (see Table II), together with producer's and user's accuracies superior to 96%, although the user's accuracy for the 20' container class, affected by 39 false-positive occurrences, scored 81%. Further investigation revealed that the majority of these false positives were due to either true containers (e.g., few 40' container objects were recognized as pairs of 20' container objects due to spectral disturbances) or real-world objects similar to ISO containers in shape and size (e.g., trucks). A qualitative comparison of the 3-D container reconstruction with the original LiDAR point cloud revealed a qualitatively "high" accuracy of height estimation per container stack (see Fig. 16).

In line with theoretical expectations about hybrid inference, these experimental results reveal that the implemented hybrid CV system can be considered in operating mode by scoring "high" in a set of minimally redundant and maximally informative Q^2Is. Estimated Q^2I values encompassed [68] accuracy ("high"; see Table II and Fig. 16), efficiency ("high," linear time), degree of automation ("high," no user interaction), timeliness from data acquisition to product generation including training data collection, to be kept low ("low," no training-from-data, physical models were intuitive to tune, etc.), and scalability to different CV problems ("high," the CV system pipeline is data- and application-independent up to the CV system's target-specific classification second stage). The conclusion is that hybrid feedback CV system design and implementation strategies can contribute to tackle the increasing demand for off-the-shelf software products capable of filling the semantic information gap from 2-D and 3-D big sensory data to high-level geospatial information products, where 2-D data are typically uncalibrated, such as images acquired by consumer-level color cameras mounted on mobile devices, including unmanned aircraft systems.

V. DISCUSSION OF THE 3-D CONTEST: THE WINNERS

The complex novel approaches proposed by the two winning teams included image modeling and processing tools for fusing 2-D and 3-D information and overcoming the different geometrical scanning modes and the limitations of each individual dataset.

1) The winning team considered the problem of road extraction (see Section III). As discussed above, automatic extraction of roads in complex urban scenes using RS sources is an open and challenging research topic, less active in comparison to the detection of buildings or trees. The innovative processing solution proposed by the winning team was found remarkable for two reasons: 1) it takes into account the characteristics of the modern urban landscape that includes many different road materials and types but also road obstacles such as speed bumps and road curbs; and 2) it applies and combines methods on different levels of processing including the laser's point clouds, grids, and regions. This approach supports the importance of LiDAR as a highly accurate RS source for road and object detection in urban areas. The integration of optical data is mainly used to exclude vegetal features such as grassy areas as well as to reduce the ambiguity in areas near the buildings.
2) The runner-up team addressed the detection of containers in the harbor of Zeebruges (see Section IV). The proposed processing scheme is complex and was conducted from three different levels of processing: enhancement, classification, and 3-D reconstruction. Main innovations are in the first and second steps. Image harmonization is a necessary condition for the implementation of a hybrid CV system, in which physical and statistical data models are combined to take advantage of each other and overcome their shortcomings. The hybrid system runs automatically without user interaction, does not involve learning of model parameters from empirically labeled data, and relies on *a priori* knowledge to account for contextual and topological visual information. This processing chain highlights the efficiency of object-based methods applied

for fused optical and LiDAR data for the detection of small objects in the urban-harbor environment.

VI. Conclusion on the Data Fusion Contest 2015

In this two-part paper, we presented and discussed the outcomes of the IEEE GRSS Data Fusion Contest 2015. In compliance with the two-track structure of the contest, we discussed the results in two parts: the 2-D contest in Part A [1] and the results of the 3-D contest in Part B (this paper). The winners of both tracks showed innovative ways of dealing with the very timely sources of data proposed: extremely high resolution RGB data and high density LiDAR point clouds.

In the 2-D contest, the emerging technology of convolutional neural networks, which is becoming a very prominent standard in computer vision, has emerged as a powerful and effective way of extracting knowledge from these complex data. The understanding of the information learned by the network is highlighted by the winning team, which provided an in-depth analysis of the properties of the deep network filters. The effectiveness in classifying land cover types was highlighted by the runner-up team, which provided a comprehensive and thorough benchmark and disclosed their evaluation ground truth to the community.[3]

In the 3-D contest, the need of working with the point cloud directly emerged as a clear need, since the precision of the DSM used for calculation proved to be fundamental for the detection tasks addressed by the participating teams. Solutions to computational problems were also deeply considered by the winning team, since high-density LiDAR point cloud calls for new standards of storage and access to data. The runner-up team showed a different kind of reasoning for data fusion, not based solely on learning from examples, but combining it, in a hybrid system, with physical knowledge of the 3-D scene and psychophysical evidence about human vision. They showed that fully automatic recognition was possible, even with uncalibrated data as the RGB image provided, in combination with a dense 3-D LiDAR dataset.

Summing up, the organizers were extremely pleased by the quality of the solutions proposed and by the variety of fusion problems addressed and processing approaches adopted by the participants to the Contest. They ranged from cutting-edge machine learning methodologies to case-specific processing pipelines and to prior knowledge-based systems, with the goal of capturing the information conveyed by extremely high resolution 2-D and 3-D remote sensing data. The organizers do hope that the concepts emerging from the 2015 Data Fusion Contest will inspire new research studies at the interface of computer vision and remote sensing and foster new joint uses of laser and optical data.

Acknowledgment

The authors would like to express their greatest appreciation to the Department of Communication, Information, Systems and Sensors of Belgian Royal Military Academy, for acquiring and providing the data used in the competition and for indispensable contribution to the organization of the Contest, and the IEEE GRSS for continuously supporting the annual Data Fusion Contest through funding and resources.

[3] The ground truth, as well as the complete data of the contest, can be downloaded at http://www.grss-ieee.org/community/technical-committees/data-fusion/2015-ieee-grss-data-fusion-contest/

References

[1] M. Campos-Taberner *et al.*, "Processing of extremely high resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS Data Fusion Contest—Part A: 2D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, submitted for publication.

[2] T. Lakes, P. Hostert, B. Kleinschmit, S. Lauf, and J. Tigges, "Remote sensing and spatial modelling of the urban environment," in *Perspectives in Urban Ecology*, W. Endlicher, Ed. Berlin, Germany: Springer, 2011.

[3] J. Jung, E. Pasolli, S. Prasad, J. Tilton, and M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 491–502, Feb. 2014.

[4] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.

[5] W. Liao *et al.*, "Processing of thermal hyperspectral and digital color cameras: Outcome of the 2014 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2984–2996, Jun. 2015.

[6] G. Petrie and C. K. Toth, "Introduction to laser ranging, profiling, and scanning," in *Topographic Laser Ranging and Scanning: Principles and Processing*, J. Shan and C. K. Toth, Eds. Boca Raton, FL, USA: CRC Press, 2008, pp. 6–24.

[7] A. Habib, "Accuracy, quality assurance, and quality control of LiDAR data," in *Topographic Laser Ranging and Scanning: Principles and Processing*, J. Shan and C. K. Toth, Eds. Boca Raton, FL, USA: CRC Press, 2008, pp. 247–360.

[8] A. Fowler and V. Kadatskiy, "Accuracy and error assessment of terrestrial, mobile and airborne LIDAR," presented at the *ASPRS Annu. Conf.*, Milwaukee, WI, USA, May 1–5, 2011, 2011.

[9] K. Schmid, "LiDAR 101: An introduction to LiDAR technology, data, and applications. revised," NOAA Coastal Serv. Center, Charleston, SC, USA, [Online]. Available: https://coast.noaa.gov/data/digitalcoast/pdf/lidar-101.pdf, 2012.

[10] J. Li and H. Guan, "3D building reconstruction from airborne LiDAR point clouds fused with aerial imagery," in *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*, X. Yang, Ed. Hoboken, NJ, USA: Wiley, 2011.

[11] A. S. Antonarakis, K. Richards, and J. Brasington, "Object-based land cover classification using airborne LiDAR," *Remote Sens. Environ.*, vol. 112, pp. 2988–2998, 2008.

[12] R. Narwade and V. Musande, "Automatic road extraction from airborne LiDAR: A review," *Int. J. Eng. Res. Appl.*, vol. 4, no. 12, pp. 54–62, 2014.

[13] X. Hu, Y. Li, J. Shan, J. Zhang, and Y. Zhang, "Road centerline extraction in complex urban scenes from LiDAR data based on multiple features," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7448–7456, Nov. 2014.

[14] Y. Chen, L. Cheng, M. Li, J. Wang, L. Tong, and K. Yang, "Multiscale grid method for detection and reconstruction of building roofs from airborne LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4081–4094, Oct. 2014.

[15] F. Lafarge and C. Mallet, "Creating large-scale city models from 3D-point clouds: A robust approach with hybrid representation," *Int. J. Comput. Vision*, vol. 99, no. 69–85, 2012.

[16] A. Börcs and C. Benedek, "Extraction of vehicle groups in airborne LiDAR point clouds with two-level point processes," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1475–1489, Mar. 2015.

[17] M. Shimoni, G. Tolt, C. Perneel, and J. Ahlberg, "Detection of vehicles in shadow areas using combined hyperspectral and LiDAR data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp.*, 2011, pp. 4427–4430.

[18] K. Singh, J. Vogler, D. Shoemaker, and R. Meentemeyer, "LiDAR-landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy," *ISPRS J. Photogramm. Remote Sens.*, vol. 74, pp. 110–121, 2012.

[19] S. Luo *et al.*, "Fusion of airborne discrete-return LiDAR and hyperspectral data for land cover classification," *Remote Sens.*, vol. 8, p. 3, 2016.

[20] C. Paris and L. Bruzzone, "A three-dimensional model-based approach to the estimation of the tree top height by fusing low-density LiDAR data and very high resolution optical images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 467–480, Jan. 2015.

[21] J. Secord and A. Zakhor, "Tree detection in urban regions using aerial LiDAR and image data," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 196–200, Apr. 2007.

[22] M. Bandyopadhyay, J. A. N. van Aardta, and K. Cawse-Nicholson, "Classification and extraction of trees and buildings from urban scenes using discrete return LiDAR and aerial color imagery," *Proc. SPIE*, vol. 8731, p. 873105, 2013.

[23] G. Zhou and X. Zhou, "Seamless fusion of LiDAR and aerial imagery for building extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7393–7407, Nov. 2014.

[24] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 2, pp. 135–149, 2007.

[25] A. Brook, M. Vandewal, and E. Ben-Dor, "Fusion of optical and thermal imagery and LiDAR data for application to 3-D urban environment and structure monitoring," in *Remote Sensing—Advanced Techniques and Platforms*, B. Escalante, Ed. Rijeka, Croatia: InTech, 2012, pp. 29–50.

[26] F. Rottensteiner and S. Clode, "Building and road extraction by LiDAR and imagery," in *Proc. Topographic Laser Ranging and Scanning: Principles and Processing*, 1st ed., J. Shan, J. Shan, C. K. Toth, and C. K. Toth, Eds. Boca Raton, FL, USA: CRC Press, 2008, pp. 445–478.

[27] Y.-W. Choi, Y.-W. Jang, H.-J. Lee, and G.-S. Cho, "Three-dimensional LiDAR data classifying to extract road point in urban area," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 725–729, Oct. 2008.

[28] P. Zhu, Z. Lu, X. Chen, K. Honda, and A. Elumnoh, "Extraction of city roads through shadow path reconstruction using laser data," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 12, pp. 1433–1440, 2004.

[29] A. Habib, M. Ghanma, M. Morgan, and R. Al-Ruzouq, "Photogrammetric and LiDAR data registration using linear features," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 6, pp. 699–707, 2005.

[30] S. Urala, J. Shana, M. A. Romero, and A. Tarko, "Road and roadside feature extraction using imagery and LiDAR data for transportation operation," presented at the ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci., Joint ISPRS Conf. PIA15+HRIGI15, Munich, Germany, Mar. 25–27, 2015.

[31] J. Hang, "An integrated approach for precise road reconstruction from aerial imagery and LiDAR data," Ph.D. dissertation, , Queensland Univ. Technol., Brisbane, Australia, 2011.

[32] L. Zhou, "Fusing laser point cloud and visual image at data level using a new reconstruction algorithm," in *Proc. IEEE Intell. Veh. Symp.*, 2013, pp. 1356–1361.

[33] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained partbased models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[34] R. Tylecek and R. Šára, "Spatial pattern templates for recognition of objects with regular structure," *Pattern Recog.*, vol. 8142, pp. 364–374, 2013.

[35] I. Stamos and P. K. Allen, "Geometry and texture recovery of scenes of large scale," *Comput. Vis. Image Underst.*, vol. 88, no. 2, pp. 94–118, 2002.

[36] I. Stamos, L. Liu, C. Chen, G. Wolberg, G. Yu, and S. Zokai, "Integrating automated range registration with multi-view geometry for the photorealistic modeling of large-scale scenes," *Int. J. Comput. Vision*, vol. 78, pp. 237–260, 2008.

[37] H. Kim, C. D. Correa, and N. Max, "Automatic registration of LiDAR and optical imagery using depth map stereo," presented at the IEEE Int. Conf. Comput. Photography, Santa Clara, CA, USA, 2014.

[38] C. Frueh, R. Sammon, and A. Zakho, "Automated texture mapping of 3D city models with oblique aerial imagery," in *Proc. Int. Symp. 3D Data Process., Visual. Trans.*, Tokyo, Japan, 2004, pp. 396–403.

[39] L. Liu, I. Stamos, G. Yu, G. Wolberg, and S. Zokai, "Multiview geometry for texture mapping 2D images onto 3D range data," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, New York, NY, USA, 2006, pp. 2293–2300.

[40] T. Schenk and B. Csatho, "Fusion of lidar data and aerial imagery for a more complete surface description," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 34, no. 3A, pp. 301–317, 2002.

[41] K. Fujii and T. Arikawa, "Urban object reconstruction using airborne laser elevation image and aerial image," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2234–2240, Oct. 2002.

[42] J. Xu, K. Kim, Z. Zhang, H.-W. Chen, and O. Yu, "2D/3D sensor exploitation and fusion for enhanced object detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recog. Workshops*, 2014, pp. 778–784.

[43] A.-V. Vo, L. Truong-Hong, and D. Laefer, "Aerial laser scanning and imagery data fusion for road detection in city scale," presented at the Int. Geosci. Remote Sens. Symp., Milan, Italy, 2015.

[44] D. Tiede, S. d'Oleire Oltmanns, and A. Baraldi, "Geospatial 2D and 3D object-based classification and 3D reconstruction of ISO-containers depicted in a LiDAR data set and aerial imagery of a harbor," presented at the Int. Geosci. Remote Sens. Symp., Milan, Italy, 2015.

[45] J. M. Collado, C. Hilario, A. De LaEscalera, and J. M. Armingol, "Detection and classification of road lanes with a frequency analysis," in *Proc. IEEE Intell. Veh. Symp.*, 2005, pp. 78–83.

[46] A. Boyko and T. Funkhouser, "Extracting roads from dense point clouds in large scale urban environment," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. S2–S12, 2011.

[47] X. Hu, C. V. Tao, and Y. Hu, "Automatic road extraction from dense urban area by integrated processing of high resolution imagery and LiDAR data," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 35, p. B3, 2004.

[48] Y.-W. Choi, Y. W. Jang, H. J. Lee, and G.-S. Cho, "Heuristic road extraction," in *Proc. Int. Symp. Inf. Technol. Convergence*, 2007, pp. 338–342.

[49] A. Alharthy and J. Bethel, "Automated road extraction from LiDAR data," in *Proc. ASPRS Annu. Conf.*, 2003, pp. 05–09.

[50] S. Clode, P. Kootsookos, and F. Rottensteiner, "The automatic extraction of roads from LiDAR data," in *Proc. ISPRS Annu. Congr.*, 2004, vol. 35, pp. 231–237.

[51] S. Clode, F. Rottensteiner, P. Kootsookos, and E. Zelniker, "Detection and vectorization of roads from LiDAR data," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 5, pp. 517–535, 2007.

[52] J. Skilling, "Programming the Hilbert curve," in *Proc. 23rd Int. Workshop Bayesian Inference Maximum Entropy Methods Sci. Eng.*, 2004, vol. 707, pp. 381–387.

[53] P. V. Oosterom and T. Vijlbrief, "The spatial location code," presented at the Int. Symp. Spatial Data Handling, Delft, The Netherlands, 1996.

[54] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The Weka data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

[55] B. Höfle and N. Pfeifer, "Correction of laser scanning intensity data: Data and model-driven approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 6, pp. 415–433, Dec. 2007.

[56] J.-e. Deschaud and F. Goulette, "A fast and accurate plane detection algorithm for large noisy point clouds using filtered normals and voxel growing," presented at the 3D Process., Visual. Trans. Conf., Paris, France, 2010.

[57] J. R. Quinlan, *C4. 5: Programming for Machine Learning*. San Mateo, CA, USA: Morgan Kauffmann, 1993.

[58] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Jan. 2008.

[59] G. Sithole, "Detection of bricks in a masonry wall," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XXXVII, part B5, Beijing, 2008, pp. 1–6.

[60] S. Müller and D. W. Zaum, "Robust building detection in aerial images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 36, no. B2/W24, pp. 143–148, 2005.

[61] L. Guo, N. Chehata, C. Mallet, and S. Boukir, "Relevance of airborne LiDAR and multispectral image data for urban scene classification using Random Forests," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 1, pp. 56–66, Jan. 2011.

[62] A.-V. Vo, L. Truong-Hong, D. F. Laefer, and M. Bertolotto, "Octree-based region growing for point cloud segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 88–100, 2015.

[63] D. F. Laefer, C. O'Sullivan, H. Carr, and L. Truong-Hong, "Aerial laser scanning (ALS) data collected over an area of around 1 square km in Dublin city in 2007," 2014. [Online]. Available: http://digital.ucd.ie/view/ucdlib:30462

[64] T. Hinks, H. Carr, H. Gharibi, and D. F. Laefer, "Visualisation of urban airborne laser scanning data with occlusion images," *ISPRS J. Photogram. Remote Sens.*, vol. 104, pp. 77–87, 2015.

[65] *ISO 668:2013 Series 1 Freight Containers—Classification, Dimensions and Ratings*. 2015. [Online]. Available: https://www.document-center.com/standards/show/ISO-668

[66] H. du Buf and J. Rodrigues, "Image morphology: From perception to rendering," *J. Interdisciplinary Image Sci.*, vol. 5, pp. 1–19, 2007.

[67] T. Matsuyama and V.-S.-S. Hwang, *SIGMA: A Knowledge-based Aerial Image Understanding System*. New York, NY, USA: Springer-Verlag, 2013.

[68] A. Baraldi, L. Durieux, D. Simonetti, G. Conchedda, F. Holecz, and P. Blonda, "Automatic spectral rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery—Part I: System design and implementation," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1299–1325, Mar. 2010.

[69] T. Blaschke *et al.*, "Geographic object-based image analysis—Towards a new paradigm," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, no. 100, pp. 180–191, 2014.

[70] D. Tiede, S. Lang, F. Albrecht, and D. Hölbling, "Object-based class modeling for cadastre-constrained delineation of geo-objects," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 2, pp. 193–202, 2010.

[71] D. Tiede, "A new geospatial overlay method for the analysis and visualization of spatial change patterns using object-oriented data modeling concepts," *Cartography Geographic Inf. Sci.*, vol. 41, no. 3, pp. 227–234, May 2014.

[72] T. Martinetz and K. Schulten, "Topology representing networks," *Neural Netw.*, vol. 7, no. 3, pp. 507–522, 1994.

[73] D. Marr, *Vision*, 1st ed. San Francisco, CA, USA: Freeman, 1982.

[74] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ, USA: Wiley, 2004.

[75] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*, 1st ed. Cambridge, U.K.: CL Engineering, 1993.

[76] S. Frintrop, *Computer Analysis of Human Behavior*, A. A. Salah and T. Gevers, Eds. London, U.K.: Springer, 2011.

[77] V. S. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*, 1st ed. New York, NY, USA: Wiley, 1998.

[78] P. M. Mather, *Computer Processing of Remotely Sensed Images: An Introduction*. New York, NY, USA: Wiley, 1992.

[79] T. Gevers, A. Gijsenij, J. van de Weijer, and J.-M. Geusebroek, *Color in Computer Vision: Fundamentals and Applications*. New York, NY, USA: Wiley, 2012.

[80] GEO and CEOS, "Group on Earth Observations (GEO)/Committee on Earth Observation Satellites (CEOS): A quality assurance framework for earth observation (QA4EO) implementation strategy and work plan March 2012," [Online]. Available: http://qa4eo.org/docs/qa4eo_implementation_strategy_for_ceos_and_geo_v0_4.pdf, Mar. 2012.

[81] L. D. Griffin, "Optimality of the basic colour categories for classification." *J. Roy. Soc., Interface*, vol. 3, no. 6, pp. 71–85, 2006.

[82] B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution (ser. The David Hume Series)*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[83] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, no. 1, pp. 84–95, Jan. 1980.

[84] B. Fritzke, "Some competitive learning methods," 1997. [Online]. Available: http://www.demogng.de/JavaPaper/t.html

[85] B. Julesz, "Texton gradients: The texton theory revisited," *Biol. Cybern.*, vol. 54, nos. 4/5, pp. 245–251, Aug. 1986.

[86] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach, Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[87] L. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 1, no. 1, pp. 3–28, Jan. 1978.

[88] R. Capurro and B. Hjørland, "The concept of information," *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 343–411, Jan. 2005.

[89] M. Baatz and A. Schäpe, "Multiresolution segmentation—An optimization approach for high quality multi-scale image segmentation" in *Angewandte Geographische Informationsverarbeitung*, J. Strobl, T. Blaschke, and G. Griesebner, Eds. Heidelberg, Germany: Wichmann-Verlag, 2000, pp. 12–23.

[90] Y. Wild, R. Scharnow, and M. Rühmann, *Container Handbook*, 1st ed. Berlin, Germany: Gesamtverband der Deutschen Versicherungswirtschaft e.V., 2005.

Authors' photographs and biographies not available at the time of publication.