

A MULTI-PURPOSE BENCHMARK FOR PHOTOGRAMMETRIC URBAN 3D RECONSTRUCTION IN A CONTROLLED ENVIRONMENT

E. Özdemir^{1,2}, I. Toschi², F. Remondino²

¹ Space Centre, Skolkovo Institute of Technology (SKOLTECH), Moscow, Russia

² 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy

Email: <eoazdemir><toschi><remondino>@fbk.eu

Web: <http://3dom.fbk.eu>

Commission I, WGI/10

KEY WORDS: benchmark, photogrammetry, orientation, matching, classification, 3D building reconstruction, evaluation, accuracy

ABSTRACT:

This paper presents a novel multi-purpose benchmark for assessing the performance of the entire image-based pipeline for 3D urban reconstruction and 3D data classification. The proposed benchmark is called “3DOMcity” photogrammetric contest and is publicly available at <https://3dom.fbk.eu/3domcity-benchmark>. Compared to previous benchmark initiatives, the innovative aspects introduced by 3DOMcity are threefold. First, it covers the entire 3D reconstruction pipeline, including image orientation, dense image matching, point cloud classification and 3D building reconstruction. Second, the performance assessment is performed within a metrological context. Finally, the provided datasets offer 2D and 3D data collected at a very high spatial resolution. Furthermore, the benchmark is multi-purpose since it involves a multiplicity of tasks, that can be either performed independently from each other’s, or grouped together. The contributions of this paper are (i) the presentation of the novel benchmark dataset and corresponding reference data, (ii) the description of the multi-purpose benchmark tasks and evaluation procedures, and (iii) the discussion of first results of image orientation, dense image matching and classification.

1. INTRODUCTION

Over the past two decades the need for up-to-date 3D data in urban scenarios has been motivating the geospatial community towards the development of advanced 3D reconstruction methods, with special focus on automation, completeness and accuracy. 2.5D digital surface models (DSM) and 3D point clouds are normally used as input data for city modelling (Haala and Kada, 2010). They can be either captured by active polar measurement systems, such as airborne LiDAR (light detection and ranging) or derived by applying photogrammetric methods to a set of airborne images. Recently, a great impetus to urban 3D mapping has been given by the revival of oblique imaging technology for geospatial applications. Since the turn of the century, the airborne market for area-wide data acquisition has enthusiastically taken up oblique multi-camera systems, that are now included in the portfolio of almost all existing geospatial companies, either as stand-alone solutions (Remondino and Gerke, 2015) or in combination with a LiDAR unit (Toschi et al., 2018). Their advantage over traditional nadir-looking cameras lies in the slanted view geometry, that allows building facades and other urban vertical objects to be visible and thus, potentially, reconstructed (Haala et al., 2015; Toschi et al., 2017). Indeed, with recent advances in dense surface generation from imagery, the geometric processing including high density image matching, filtering and meshing, can now be performed in “true” 3D space, thus fully exploiting the vantage viewpoint offered by oblique image acquisitions (Haala and Rothermel, 2015). However, this all comes at a cost, i.e. extending traditional photogrammetric approaches to cope with new challenges in airborne applications, e.g.: large variations in image scale and illumination, multiple occlusions, an increased disparity search space and higher variances of observations quality across depth maps.

Once structured or unstructured point clouds are available, 3D city modelling approaches normally tackle the reconstruction problem by semantically analysing the pointwise data. Indeed, 3D scene understating allows for the automatic extraction of meaningful urban information, thus reducing the need for external ancillary information, like building footprints, DSM and DTM (digital terrain model). Many approaches for point cloud

classification exist, varying from classic machine learning methods such as SVM (support vector machine) and RF (random forest), to deep learning approaches for classification, object detection and point-wise segmentation (Griffiths and Boehm, 2019). In both cases, a key challenge is still the lack of adequate, high-quality and generalized training data, as well as the extremely varying and non-uniform distribution of the input data. In a last step, urban objects like buildings and trees are then represented in 3D using template(s) fitting and meshing approaches. Several methods have been developed to automatically get the 3D geometry of varying urban environments at the required level of abstraction (Haala and Kada, 2010). However, most of them still rely on strong urban priors (e.g. symmetry and roof typology) and on pre-defined template libraries, that can hardly be generalized to urban scenes with high variations in roof shapes, as well as to input data with different point densities.

1.1 Related benchmark activities

These recent potentials and challenges in the sector of urban 3D reconstruction have motivated the adoption of several scientific initiatives, aimed at evaluating the current status of the available processing tools, and at boosting further research in this field. Particularly, three ISPRS/EuroSDR benchmark activities have recently provided publicly available 2D/3D data and corresponding ground truth, to evaluate the performance of:

- aerial triangulation (AT) and dense image matching (DIM) algorithms within airborne oblique acquisition scenarios (Cavegn et al., 2014; Nex et al., 2015);
- urban classification and 3D building reconstruction approaches when applied to different urban landscapes (Rottensteiner et al., 2012).

These efforts, that encouraged developers and users to deliver comparative performance analyses on the provided datasets, are part of a broader landscape of benchmark actions, that have been assessing the performance of image-based 3D geometry reconstruction and scene classification algorithms over the last twenty years. An overview of these state-of-the-art initiatives

undertaken by both the computer vision community and the geospatial one is given in Table 1, while a complete review of publicly available benchmark datasets for deep learning evaluation is discussed by Griffiths and Boehm (2019a).

1.2 The 3DOMcity benchmark

Within this framework, this paper presents a novel multi-purpose benchmark for assessing the performance of the entire image-based pipeline for 3D urban reconstruction and 3D data classification. The proposed benchmark is called *3DOMcity* photogrammetric contest and is publicly available at <https://3dom.fbk.eu/3domcity-benchmark>. Compared to previous benchmark initiatives, the innovative aspects introduced by *3DOMcity* are threefold:

- it allows and encourages developers and users in the 3D imaging field to evaluate their own processing tools throughout the entire 3D reconstruction pipeline, covering image orientation, dense image matching, point cloud classification and 3D building reconstruction;
- it permits algorithms evaluation within a metrological context, being the 3D scene captured in a controlled laboratory environment. This controlled setup offers a privileged context

for traceable measurements where the accuracy provided by a particular technique can be properly evaluated against several reference datasets and without worrying about the effect of the environment. In this way, such an environment provides the perfect setting to determine the best results achievable with a particular 3D reconstruction and classification method;

- in contrast to previous benchmarks in the geospatial sector, the provided datasets offer 2D and 3D data collected at a very high spatial resolution, defined by the GSD (ground sampling distance) of the imagery and the density of the 3D point clouds. This enables the assessment of the algorithms' performance to provide for a detailed 3D surface reconstruction. At this point, computational efficiency of the algorithms (i.e. memory and time consumptions), would be also a critical challenge.

In the following, we present (i) the novel benchmark dataset characteristics and corresponding reference data (Section 2), (ii) the description of the multi-purpose benchmark tasks and evaluation procedures (Section 3), and (iii) the discussion on first results of image orientation, dense image matching and classification (Section 4). Conclusive remarks on future benchmark improvements will conclude the article (Section 5).

	Benchmark	Reference	Scenario	Assessment
Small / medium scale	Middlebury, stereo-view (http://www.middlebury.edu/stereo/)	Scharstein and Szeliski, 2003; Scharstein et al., 2014	Laboratory	Stereo correspondence algorithms
	Middlebury, multi-view (http://vision.middlebury.edu/mview/)	Seitz et al., 2006	Laboratory	Multi-view stereo algorithms
	Multi-view evaluation (http://cvlab.epfl.ch/data)	Strecha et al., 2008	Outdoor (building)	Camera calibration and multi-view stereo algorithms
	KITTI Vision Benchmark Suite (www.cvlibs.net/datasets/kitti)	Geiger et al., 2012	Outdoor (street)	Stereo, optical flow, visual odometry / SLAM and 3D object detection algorithms
	Roboimagedata (http://roboimagedata.imm.dtu.dk/)	Jensen et al., 2014	Laboratory	Multi-view stereo algorithms
	Tanks and Temple (https://www.tanksandtemples.org/)	Knapitsch et al., 2017	Outdoor	3D object reconstruction
	ETH3D (http://www.eth3d.net)	Schöps et al., 2017	Laboratory and outdoor	Multi-view and two-view stereo algorithms
Large scale	Benchmark on Building Extraction	Kaartinen et al., 2005	Outdoor (city)	3D building modelling algorithms
	Benchmark on DEM generation (http://www2.isprs.org/commissions/comm1.html)	Reinartz et al., 2010	Outdoor (urban, rural and forested areas)	DEM generation algorithms from stereo satellite data
	Benchmark on urban object classification and 3D building reconstruction (http://www2.isprs.org/commissions/comm3/wg4/tests.html)	Rottensteiner et al., 2012	Outdoor (city)	Urban object extraction from airborne sensors
	Benchmark on High Density Image Matching for DSM Computation – nadir images (https://ifpwww.ifp.uni-stuttgart.de/ISPRS-EuroSDR/ImageMatching/index.en.html)	Haala, 2013	Outdoor (city)	DSM generation from nadir airborne images
	Benchmark on High Density Image Matching for DSM Computation – oblique images (https://ifpwww.ifp.uni-stuttgart.de/ISPRS-EuroSDR/ImageMatching/index.en.html)	Cavegn et al., 2014	Outdoor (city)	DIM from oblique images
	Benchmark for multi-platform photogrammetry (http://www2.isprs.org/commissions/comm1/icwg15b/benchmark_main.html)	Nex et al., 2015	Outdoor (city)	AT and DIM from airborne oblique, UAV-based and terrestrial images
	SynthCity (http://www.synthcity.xyz)	Griffiths and Boehm, 2019b	Outdoor (city)	Urban classification
	DublinCity (https://v-sense.scss.tcd.ie/dublincity/)	Zolanvari et al., 2019	Outdoor (city)	Urban classification

Table 1. Overview of state-of-the-art benchmark initiatives undertaken by the computer vision and geospatial communities in the fields of image-based 3D geometry reconstruction and scene analysis.

2. BENCHMARK DATASET

The benchmark is realized within the FBK-3DOM metrological laboratory, built specifically for research and development activities in the field of traceable measurements using 3D imaging technologies. As test-object, an ad-hoc 3D artefact is employed (Figure 1) that simulates a typical urban scenario by means of a mock 3D city model (approximately 0.8 x 0.8 m). It is characterized by different textures and materials, including opaque vs transparent plastic surfaces on buildings, a spongy rubber layer on roads, and smooth plastic granulates on grass-, bush-, and tree-like objects. Furthermore, the presence of detailed elements on the surfaces (e.g. small openings in the building facades, roof tiles, tiny city billboards, etc.) and significant depth variations at the adopted imaging scale, make the test field a realistic (and challenging) representation of an urban landscape. The 3D scene contains:

- the 3D object, firmly anchored on the ground and featuring the most typical elements of an urban scenario, i.e. buildings of varying shapes and heights, streets, gardens and trees;
- different kinds of references, i.e. coded/circular targets, resolution charts and scale bars laser printed on a semi-rigid and multi-layer surface, two calibrated invar rods, a calliper and a rigid measuring tape.

A controlled setup is realized for image acquisition (Figure 1, left), where a diffused and controlled ambient light (fluorescent) illuminates the scene without cast shadows. A set of 420 images are acquired with a Nikon D750 digital camera (6016 x 4016 pixels, pixel size $\sim 5.9 \mu\text{m}$) equipped with a fixed 50 mm focal length lens. A classical airborne flight is simulated by advancing the camera along a rigid bar that is sequentially moved over the scene as to define parallel flight lines. The resulting photogrammetric network is displayed in Figure 3, left. Each camera station consists of six images, two nadir looking views (one in portrait mode and one in landscape mode w.r.t. the along track direction), and four oblique looking views (forward, backward, left and right w.r.t. the along track direction), acquired under a 45° tilting of the camera principle axis. Examples of nadir and oblique images are shown in Figure 1, centre and right respectively. The overlap is 80/65% (along/across-track directions) in the landscape nadir images, while it is 85/70% for the oblique images. At a mean acquisition height of 1.03 m, the GSD is 0.124 mm in the nadir images and varies from 0.128 mm to 0.273 mm in the oblique views. By upscaling this flight plan to real-world dimensions, it would result in an approximate 1 cm GSD in the nadir images, thus allowing to fully exploit the high resolution and fine surface detail provided by the image-based 3D reconstruction.

This image set represents the rawest level of data, publicly available within the *3DOMcity* photogrammetric contest. Starting from it, further processed datasets are generated and

provided to the users as part of the benchmark: they will be described in Section 3, along with the contest's tasks and evaluation procedures.

2.1 Reference dataset

The evaluation tests are performed using the following reference data: point clouds acquired with a triangulation-based laser scanner, a set of targets, two calibrated scale bars, two printed scale bars, one calliper and one rigid measuring tape. While the assessment procedures will be described in Section 3, an error budget estimated here, that accounts for the major components of measurement uncertainty. Indeed, this represents a crucial step, if one prefers to perform an inter-comparison or an evaluation of a particular 3D imaging technique in a metrological context.

The 3D scene includes a set of circular and coded targets (115 in total), of varying dimensions, and well distributed all around the test object (Figure 1, left). In order to avoid any additional source of uncertainty that may introduce deformation effects within the image block, these targets are only used to evaluate the precision of the adjustment in image space. This is done by computing the residuals of their image coordinates and corresponding statistics. Thus, the localization errors can be expected to originate only from the adjustment process and the algorithms used to automatically locate their centroid in the image, and not from any external technique used to measure the 3D coordinates of their centre. Since the length of the two calibrated scale bars (one of which is visible in Figure 1, centre) and the distance between the circular targets (L1-R1) of the rigid measuring tape are the only reference data where both uncertainty and traceability can be attached to, they are used to solve the scale ambiguity of the photogrammetric network. The expected uncertainty of length measurements for the rods is 0.003 mm, with a retro target 95% confidence interval of 0.0042 mm, whereas the *a-priori* accuracy of the L1-R1 distance is estimated at 0.012 mm. As check information, the distances between the circular targets of the two printed scale bars, and the ones printed on the calliper are adopted. Unfortunately, no calibration is provided for them.

A multi-stripe triangulation-based laser scanner, NextEngine 3D Scanner, is used to acquire a 3D model of two buildings of interest (highlighted in red in Figure 1, right). No attempt is made to evaluate the image-based reconstruction procedures over the whole test area, and only two small volumes are measured in order to estimate the accuracy over a single area using a limited number of 3D point clouds. The two buildings, hereinafter referred to as "border building" and "centre building", are selected among the others based on their position in the centre and at the border of the scene. This allows the assessment of DIM performance under different scenarios w.r.t. (i) redundancy and geometric configuration of image rays (centre vs border of the image block), and (ii) occlusion levels.



Figure 1. The test-object and acquisition scenario adopted for the *3DOMcity* photogrammetric contest: setup for data collection (left) and examples of nadir (centre) and oblique (right) images. Red ellipses highlight the buildings selected for DIM evaluation.

A lateral resolution (single scan value) of 0.120 mm is achieved, and the instrument datasheet reports a dimensional accuracy of 0.100 mm when the macro lens is used. The alignment of two 3D point clouds acquired from two distinct scanner-object positions resulted in an error (1 σ value) of about 0.124 mm: this value is a realistic representation of the noise level present in the 3D point clouds and confirms the datasheet expectations.

In order to refine the registration among the point clouds and get a complete 3D model of the object, the incremental ICP-based registration algorithm available within the open source Point Cloud Library is here adopted (PCL, 2019). It finds the best transformation between each consecutive point cloud and accumulates these transformations over the whole set of clouds. The final refinement error (mean value) resulted in 0.218 mm for the border building and 0.263 mm for the centre building.

3. BENCHMARK TASKS AND EVALUATION

The *3DOMcity* photogrammetric contest is based on a modular concept that allows the user to evaluate the performance of any algorithm of interest when applied to a specific step of the photogrammetric pipeline and starting from raw or processed data. The benchmark framework, as illustrated in Figure 2, is thus multi-purpose since it involves a multiplicity of tasks that can be performed independently from each other's. Tasks can also be possibly grouped together, e.g. users may want to first assess the accuracy of their image orientations, and then start from these to compute and evaluate the subsequent DIM result. The benchmark tasks and corresponding evaluation procedures are described in the followings.

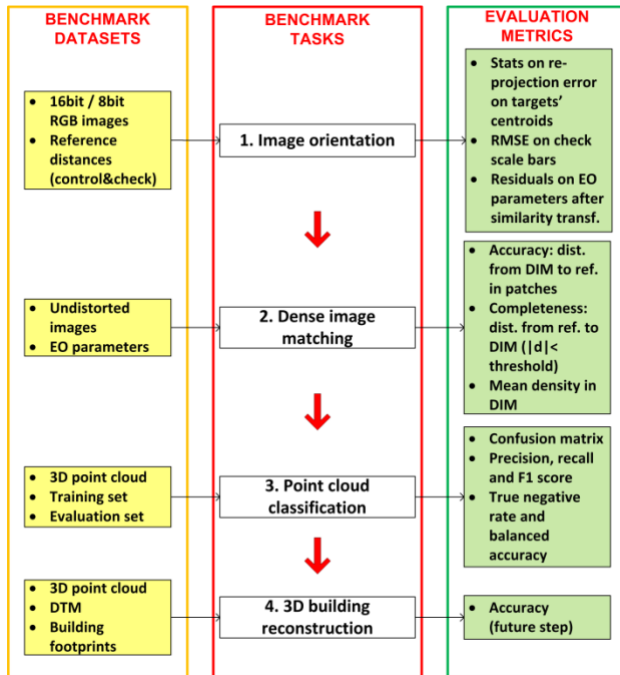


Figure 2. The *3DOMcity* photogrammetric contest: datasets, tasks and corresponding evaluation metrics.

3.1 Image orientation

The photogrammetric network adjustment represents the first task of the contest. As input data, the 420 RGB images, in 16bit tiff and 8bit jpg file format, are provided. As control data to solve the scale ambiguity, three reference lengths are available, i.e.: the distance between the retro-reflective targets of the two calibrated

scale bars (yellow invar rods), and the distance between the circular targets R1 and L1 glued on the rigid measuring tape.

The evaluation procedure consists of three parts:

- *Precision assessment.* Precision defines the statistical noise of an adjustment, i.e. it models the inner accuracy of a given system (Luhmann, 2011). To evaluate precision in image space, the residuals of the image coordinates of 115 targets' centroids are used, together with corresponding statistics (RMS and max value).
- *Accuracy assessment.* Accuracy models the deviation of a measured value from an independent, reference measurement. This nominal measurement is defined by a measurement system of higher accuracy, a calibrated reference body or a true value (Luhmann, 2011). To evaluate the accuracy of the adjusted network, three reference lengths are here provided, i.e.: the distance between the circular targets of the two laser printed scale bars, and the distance between the circular targets glued on the caliper. Starting from the orientation results, the distances between the forward-intersected targets' centroids are compared against the nominal ones, thus deriving their residuals and corresponding statistics.
- *Relative accuracy assessment.* No independent nominal values are available for image external orientation parameters (EO - poses and attitudes). Experimental values are therefore estimated by performing a self-calibrating bundle block adjustment (BBA) of the image network. A SIFT-like operator is adopted to automatically extract a large number of tie points (more than 5 million points) that are then exported in the form of both image observations and corresponding 3D coordinates. These data are later filtered and regularized by applying an in-house developed tool, based on a weighted combination of three criteria: re-projection error, intersection angle of image rays and their multiplicity (Farella et al., 2019). The filtered and more reliable correspondences (almost 2 million points) are subsequently imported as image observations into the bundle adjustment, where a free-network self-calibrating adjustment is again performed. The reference control distances are finally used to rigidly scale the image network. The resulting adjusted image network, hereinafter referred to as "reference image network", is shown in Figure 3 left, whereas the corresponding BBA statistics are listed in Table 2.

Mean image multiplicity per each tie point	3.284
Mean intersection angle per each tie point – [°]	17.618
RMS re-projection Error for automatically extracted tie points – [px]	0.309
Max. re-projection error for automatically extracted tie points – [px]	1.925
RMS re-projection Error for automatically extracted target centroids – [px]	0.204
Max. re-projection error for automatically extracted target centroids – [px]	0.905
RMSE on control scale bars – [mm]	0.005
RMSE on check scale bars – [mm]	0.066

Table 2. Statistical results of the assessment performed on the reference image network.

Starting from the EO results adjusted by the contest's participants, the camera EO parameters of the reference image network are thus exploited to compute a rigid similarity transformation (6 DOF), that maps the submitted network to the reference one. Lastly, errors of this transformation and

corresponding residuals on EO parameters are used to estimate the relative accuracy of the submitted adjustment results.

3.2 Dense image matching

The second task of the *3DOMcity* photogrammetric contest is to generate a dense 3D reconstruction of the scene, by applying high density image matching algorithms to the set of oriented images. The participants may either start from the reference image network dataset (provided in the form of undistorted images and corresponding adjusted EO parameters), or apply their SfM-DIM pipeline to the raw image set. In the second case, participants are asked to submit also their adjusted EO parameters, so that the similarity transformation described in Section 3.1 can be used to align the submitted DIM result to the reference scanner-based point cloud (see Section 3.3 for the registration of the latter within the coordinate system of the reference image network). Up to now, the assessment is limited to two buildings of interest, while further reference datasets will be acquired in the next future. Following the work of Seitz et al. (2006) and Jensen et al. (2014), two evaluation criteria are here checked, namely accuracy and completeness. The former estimates how close the submitted result is to the reference one, the latter quantifies how much of the reference result is reconstructed in the submitted one. Both metrics can be affected by the following issues, that may hamper a rigorous assessment:

- varying textural and geometric complexity and challenges in the 3D scene, that can influence the quality of both the DIM points and the ones collected by the triangulation-based scanner;
- varying point density between submitted and reference clouds;
- missing data and occlusions in both the submitted and the reference clouds;
- presence of outliers in the submitted DIM result.

The following strategies are adopted in order to tackle these problems and lead to possibly unbiased evaluation results:

- *Accuracy assessment.* Accuracy is estimated by measuring the distance from the submitted DIM result to the reference one. First, the mean spatial resolution of the submitted DIM cloud is estimated for each of the two building and adopted as subsampling threshold for the reference cloud. Incidentally, if the DIM cloud is denser than the reference one, the former is decimated according to the latter. This decimation process ensures a consistent comparison between the two point sets, reducing the impact of their varying point density. Second, the reference cloud is segmented and a number of well distributed patches are extracted, such as to (i) exclude areas of missing data, boundary areas and possibly noisy areas, where challenging surface materials may have affected the performance of the active-based 3D imaging technique, and (ii) include areas that differ in texture and geometric complexity. Third, the accuracy assessment is performed on the selected patches, i.e. only DIM points that map on those regions are

considered within the evaluation procedure. For it, the distances between the points in the DIM result and their closest points in the extracted reference patches are measured. In order to assess whether the submitted reconstruction tends to under- or over-estimate the nominal shape of the object, the signed distances are computed, based on the normal vectors at the nearest points in the reference. Finally, given the sample of the signed distances remaining after outliers removal ($3 \cdot \sigma_{MAD}$ rule), their distribution is visualized and the following statistics are computed: Mean, σ , Median and $\sigma_{MAD} = 1.4826 \cdot MAD$. An example of such evaluation result is given in Figure 4 and will be commented in Section 3.3.

- *Completeness assessment.* After applying the same subsampling approach adopted for the accuracy assessment, completeness is quantified by measuring the distance from the reference point cloud to the submitted DIM one. Indeed, it is based on the intuitive concept that, if a reference point maps “closely” to a DIM point. This implies that it is also reconstructed by the multi-view matching algorithm. Contrary to the patch-based approach previously used, the whole population of reference points is included in the evaluation, in order to give a more general idea of the DIM performance to achieve a complete 3D reconstruction over the entire volume of interest. Incidentally, to ensure a fair comparison, we deleted the reference points falling in areas where the intersection of image rays is denied due to occlusion issues (e.g. the portions beneath balconies and awnings). Furthermore, while computing the distance between each reference point and its closest neighbor in the DIM result, its absolute value is here considered, to evaluate if it is “small” enough to mark that point as present also in the submitted result. The challenge here is to distinguish the “true” missing parts in the DIM point cloud from its noisy reconstructed areas, i.e. those DIM points that have been matched far away from their actual position. To tackle this issue, inspired by the work of Seitz et al. (2006), completeness is here defined as the fraction of reference points that are within an allowable distance d from their neighbors in the DIM point cloud. The maximum acceptable distance is set to three times the mean spatial resolution adopted during the initial decimation process. This comparison-specific threshold is chosen as a good tradeoff to avoid penalizing sparser (although complete) DIM point clouds and their “reasonable” errors, i.e. erroneously matched points in the submitted DIM result. For the sake of transparency, the estimated mean spatial resolution of the DIM cloud is listed as well among the evaluation metrics for each of the two buildings.

The above strategies for the evaluation of DIM results were applied to our reference point cloud, that is provided as input data for the point cloud classification task (see Section 3.3). It was generated by applying a semi-global DIM algorithm at the second-level image pyramid, starting from the reference image network.



Figure 3. Examples of the datasets available within the “*3DOMcity*” photogrammetric contest: reference image network (left), reference DIM point cloud (overview in center and close-up view on the right).

For this reason, it is called “reference DIM point cloud” (Figure 3, center and right). It consists of approximately 30 million points and features an overall mean spatial resolution of 0.158 mm, that corresponds to a point density of around 1700 points/m² in a real-world scenario.

Its quality in terms of accuracy is graphically visualized by the histogram of the signed distances to the reference scanner-based result (Figure 4). Statistics report a median of 0.119 mm and σ_{MAD} of 0.056 mm for the border building, whereas a median of 0.133 mm and σ_{MAD} of 0.082 mm are estimated for the centre building. Finally, their completeness levels reach 88.3 % and 80.9% (border and centre, respectively), that account for the fraction of points in the scanner-based reference data that are within the allowable distance from the DIM point cloud. Overall, the accuracy levels achieved in both cases are consistent between each other’s, and in the range of the mean GSD of the original imagery. The difference in completeness is due to the presence of trees that are partially occluding the facades of the centre building.

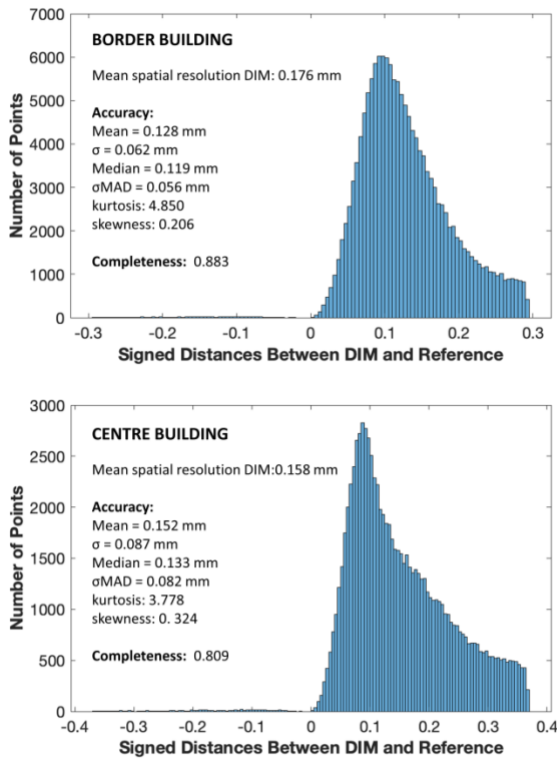


Figure 4. Result of the accuracy and completeness assessments performed on the reference DIM point cloud.

3.3 Point cloud classification

The classification of the 3D point cloud is the third task of the 3DOMcity photogrammetric contest. As input, the reference DIM point cloud (including X, Y, Z, R, G, B and normal vector per each point) is provided. It serves two main purposes within the present benchmark:

- it is used as reference data for the registration of the NextEngine point clouds to the coordinate system of the reference image network. This is accomplished by applying an ICP-based registration refinement, resulting in a mean transformation error of 0.363 mm.
- it is publicly provided as a part of the input datasets for the classification task. The benchmark dataset includes a training (~10% of all points) and an evaluation set (~10% of all points) of manually labelled points (Figure 5). Participants are asked to classify the point cloud into the following classes of

interests: ground, low vegetation, shrub, high vegetation, façade and roof.

The classification results are then analysed for accuracy assessment by computing: confusion matrix, precision, recall, F1 score, true negative rate (TNR) and balanced accuracy (Equations 1-5). A classification example is discussed in Section 4, together with its corresponding evaluation results.

$$Precision = \frac{tp}{tp + fp} \quad (1)$$

$$Recall = \frac{tp}{tp + fn} \quad (2)$$

$$F1 = 2 * Precision * \frac{Recall}{Precision + Recall} \quad (3)$$

$$TNR = \frac{tn}{tn + fp} \quad (4)$$

$$Balanced Accuracy = \frac{Recall + TNR}{2} \quad (5)$$

where, tp: true positive, tn: true negative, fp: false positive, fn: false negative.

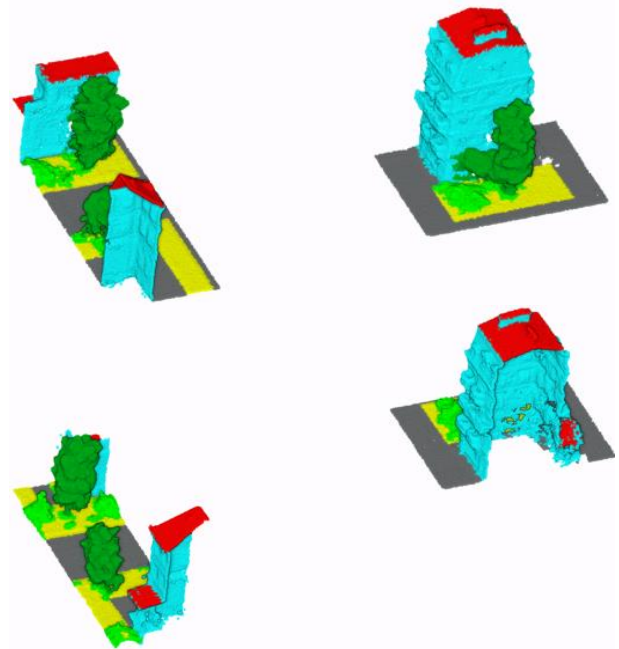


Figure 5. Training set (top) and evaluation set (bottom) for 6 classes: ground (grey), grass (yellow), shrub (light green), tree (dark green), facades (cyan), roof (red).

3.4 Building reconstruction

The last task of the 3DOMcity photogrammetric contest is to generate both LOD1- and LOD2-compliant building models, by processing the reference DIM point cloud, together with the DTM of the scene and the footprints of the buildings, as auxiliary datasets. The ground truth data and evaluation procedure will be provided in a future step of this benchmark initiative.

4. FIRST CLASSIFICATION TEST

4.1 Classification procedure

Our classification procedure (Figure 6) is based on a deep learning (DL) method. However, unlike common way of using DL, we extract features for the points in the point cloud, and also

use these features as an input. Considering the computational power necessities for commonly used DL approaches, our approach can be run on a mid-level laptop, due to its light design, which is possible with including features in the input.

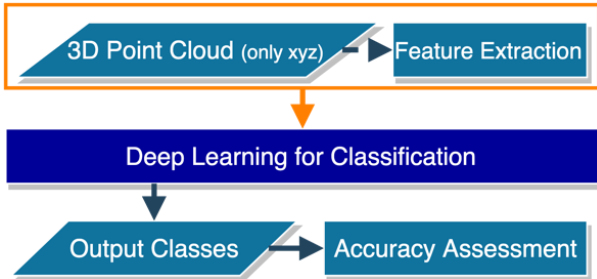


Figure 6. The used deep learning classification approach.

The extracted features include linearity, planarity, surface variation and sphericity from eigen features. In addition to these roughness, local planarity, elevation change, height above ground and z- component of the normal vector are used. In their work, Hackel et al. (2016) details the used eigen-features. The features of local planarity and elevation change are given in Özdemir and Remondino (2019), and height above ground is explained in Özdemir et al. (2019).

Our deep neural network (DNN) is a 2D convolutional neural network (CNN) and it is shown in Figure 7.

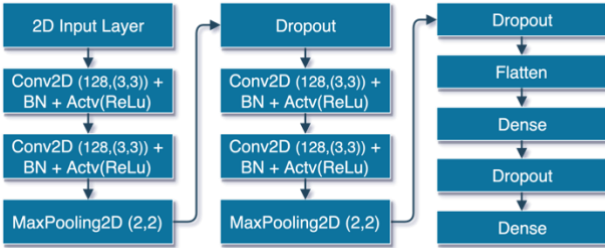


Figure 7. The employed 2D-CNN architecture for the point cloud classification task.

4.2 Classification result and evaluation

Classification results for the evaluation set is shown in Table 3 with confusion matrix and Table 4 with accuracy metrics. The corresponding classified point cloud is shown in Figure 7 (left: evaluation set; right: whole point cloud).

Our classification workflow takes ~40 minutes for feature extraction (total time for training and evaluation sets, on laptop CPU) and ~4 hours for training (on a mid-level laptop GPU).

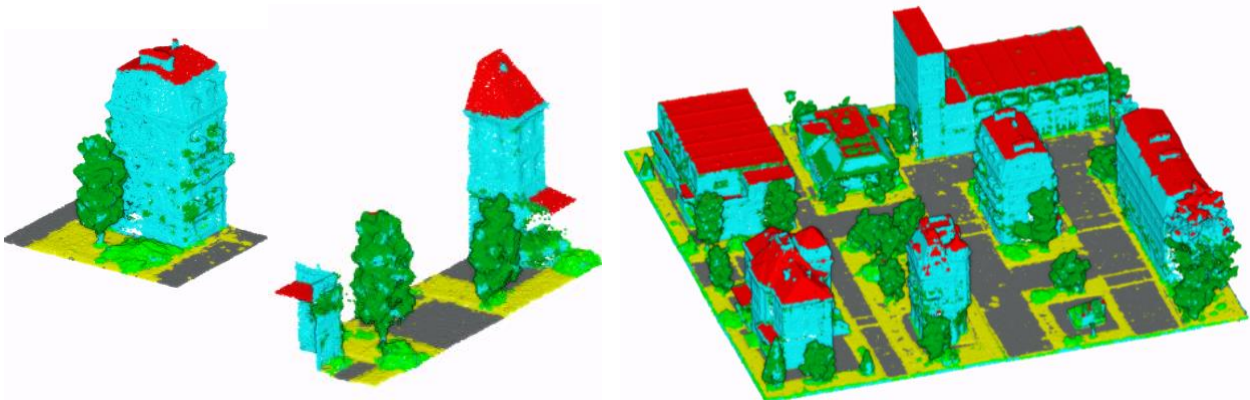


Figure 8. Classified evaluation data (left) and classified point cloud (right) for 6 classes: ground (grey), grass (yellow), shrub (light green), tree (dark green), facades (cyan), roof (red).

Considering the point cloud's density, we initially down sampled the original DIM cloud to ~7% of it (<2million points after down sampling), as the original DIM cloud is denser than our needs. Our classification results (Tables 3 and 4, Figure 8) show that our approach is able to provide useful results for this 6-class classification challenge. Interpreting the confusion matrix (Table 3), our future work will be focusing on designing better features, feature selection, and a more enhanced DL, so as to define each point within the global space in a better way, hence resulting with less confusions.

Classes	Ground	Grass	Shrub	Tree	Facade	Roof
Ground	15823	3609	608	9	223	0
Grass	194	11211	1356	0	3	0
Shrub	19	770	12646	379	613	0
Tree	0	12	191	43671	3530	32
Facade	79	65	1911	13477	54475	282
Roof	0	0	0	841	913	8758

Table 3. Confusion matrix for the classification result on evaluation set (overall accuracy: 83.4%).

Classes	Precision	Recall	F1_Score	TNR	Blncd. Acc.
Ground	78.1%	98.2%	87.0%	97.2%	97.7%
Grass	87.8%	71.6%	78.9%	99.0%	85.3%
Shrub	87.7%	75.7%	81.2%	98.9%	87.3%
Tree	92.1%	74.8%	82.5%	96.8%	85.8%
Facade	77.5%	91.2%	83.8%	86.4%	88.8%
Roof	83.3%	96.5%	89.4%	98.9%	97.7%
Avg.	84.4%	84.7%	83.8%	96.2%	90.4%

Table 4. Per class accuracy metrics, for the classification result on evaluation set.

5. CONCLUSIONS

This paper introduced the *3DOMcity* photogrammetric contest, a multi-purpose benchmark for assessing the performance of the entire image-based pipeline for 3D urban reconstruction and 3D data classification. Within a broader context of international benchmark actions, this initiative strives to provide:

- a modular framework, involving a multiplicity of tasks (from image orientation to 3D building reconstruction) that can be either performed independently from each other's, or grouped together;

- an experimental scenario specifically designed to assess the algorithms' performance throughout the entire 3D reconstruction pipeline, and that within a metrological context. The metrological approach adopted here is based on reference instruments and materials, whose uncertainty is estimated within an a-priori error budget;
- 2D and 3D publicly available data collected at a very high spatial resolution, which would also be a challenge for the computational efficiency of the algorithms.

As future improvements, further reference data will be collected for the assessment of DIM accuracy and completeness, thus including further buildings in the evaluation procedure. This will generalize the analysis to account for a larger variety of building shapes, surface details, texture and material complexity, and occlusions levels. Finally, the ground truth data and evaluation procedure for the 3D building reconstruction task will be provided soon in order to complete the whole *3DOMcity* benchmark framework and make it fully operational.

REFERENCES

- Cavegn, S., Haala, N., Nebiker, S., Rothermel, M. and Tutzauer, P., 2014. Benchmarking high density image matching for oblique airborne imagery. In: *ISPRS Int. Archives of Photogram., Remote Sens. & Spatial Inf. Sci.*, Vol. 40(3), pp. 45-52.
- Farella E. M., Torresani A. and Remondino F., 2019. Quality features for the integration of terrestrial and UAV images. In: *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42(2/W9), pp. 339-346.
- Geiger, A., Lenz, P. and Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In: *Proc. IEEE CVPR*, pp. 3354-3361.
- Griffiths, D. and Boehm, J., 2019a. A review on deep learning techniques for 3D sensed data classification. In: *Remote Sensing*, Vol. 11(12), p.1499.
- Griffiths, D. and Boehm, J., 2019b. SynthCity: A large scale synthetic point cloud. *arXiv preprint arXiv:1907.04758*.
- Haala, N. and Kada, M., 2010. An update on automatic 3D building reconstruction. In: *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 65, pp. 570-580.
- Haala, N. and Rothermel, M., 2015. Image-based 3D data capture in urban scenarios. In: *Proc. Photogrammetric Week 2015*, D. Fritsch (Ed.), pp. 119-130.
- Haala, N., 2013. The landscape of dense image matching algorithms. In: *Photogrammetric Week '13*, pp. 271-284.
- Haala, N., Rothermel, M. and Cavegn, S., 2015. Extracting 3D urban models from oblique aerial images. In: *Proc. of the 2015 Joint Urban Remote Sensing Event (JURSE)*, pp. 1-4.
- Hackel, T., Wegner, J.D. and Schindler, K., 2016. Fast semantic segmentation of 3D point clouds with strongly varying density. In: *ISPRS Annals of Photogram., Remote Sens. & Spatial Inf. Sci.*, Vol. III-3, pp. 177-184.
- Jensen, R., Dahl, A., Vogiatzis, G., Tola, E. and Aanæs, H., 2014. Large scale multi-view stereopsis evaluation. In: *Proc. IEEE CVPR*, pp. 406-413.
- Kaartinen, H., Hyypä, J., Guelch, E., et al., 2005. Accuracy of 3D city models: EuroSDR comparison. In: *ISPRS Int. Archives of Photogram., Remote Sens. & Spatial Inf. Sci.*, Vol. 36(3/W19), pp. 227-232.
- Knapitsch, A., Park, J., Zhou, Q.Y. and Koltun, V., 2017. Tanks and Temples: benchmarking large-scale scene reconstruction. In: *ACM Transactions on Graphics*, Vol. 36(4).
- Luhmann, T., 2011. 3D imaging: how to achieve highest accuracy. In: *Videometrics, Range Imaging, and Applications XI (Proc. SPIE)*, Vol. 8085.
- Nex, F., Gerke, M., Remondino, F., Przybilla, H.-J., Bäumker, M. and Zurhorst, A., 2015. ISPRS benchmark for multi-platform photogrammetry. In: *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. II-3/W4, pp. 135-142.
- Özdemir, E. and Remondino, F., 2019. Classification of aerial point clouds with deep learning. In: *ISPRS Int. Archives of Photogram., Remote Sens. & Spatial Inf. Sci.*, Vol. 42(2/W13), pp. 103-110.
- Özdemir, E., Remondino, F. and Golkar, A., 2019. Aerial point cloud classification with deep learning and machine learning Algorithms. In: *ISPRS Int. Archives of Photogram., Remote Sens. & Spatial Inf. Sci.* (in press).
- PCL - Point Cloud Library, 2019. Incremental ICP - http://pointclouds.org/documentation/tutorials/pairwise_incremental_registration.php (Last access: August 2019)
- Reinartz, P., d'Angelo, P., Krauß, T., Poli, D., Jacobsen, K. and Buyuksalih, G., 2010. Benchmarking and quality analysis of DEM generated from high and very high resolution optical stereo satellite data. In: *ISPRS Int. Archives of Photogram., Remote Sens. & Spatial Inf. Sci.*, Vol. 38(1).
- Remondino, F. and Gerke, M., 2015. Oblique Aerial Imagery – A Review. In: *Proc. Photogrammetric Week 2015*, D. Fritsch (Ed.), pp. 75-83.
- Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S. and Breitkopf, U., 2012. The ISPRS benchmark on urban object classification and 3D building reconstruction. In: *ISPRS Ann. Photogram., Remote Sens. & Spatial Inf. Sci.*, Vol. I(3), pp.293-298.
- Scharstein, D. and Szeliski, R., 2003. High-accuracy stereo depth maps using structured light. In: *Proc. IEEE CVPR*, Vol. 1, pp. 1195-202.
- Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X. and Westling, P., 2014. High-resolution stereo datasets with subpixel-accurate ground truth. In: *German conference on pattern recognition*, pp. 31-42.
- Schöps, T., Schonberger, J.L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M. and Geiger, A., 2017. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In: *Proc. IEEE CVPR*, pp. 3260-3269.
- Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *Proc. IEEE CVPR*, Vol. 1, pp. 519-528.
- Strecha, C., Von Hansen, W., Van Gool, L., Fua, P. and Thoennessen, U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: *Proc. IEEE CVPR*, pp. 1-8.
- Toschi, I. et al., 2017. Oblique photogrammetry supporting 3D urban reconstruction of complex scenarios. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. 42(1/W1), pp. 519-526.
- Toschi, I., Remondino, F., Rothe, R. and Klimek, K., 2018. Combining airborne oblique and LiDAR sensors: investigation and new perspectives. In: *ISPRS Int. Archives of Photogram., Remote Sens. & Spatial Inf. Sci.*, Vol. 42(1), pp. 437-444.
- Zolanvari, S.M.I., Ruano, S., Rana, A., Cummins, A, da Silva, R.E., Rahbar, M., Smolic, A., 2019. DublinCity: Annotated LiDAR Point Cloud and its Applications. In: *Proc. 30th BMVC*.