

SUPERVISED CLASSIFICATION AND ITS REPEATABILITY FOR POINT CLOUDS FROM DENSE VHR TRI-STEREO SATELLITE IMAGE MATCHING USING MACHINE LEARNING

A-M. Loghin^{1,*}, N. Pfeifer¹, J. Otepka-Schremmer¹

¹Department of Geodesy and Geoinformation, Technische Universität Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria – (ana-maria.loghin, Norbert.Pfeifer, Johannes.Otepka)@geo.tuwien.ac.at

Commission II, WG II/6

KEY WORDS: Machine Learning, Supervised Classification, Dense Image Matching, Very High Resolution (VHR) Satellite Imagery, Tri-stereo

ABSTRACT:

Image matching of aerial or satellite images and Airborne Laser Scanning (ALS) are the two main techniques for the acquisition of geospatial information (3D point clouds), used for mapping and 3D modelling of large surface areas. While ALS point cloud classification is a widely investigated topic, there are fewer studies related to the image-derived point clouds, even less for point clouds derived from stereo satellite imagery. Therefore, the main focus of this contribution is a comparative analysis and evaluation of a supervised machine learning classification method that exploits the full 3D content of point clouds generated by dense image matching of tri-stereo Very High Resolution (VHR) satellite imagery. The images were collected with two different sensors (Pléiades and WorldView-3) at different timestamps for a study area covering a surface of 24 km², located in Waldviertel, Lower Austria. In particular, we evaluate the performance and precision of the classifier by analysing the variation of the results obtained after multiple scenarios using different training and test data sets. The temporal difference of the two Pléiades acquisitions (7 days) allowed us to calculate the repeatability of the adopted machine learning algorithm for the classification. Additionally, we investigate how the different acquisition geometries (ground sample distance, viewing and convergence angles) influence the performance of classifying the satellite image-derived point clouds into five object classes: *ground, trees, roads, buildings, and vehicles*. Our experimental results indicate that, in overall the classifier performs very similar in all situations, with values for the F1-score between 0.63 and 0.65 and overall accuracies beyond 93%. As a measure of repeatability, stable classes such as buildings and roads show a variation below 3% for the F1-score between the two Pléiades acquisitions, proving the stability of the model.

1. INTRODUCTION

Point cloud classification, i.e. semantic segmentation of point clouds, has always been an essential and challenging task with applications in 3D city modelling, urban planning, monitoring, autonomous-driving, virtual/augmented reality, and robotics. The two important sources for 3D point cloud acquisition over large areas are image matching of aerial or satellite images and Airborne Laser Scanning (Wehr et al., 1999; Remondino et al., 2014). For more than thirty years, the interest of the scientific community has turned to the great potential of stereo satellite imagery for providing 3D geospatial information over large areas in a timely and cost-effective manner (D'Angelo et al., 2008). Therefore, due to the new dense matching techniques and their improved spatial resolution, satellite imagery is becoming an important data source for image-derived 3D point clouds (Xie et al., 2019). Moreover, VHR optical satellite sensors are able to collect not only stereo but tri-stereo images of the same area during a single flight path, from different viewing angles (along-track): forward (F), close to nadir (N) and backward (B). This ability has many advantages with regard to point cloud completeness and Digital Surface Model (DSM) derivation (Panagiotakis et al., 2018, Piermattei et al., 2018). For understanding the 3D scenes generated by image matching and to make use of them, classification plays a key role. In our case, by classification we refer to the assignment of semantic

labels to points, on a per-point basis (Otepka et al., 2013). For this purpose, many different classification algorithms are available and there has been much progress in machine learning in the recent years (Grilli et al., 2017). Due to the existence of noise, occlusions and different objects types with various sizes and shapes, the classification of 3D point clouds is a challenging task.

Most of the previous studies on classification of 3D point clouds generated by image matching use rasterization or voxelisation, which reduces their full 3D content. For example, Gerke et al. (2013) converted the point clouds into a voxel representation and segmented them by adopting a 'Random trees' machine learning technique and a supervised method (Markov-Random-Field). Finally, point clouds derived from image matching of airborne oblique images over two urban areas were classified into the following classes: façade, roof, rubble, sealed ground, and trees. Modiri et al. (2015) propose a region-growing technique to classify buildings and vegetation from stereo UltraCam-X matched point clouds, by using colour information and vegetation index. In contrast, in their approach Tran et al. (2018) analyse the results of two supervised classification algorithms on original 3D point clouds derived from high-resolution aerial images over an urban area with Ground Sample Distance (GSD) of 6 cm.

Benchmarking datasets containing reference data are of great importance in evaluation tasks, being also used as training data

* Corresponding author

in classification problems. Well known available benchmarks for 3D point cloud semantic segmentation are the following: the Oakland 3D point cloud dataset, which contains laser data collected from a moving platform in an urban environment (Munoz et al., 2009); the Sydney Urban Objects data set (Deuge et al., 2013) and IQmulus & TerraMobilita Contest (Vallet et al., 2015) use mobile laser scanning for the 3D point cloud acquisition in dense urban environments; the ISPRS benchmark on urban object classification and 3D building reconstruction (Rottensteiner et al., 2012), and Semantic3D.Net: a new large-scale point cloud classification benchmark for natural and urban scenes (Hakel et al., 2017).

However, there is a small amount of work exploring the semantic segmentation of satellite image-derived point clouds. In their investigation, Leotta et al. (2019) develop an end-to-end system for segmenting buildings and bridges from terrain, by using point clouds derived from WorldView-3 multi-view satellite imagery. A reason for the reduced research in this direction would be the low 3D quality of the obtained point clouds, caused generally by the smoothing effects of the used dense image matching algorithm, as well as the 3D reconstruction difficulties encountered in the occluded, non-textured areas, water surfaces and repetitive patterns.

Our aim is to understand the quality of satellite image derived point clouds for semantic segmentation. We chose classes of decreasing occurrence probability. Therefore, we apply a supervised machine learning algorithm using decision trees, for classifying tri-stereo satellite image-derived 3D point clouds into ground, trees, roads, buildings, and vehicles. The VHR sensors that we use are Pléiades and WorldView-3 with GSD of 0.70 m and 0.30 m, respectively. Our study is, to the best of our knowledge, the first to assess the performance of a supervised machine learning algorithm for classification of 3D point clouds derived from Pléiades and WorldView-3 tri-stereo VHR satellite imagery. Specifically, our research investigation has as main purpose finding the answers to the following questions:

- (1) How repeatable is the process from acquisition to classification result?
- (2) What is the impact of acquisition geometry and GSD on the classification result?

With three acquisitions over the same area indicative answers will be given.

The rest of this paper is organised as follows: first, we describe the study site and the tri-stereo satellite image datasets (Section 2). Next, we present the photogrammetric workflow for 3D reconstruction from tri-stereo scenes and the machine learning approach for point cloud classification (Section 3). Our results are presented in Section 4 and finally we conclude and summarize this paper in Section 5.

2. STUDY AREA AND IMAGE DATASETS

The study area for our investigations is located in Waldviertel, a hilly region in Lower Austria (48° 30' 30"N; 15° 08' 34"E; WGS84), with elevations ranging from 537 to 846 m above sea level (Figure 1). On 13 June 2017 the territory was captured with the Pléiades-1B sensor, in a tri-stereo mode, covering a surface on the ground of 159 km². A week later, on 20 June 2017 the same platform acquired 383 km², having an overlap of 45 km² with the first dataset. The baseline to height ratios (B/H) are of 0.24 and 0.25 for the two acquisitions, corresponding to Forward-Backward (FB) image combination. For analysing the impact of a different GSD and on the performance of the classifier, we tasked a new tri-stereo WorldView-3 dataset in

the same area. The tri-stereo images were acquired on 8 April 2018.

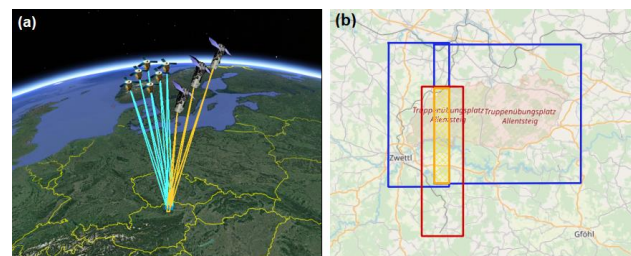


Figure 1. Study area (a) Acquisition geometries of Pléiades (blue) and WorldView-3 sensors (orange) in Google Earth; (b) acquisition footprints with marked common study area (yellow) overlaid on Open Street Map (OSM)

All images were delivered as pan-sharpened with four bands (Red, Green, Blue and Near-infrared). Depending on the different viewing angles, the spatial resolutions are varying between 0.70 - 0.71 m and 0.31 - 0.32 m for Pléiades and WorldView-3 images, respectively. For each image, the Rational Polynomial Coefficients (RPCs), allowing the conversion between image and object space, were provided by the supplier. Detailed information regarding the acquisition properties for all the tri-stereo datasets are shown in Table 1. The acquisition geometries of both Pléiades tri-stereo datasets are very similar w.r.t the values for the B/H ratio and convergence angles on the ground. In contrast, the WorldView-3 acquisition shows values approximately two times larger.

Sensor type & acq. date	View	GSD (m)	Viewing Angles (°)		B / H Ratio	Convergence Angle (°)
			Across	Along		
Pléiades 13-06-2017	F	0.71	3.15	-5.66	0.13	7.5 (FN)
	N	0.70	3.37	0.46	0.11	6.3 (NB)
	B	0.71	3.62	5.19	0.24	13.8 (FB)
Pléiades 20-06-2017	F	0.71	3.14	-5.49	0.13	7.5 (FN)
	N	0.70	1.45	0.06	0.12	6.4 (NB)
	B	0.71	0.07	6.09	0.25	13.9 (FB)
WorldView-3 08-04-2018	F	0.32	7.70	-11.0	0.22	12.7 (FN)
	N	0.31	7.23	-0.62	0.22	12.8 (NB)
	B	0.32	6.72	12.20	0.45	25.5 (FB)

Table 1. Acquisition properties for the satellite images

For a comparative analysis in the further investigations, we considered the overlapping area of the three datasets, an area of 24 km², as highlighted in Figure 1. The rural region is characterized by forests and open areas. In the south more villages are found.

3. METHODOLOGY AND PROCESSING WORKFLOW

3.1 Satellite image processing and 3D reconstruction

The tri-stereo Pléiades images were provided as primary product, corrected only from sensor distortion, whereas the WorldView-3 images were delivered at an OR2A processing level, with relative radiometrically-corrected image pixels. Hence, as a pre-processing step, an optical radiometric calibration of the images is required before we make any comparison between them. Depending on the season, the atmospheric conditions, the sun's azimuth and elevation, the energy that satellite sensors record is different from the actual energy emitted or reflected from a surface on the ground. Since the value recorded in each pixel includes not only the reflected radiation from the surface but also the radiation scattered and

emitted by the atmosphere, the pixel values, i.e. Digital Numbers (DN), need to be corrected. For the Pléiades images we use the open source software Orfeo ToolBox (OTB Development Team, 2019) to perform the radiometric calibration. In this step, the pixels were corrected by the influence of the parameters like: sensor gain, spectral response, solar illumination, atmospheric pressure, optical thickness of the atmosphere, ozone and water vapour amount, and composition and amount of aerosol gasses. Since the OTB software does not support the WorldView-3 sensor, we computed the absolute radiometric calibration for each image (and independently for each band), by using the equations and parameters found in the technical sensor description (Kuester, 2016). This comprises two steps: (1) conversion from DN (raw data from the sensor) to top-of-atmosphere spectral radiance and (2) conversion from top-of-atmosphere spectral radiance to top-of-atmosphere reflectance. Optical radiometric calibration is important in training the classifier without dependence to the atmospheric conditions at the acquisition time, making it valid for a wider range of applications and not particular to one study case only. The optical calibration of the images is definitely required when investigating the transferability performance of the classifier across different acquisition times. We check the cross-time performance by applying the classifier trained on Pleiades 1st acquisition to the Pleiades 2nd acquisition.

The VHR satellite imagery from each tri-stereo acquisition are overlapping, a fact that makes the extraction of 3D information possible by applying photogrammetric techniques. The workflow for 3D point cloud derivation from the Pléiades and WorldView-3 imagery was performed in the Trimble Inpho software and comprises the following main steps:

- (1) image with metadata information (RPCs) and Ground Control Point (GCP) import;
- (2) GCP measurement in image space;
- (3) orientation refinement based on GCPs and Tie Points (TPs);
- (4) dense image matching for 3D reconstruction.

For improving the image orientation provided through the RPCs (to obtain a sub-meter accuracy), GCPs with known 3D coordinates are needed. Thus, in total, we employed 43, 73, and 36 GCPs for Pléiades first, second, and WorldView-3 acquisition, respectively. The points were measured by means of Real Time Kinematic (RTK) GPS with high accuracy of approx. 1 cm. During satellite triangulation Tie Points (TPs) were automatically extracted in all images, and they were further used together with the GCPs in a bundle block adjustment, to refine the initial values of the RPCs. From the resulting statistics, we considered the points with image residuals greater than one pixel as blunders, filtered them out and performed the refinement again.

The automatic computation and extraction of 3D information from the tri-stereo satellite imagery is possible through dense image matching. In our case, the algorithm finds the corresponding pixels between the images collected from the three different viewing points: forward, nadir and backward. For this purpose, Match T-DSM module of the Trimble Inpho software was used. During processing, ten pyramid levels are generated: the higher seven levels adopt a Feature Based Matching (FBM) while the lower three ones use a Cost Based Matching (CBM) strategy.

The ground coordinates of the corresponding image pixels (retrieved from image matching) are computed by applying forward spatial intersections by means of a least squares approach. This results in a “cloud” containing 3D points regularly distributed on the ground surface.

3.2 Manual labelling

To obtain training and testing data the three point clouds of the entire study area were manually classified using the application TerraScan from Terrasolid software. With its versatile visualisation options, the application allowed the manually labelling of the 3D point clouds into ground, trees, roads, buildings, and vehicles. Additionally, the corresponding orthophotos for each data acquisition were used in the thorough visual analysis.

3.3 Supervised classification

Once the satellite image-derived 3D point clouds are generated, we apply a tree based classification algorithm using machine learning for labelling the points into the five following classes: ground, trees, roads, buildings, and vehicles. In our case, the class ground comprises not only the bare-soil but also the covered ground, such as the agricultural- or grass-lands and all points not being included in the other classes.

The classification model comprises the following three main steps (Figure 2): (1) feature extraction, (2) training and (3) application of the trained model.

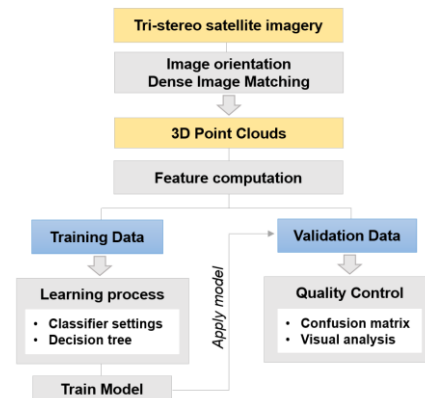


Figure 2. Overview of the general workflow

In a first step, additional geometric features were computed for each 3D point in the matched cloud. For this, in the spatial query of a point neighbourhood, we considered an infinite cylinder with 7 m search radius. The computed features include the normal vector components, features derived from the structure tensor, vertical point distribution, and surface roughness. These additional attributes describe the point distribution and are required for the separability of classes. Based on structure tensor eigenvalues, features like linearity, planarity or omnivariance could be computed. The first two have values between zero for non-linear/non-planar objects and one for linear/planar point distributions, respectively. The omnivariance feature gives information about the volumetric distribution of points. Features like EchoRatio, ZRange and ZRank describe the vertical point distribution, the maximum height difference between neighbouring points, and the rank of the point corresponding to its height, respectively. In addition to the hand-crafted 3D geometric features, the RGB colour information in each point was used for training the classifier. For the model learning process we used a total of 17 features: Red, Green, Blue, EchoNumber, Number of echos, Amplitude, Normal X, Normal Y, Normal Z, Normal sigma0, linearity, planarity, omnivariance, EchoRatio, NormalizedZ, dZRange, and dZRank (Bachhofner et al., 2020; Waldhauser et al., 2014). Like all supervised classification methods, the adopted decision tree requires training data, which are used by the machine

learning algorithm to build the classification model. In particular, we use the approach described in Waldhauser et al. (2014), where they apply CART for classifying point clouds from airborne laser scanning. In the training phase, the following information is required: (i) list of classes, (ii) hand-crafted 3D geometric and colour features and (iii) reference labelled dataset as training data. The classification tree seeks to partition the entire feature space of a data set, one variable at a time, by selecting a variable and an appropriate splitting value (Waldhauser et al., 2014). The decision tree is trained with the following hyperparameters: 0.00001 complexity factor, minimum 20 observations existent in a node (for splitting), at least 7 observations for each leaf node, a maximum depth of 30 for the tree, 5 competitor splits, and 5 surrogate splits (Bachhofner et al., 2020). Finally, to estimate how accurately our predictive model is, we test it on the validation dataset, containing labelled points.

3.4 Experiment design

The experiment is designed to characterize the repeatability of the semantic segmentation for satellite derived 3D point clouds. Thus the same area is studied, and also the same areas are used for training the classifier and testing the model. The setup of the experiment is to compare the classification results over the same areas for the different acquisitions.

For evaluating the precision and performance of the classifier, the variation of the results for three scenarios is analysed. For this, firstly we split the point clouds generated by image matching into five line-patches each with the size of 0.4 x 7 km and repetitively train and validate a new model by employing distinct combinations for the test and training data, like shown in Figure 3. Therefore, our model is trained on 40% of the data, while 60% is being used for validation. To obtain more stable and robust results, the experiments were repeated 5 times for each scenario and data acquisition and classification metrics are associated with standard deviations.

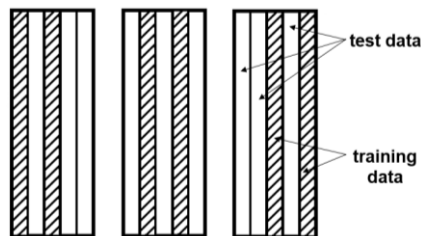


Figure 3. Overview of the different scenarios with the selections for the training and test data

The processing workflow for geometric features computation and classification is performed using the Opals (Orientation and Processing of Airborne Laser Scanning Data) software (Pfeifer et al., 2014). The approach for classification utilises CART (Classification and Regression Trees) modelling as implemented in the software package R (Therneau et al., 2019). The same processing chain (comprising the 3D point cloud extraction from the satellite imagery and the machine learning supervised classification algorithm) was applied independently in each scenario and for each of the three datasets: Pléiades first, second and WorldView-3 acquisitions.

As mentioned above, we chose classes that have different occurrences. In such situations, the overall accuracy measure alone is inappropriate, since the big number of examples from the majority classes overwhelms the number of examples in the minority classes. Therefore, to counter class imbalance we employ metrics such as (average) Precision, (average) Recall,

and F1-score for evaluating the prediction performance of the classifier.

4. RESULTS AND DISCUSSION

4.1 Reference data

The distributions of the number of points per class in each scenario and for each image acquisition are given in Table 2 as percentage values. The number of instances included in classes Buildings, Roads and Vehicles is far less than the ones included in Ground and Trees.

The distribution of the classes is similar between the acquisitions and also between the scenarios. The largest variation in absolute numbers is found between the two Pléiades point clouds on the one hand and the WorldView 3 point cloud on the other hand for the classes ground and trees. For the classes, roads and buildings the variation between the scenarios is larger than the variation caused by acquisition time.

Class	Percentage of points / Class					
	Pléiades (1 st acq.)		Pléiades (2 nd acq.)		WorldView-3	
	Train.	Test	Train.	Test	Train.	Test
Scenario 1						
Ground	52.748	60.257	52.887	60.376	66.474	71.664
Trees	45.351	38.119	45.250	38.064	31.682	26.708
Roads	1.641	1.430	1.610	1.366	1.544	1.423
Buildings	0.257	0.189	0.250	0.191	0.297	0.203
Vehicles	0.003	0.004	0.002	0.002	0.002	0.002
Scenario 2						
Ground	57.611	57.136	57.726	57.272	71.367	68.445
Trees	40.569	41.194	40.516	41.106	26.854	29.889
Roads	1.542	1.492	1.479	1.449	1.482	1.462
Buildings	0.276	0.174	0.279	0.169	0.295	0.201
Vehicles	0.003	0.004	0.001	0.003	0.001	0.003
Scenario 3						
Ground	58.007	56.868	58.166	56.975	69.159	69.971
Trees	40.576	41.185	40.492	41.119	29.393	28.137
Roads	1.232	1.704	1.163	1.665	1.248	1.622
Buildings	0.179	0.241	0.176	0.241	0.199	0.268
Vehicles	0.006	0.002	0.003	0.001	0.002	0.002

Table 2. Class distribution for each scenario and image acquisition (values given in percentages)

4.2 3D point cloud reconstruction from Pléiades and WorldView-3 tri-stereo satellite imagery

For each tri-stereo acquisition, the bundle adjustment was performed by employing all three images together with their RPCs, the RTK GCPs and the automatically extracted TPs. For the transformation in image space, the software estimated a correction model that contains two shifts (in X and Y) and a scale in Y-direction. An additional shift in X-direction was computed only for the Pléiades-second acquisition. The final standard deviations of the bundle block adjustment were at sub-pixel level, 0.54, 0.57 and 0.46 pixels for Pléiades first, second and WorldView-3 acquisitions, respectively. Through forward intersections, the three-dimensional positions of the points in object space (i.e. X, Y, Z) are determined.

The photogrammetric point clouds obtained from the tri-stereo satellite image matching contain not only the 3D coordinates, but also the reflectance information from the three spectral bands (Red, Green and Blue). In las file format and with a regular distribution (one point per each image pixel) the resulting point clouds have densities of 4 points/m² for Pléiades

and 12 points/m² for WorldView-3. They describe the terrain surface (bare soil, agricultural-, grass-lands) and the upper surface of the objects on it (vegetation and individual structures such as buildings, bridges). The resulted 3D point clouds are characterized by a poor geometry with smoothing effects in areas with a rapid change elevation (especially met at buildings roofs edges that look bevelled) and missing elevation information for small individual objects. In contrast to Pléiades, the WorldView-3 point cloud shows a better preservation of information and object details on the terrain surface (Figure 4).

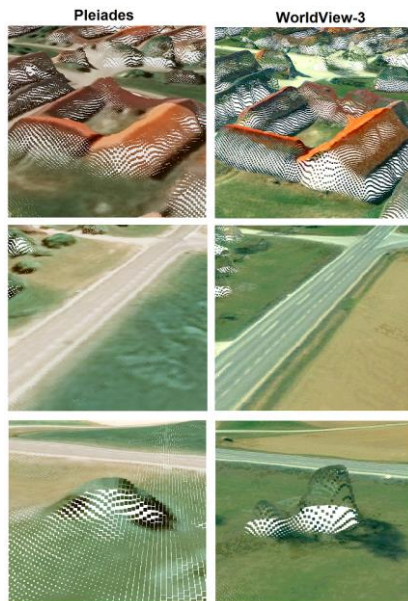


Figure 4. Comparative examples of Pléiades (left) and WorldView-3 (right) point clouds for buildings (first row), roads (second row) and trees (third row)

This can be explained by the better spatial resolution (0.31 m GSD) and the acquisition geometry with higher values for the viewing angles and a larger convergence angle on the ground (25°). All these factors allow a better 3D reconstruction resulting in a denser point cloud with reduced smoothing effect compared with Pléiades point clouds.

The total number of 3D points per scenario is around 54 mil. for Pléiades first and second acquisition, while for World View-3 is approximately 3 times higher (of 151 mil.).

4.3 Supervised classification and its repeatability for point clouds derived from Pléiades and WorldView-3 tri-stereo imagery

During classification, an object class label is assigned to each point, resulting in a 3D labelled point cloud. From a visual inspection, Figure 5 depicts a build-up area with the reference and classification results from all three acquisitions. Overall, the classifier identifies ground, trees, roads, and buildings classes. The only exemption are points belonging to vehicles, which were incorrectly classified as ground or trees. From a comparative point of view, no significant differences between the results for the two Pléiades acquisitions can be observed. In contrast, for WorldView-3 we can see some distinctions. For instance, as shown in the marked area A in Figure 5, in the WorldView-3 acquisition points are correctly classified as road, whereas in the Pléiades data they are wrongly labelled as ground. This is mainly because of the low Pléiades color contrast in this area. Road points have very similar geometric features with ground points and only color information is the distinctive attribute used for classification. Another notable difference appears in the marked area B in Figure 5. Whereas in the Pléiades acquisitions points are correctly identified as trees, in the World View-3 data only few of them are classified as trees and the rest as ground.

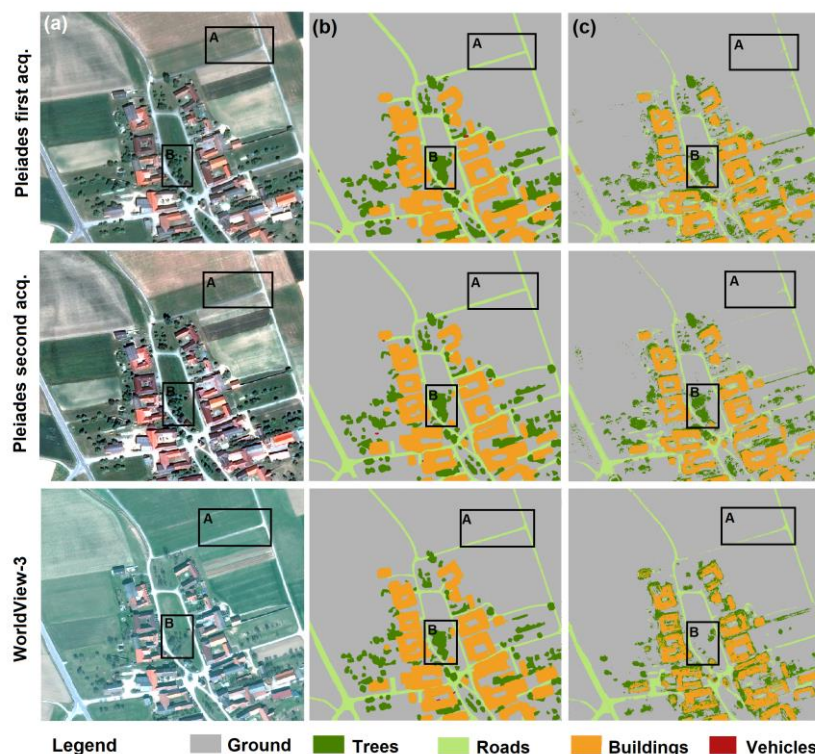


Figure 5. Top detail-view of Kleinschönau village (a) Pléiades and WorldView-3 ortho-rectified images; (b) the reference labelled 3D point cloud generated by image matching; (c) the classification result

This situation is caused by the different acquisition times: the two Pleiades in June 2017 (leaf-on conditions) and the WorldView-3 in April next year (leaf-off conditions). The leafless appearance in the WorldView-3 images caused difficulties in the 3D reconstruction, resulting in a point cloud with no or reduced elevation information. Hence, the points were misclassified as ground.

The confusion matrices are built for each scenario and iteration of the datasets and the achieved average results are shown in Table 3. In all tests, we achieve very high values for the accuracy (between 93.45% and 95.81%), but this alone is not a reliable performance metric to use, because of the imbalanced dataset.

Evaluation Metrics (%)		Pléiades (1 st acq.)	Pléiades (2 nd acq.)	WV-3
Average Accuracy	Scenario 1	94.58	94.51	95.81
	Scenario 2	94.82	95.02	95.54
	Scenario 3	93.45	94.12	95.50
Average Precision	Scenario 1	60.17	60.11	59.74
	Scenario 2	62.21	62.81	60.18
	Scenario 3	60.04	59.53	57.47
Average Recall	Scenario 1	68.84	67.15	68.23
	Scenario 2	65.58	65.81	67.03
	Scenario 3	66.78	68.12	69.28
Average F1 score	Scenario 1	64.18	63.54	63.72
	Scenario 2	63.75	64.35	63.39
	Scenario 3	63.05	63.53	63.03

Table 3. Comparison of evaluation metrics for all scenarios. In bold are the highest and the lowest values

Overall, the results are very consistent for the shown metrics, which integrate over the results of all the classes, independent of sensor and acquisition date. For the Pleiades data the results suggest that the scenarios, i.e. the choice of test data region, has more impact than the acquisition date. The WorldView results deviate stronger from the Pleiades data. At this stage it cannot be concluded if this is an effect of GSD, geometric accuracy, or acquisition time.

The qualitative results show that WorldView-3 classification metrics vary when compared with the two Pleiades acquisitions. For instance, the highest accuracy (95.81%) and average recall (69.28%) are reached in scenarios 1 and 3, respectively. Even higher variations between WorldView-3 and Pleiades can be observed at the individual class level (Figure 6). For the ground and road class both recall and F1-measure achieve better values in the WorldView-3 model. This confirms the visual analysis given above. The average recall raises with 3% for ground and with 8% for roads compared to Pleiades 1st acquisition and with 5% compared to Pleiades 2nd acquisition (for Scenario 1).

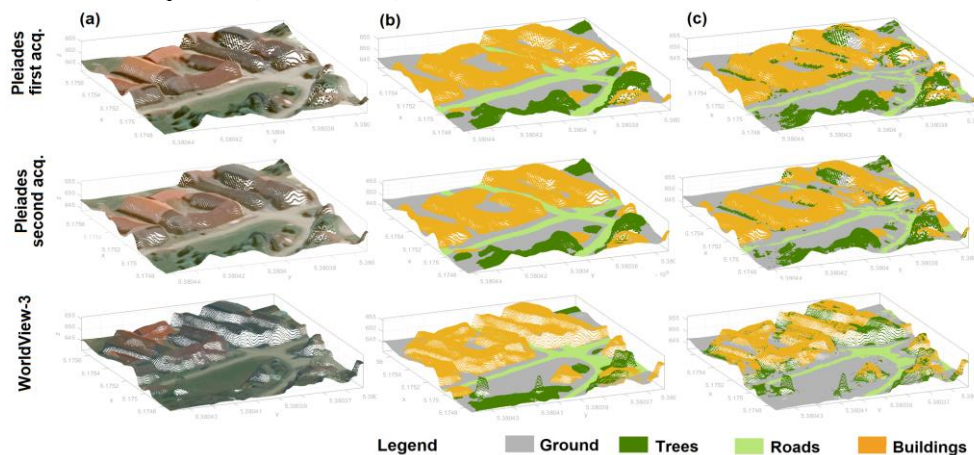


Figure 7. Detailed view of classification results (a) RGB point cloud; (b) reference labelled 3D point cloud; (c) classification result

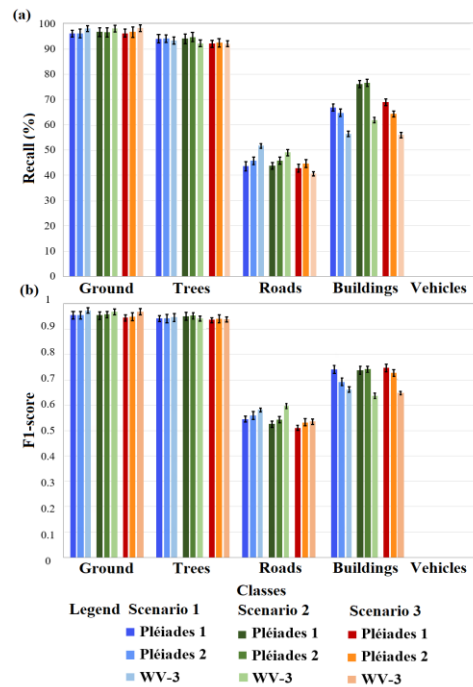


Figure 6. Comparison of average classification metrics with standard deviations (below 3%) for each class in all three scenarios: (a) recall and (b) F1-score

However, the statistics for buildings drop from the Pleiades to the WorldView results by a maximum of 15% (Figure 6). Given the clearer appearance of buildings in Figure 5, this is unexpected. This can be explained, on one hand, by the leafless trees in the WorldView-3 images, which have similar radiometric properties as buildings and are therefore misclassified as buildings. On the other hand, in all three acquisitions many points surrounding buildings are misclassified as trees (Figure 7c). These particular areas are raising difficulties for the classifier, because the two classes show similar geometric features. Even colour information does not make the difference here, due to shadowing effects.

At the individual class level, the unbalanced data makes the classifier biased toward the ground and trees classes, while dwarfing the buildings, roads, and vehicles. For the vehicle class, the precision, recall, and F1-score are 0.00% in all scenarios. This is because they were no points correctly classified as vehicles (no true positives). For the class roads the highest variation of ~7% F1-score is obtained between Scenario 2 and Scenario 3 when using the WorldView-3 data.

In the context of our investigation, repeatability refers to the agreement between the independent results, obtained by applying the same classifier to the same data within short interval of times. This was possible for the Pléiades acquisitions due to the temporal difference of only 7 days. Since land cover might change and trees appearance in images can differ, we considered buildings and roads as stable classes in time, with well-defined shapes and boundaries. Whereas the accuracy changes are small (below 0.15%) for both buildings and roads classes, the F1-score varies between -2.1 and 2.6 in scenario 3 (Figure 8). Besides the quality of the classifier, the repeatability results suggest also the stability of the Pléiades sensor itself, with respect to the spectral geometric precision and geolocation accuracy.

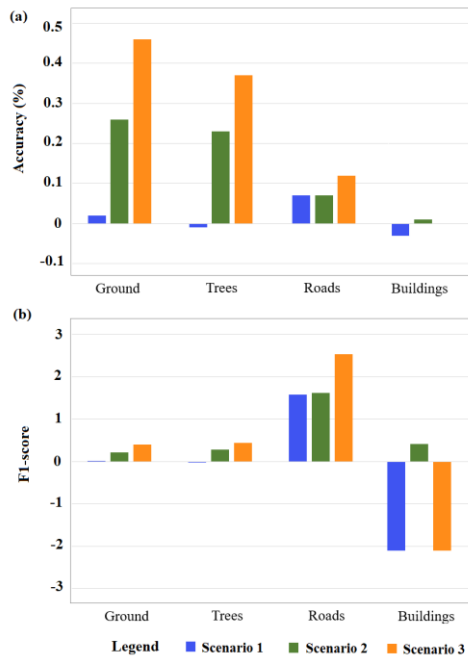


Figure 8. Comparison of classification metrics changes for each class (except vehicles) in all three scenarios: (a) accuracy and (b) F1-score changes between the two Pléiades acquisitions

The transferability performance of a classifier across different acquisition times is an important question. How does the classifier trained on a dataset perform on a new dataset for the same area, but from a different acquisition time? For testing the cross-time performance, we applied the classifier trained on the Pleiades 1st acquisition on the test set of Pleiades 2nd acquisition and vice versa and compared the prediction performance to the one achieved when training and test set are from the same acquisition time. The results are shown in Table 4, for the F1-score in each scenario.

		Tested on	
		Pleiades 1 st acq.	Pleiades 2 nd acq.
Trained on	Pleiades 1 st acq.	Scenario 1	
		Pleiades 2 nd acq.	64,26
	Pleiades 2 nd acq.	Scenario 2	
		Pleiades 1 st acq.	63,87
	Pleiades 1 st acq.	Scenario 3	
		Pleiades 2 nd acq.	61,82
Pleiades 2 nd acq.	Pleiades 1 st acq.	63,19	62,92
	Pleiades 2 nd acq.	60,97	63,56

Table 4. F1-score result for the cross-time performance

The F1-measures do not change significantly when the classifier is trained with data from another acquisition time. That proves a good transferability in time, which allows utilising the same classifier for future acquisition datasets for the same area.

Overall, the evaluation metrics show a slightly higher variation between the three scenarios when compared with the changes between the two Pléiades acquisitions (Figure 5). Hence, the selection of the training data for building the classifier has a higher influence on the performance, than the different acquisition times.

Another critical factor that influences the performance of the classifier is the correctness of the manual ground truth annotations in all three datasets. Due to the smoothed geometric appearance of the point clouds generated by image matching, with unclear object contours labelling is a challenging task and some uncertainty may occur.

Besides the variation of the classification metrics between Pléiades and WorldView-3, the higher GSD of the latter leads to higher computation times. With the same hardware configuration, an 2 x AMD EPYC 7302, 3GHz, 16-Core Processors, 512 GB RAM memory, (64-bit operating system), the processing times are approximately three times higher for World View-3 point clouds compared to Pléiades (Table 5).

Processing Step	Processing Time / Scenario		
	Pleiades (1 st acq.)	Pleiades (2 nd acq.)	WorldView-3
Feature computation	9h 42'	9h 11'	31h 40'
Train model	3h 17'	2h 54'	10h 46'
Apply model	28'	31'	1h 20'
Total time/iteration	13h 27'	12h 36'	43h 46'
Total (5 iterations – train and apply)	~18h 45'	~17h 5'	~60h 30'

Table 5. Processing times for classification

5. CONCLUSIONS AND FUTURE WORK

In this work, we analysed the performance of a supervised machine learning classification algorithm that exploits the full 3D content of point clouds derived from dense image matching of tri-stereo Very High Resolution (VHR) satellite imagery. The tree based classification method has been already successfully applied to aerial laser scanning data and point clouds from aerial image matching, but in contrast to previous research, in our study, we trained the decision tree on the geometric and color features of the satellite image driven 3D point clouds. This was a challenging investigation, since the geometric information precision of the satellite-driven point clouds is lower compared to the accurate geometric 3D position of laser scanning point datasets.

For evaluating the performance and quality of the classifier, we comparatively investigated the variation of the results by adopting three different scenarios for three different data sources: Pléiades (6 June 2017), Pléiades (20 June 2017) and WorldView-3 (8 April 2018). In each scenario, we have used 40% of the data for training and the rest of 60% for validation. The result of the classification are the labelled 3D point clouds; each point assigned to one of the following classes: ground, trees, roads, buildings, and vehicles. The unbalanced data leads to better classification metrics for the ground and trees classes, while dwarfing the buildings, roads, and vehicles classes. Due to the extremely low number of points belonging to the vehicle class, the algorithm was not able to recognize them. Our results

show that the adopted tree based supervised learning shows a good performance with overall accuracies

The agreement between the outcomes for the two Pléiades datasets with very low variations, assure that the reported quality measures did not result by chance. Moreover, it confirms the stability of the sensor itself, in terms of spectral geometric precision and geolocation accuracy. Additionally, we show the behaviour of the classifier when using a different sensor (WorldView-3) with a higher resolution and a distinct acquisition geometry. From our investigations, a factor of three could be established between WorldView-3 and Pléiades data, with respect to point cloud densities (the total number of points) and processing times. This lead to a moderate improvement for some classes (ground, roads). For the class building the higher resolution and geometric sharpness provided by WorldView-3 did not lead to higher quality.

The large training sets resulted in long training computation times. Therefore, in future work, Graphic Processing Unit (GPU) programming and parallelization schemes will have to be exploited to further reduce the computing time.

It will also be necessary to include strategies to counter the class imbalanced (data augmentation, resampling, or oversampling), for improving the performance of the model. Another direction of further research would be the use of the classified point clouds for 3D object reconstruction and modelling.

In summary, it can be stated that the adopted classifier is stable, providing a high potential in VHR satellite tri-stereo point clouds scene classification.

ACKNOWLEDGEMENTS

The authors express their gratitude to the company Vermessung Schmid ZT GmbH (DI. Karl Zeitberger) for providing the manually annotated point clouds used in the present study.

REFERENCES

Bachhofner, S., Loghin, A. M., Otepka, J., Pfeifer, N., Hornacek, M., Siposova, A., Hochreiter, R. 2020. Generalized Sparse Convolutional Neural Networks for Semantic Segmentation of Point Clouds Derived from Tri-Stereo Satellite Imagery. *Remote Sensing*, 12(8), 1289.

D'Angelo, P., Kusch, G. 2012. Dense multi-view stereo from satellite imagery. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, 6944-6947.

D. Deuge, M., Quadros, A., Hung, C., Douillard, B., 2013. Unsupervised feature learning for classification of outdoor 3D scans. *Australas Conference on Robotics and Automation*, (2)1.

Gerke, M., Xiao, J., 2013. Supervised and unsupervised MRF based 3D scene classification in multiple view airborne oblique images. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, II-3/W3, 25-30, doi.org/10.5194/isprsannals-II-3-W3-25.

Griffiths, D., Boehm, J., 2019. A Review on Deep Learning Techniques for 3D Sensed Data Classification. *Remote Sensing*, 11(12), 1499.

Grilli, E., Menna, F., Remondino, F., 2017. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 339.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., Pollefeys, M., 2017. Semantic3D.Net: A new large-scale point cloud classification benchmark. *ISPRS Annals of*

Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-1-W1, 91-98.

Kuester, M., 2016. Radiometric use of WorldView-3 imagery - Technical Note.

Leotta, M. J., Long, C., Jacquet, B., Zins, M., Lipsa, D., Shan, J., Purri, M., 2019. Urban Semantic 3D Reconstruction from Multiview Satellite Imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

Modiri, M., Masumi, M., Eftekhari, A., 2015. Automatic classification of point clouds extracted from UltraCam stereo images. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 40.

Munoz, D., Bagnell, J. A., Vandapel, N., Hebert, M., 2009. Contextual classification with functional max-margin Markov networks. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 975-982.

OTB Development Team, 2019. Orfeo ToolBox processing of remote sensing images software, Version 7.0.0. Open Source orfeo-toolbox.org (15 November 2019).

Otepka, J., Ghuffar, S., Waldhauser, C., Hochreiter, R., Pfeifer, N., 2013. Georeferenced point clouds: A survey of features and point cloud management. *ISPRS International Journal of Geo-Information*, 2(4), 1038-1065.

Panagiotakis, E., Chrysoulakis, N., Charalampopoulou, V., Poursanidis, D., 2018. Validation of Pleiades Tri-Stereo DSM in urban areas. *ISPRS International Journal of Geo-Information*, 7(3), 118.

Pfeifer, N., Mandlbürger, G., Otepka, J., Karel, W., 2014. OPALS – A Framework for Airborne Laser Scanning Data Analysis. *Computers, Environment and Urban Systems* 45: 125–136.

Piermattei, L., Marty, M., Karel, W., Ressel, C., Hollaus, M., Ginzler, C., Pfeifer, N., 2018. Impact of the acquisition geometry of very high-resolution Pléiades imagery on the accuracy of canopy height models over forested alpine regions. *Remote Sens-Basel*, 10, 1542.

Remondino, F., Spera, M. G., Nocerino, E., Menna, F., Nex, F. 2014. State of the art in high density image matching. *The photogrammetric record*, 29(146), 144-166.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012. The ISPRS Benchmark on Urban Object Classification and 3D Building Reconstruction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, I-3/ 1, 1(1), 293-298.

Therneau, T. M., Atkinson, E. J., 2019. An introduction to recursive partitioning using the RPART routines.

Vallet, B., Brédif, M., Serna, A., Marcotegui, B., Paparoditis, N., 2015. TerraMobilita/iQmulus urban point cloud analysis benchmark. *Computers & Graphics*, 49, 126-133.

Waldhauser, C., Hochreiter, R., Otepka, J., Pfeifer, N., Ghuffar, S., Korzeniowska, K.; Wagner, G., 2014. Automated classification of airborne laser scanning point clouds. In *Solving Computationally Engineering Problems*, Springer, 269–292.

Wehr, A., Lohr, U., 1999. Airborne laser scanning—an introduction and overview. *ISPRS Journal of photogrammetry and remote sensing*, 54(2-3), 68-82.

Xie, Y., Tian, J., Zhu, X. X., 2019. A review of point cloud semantic segmentation. *arXiv preprint arXiv:1908.08854*.