EVALUATION AND OPTIMISATION OF CROWD-BASED COLLECTION OF TREES FROM 3D POINT CLOUDS

Volker Walter[*], Michael Kölle, Yifey Yin

Institute for Photogrammetry (ifp), Universitaet Stuttgart, Germany
Geschwister-Scholl-Strasse 24D, 70174 Stuttgart
firstname.lastname@ifp.uni-stuttgart.de

Commission IV, WG 4

KEY WORDS: Crowd, Paid Crowdsourcing, Data Collection, 3D Point Clouds, Quality Improvement

ABSTRACT:

The term "Crowdsourcing" goes back to Jeff Howe (Howe, 2006) and represents a neologism of the words "crowd" and "outsourcing". Unlike outsourcing, where companies outsource certain tasks to known third parties, crowdsourcing outsources tasks to unknown workers (crowdworkers) on the Internet. This allows companies to access large numbers of workers who would otherwise not be available. In this paper, we will discuss an approach for the crowd-based collection of trees by means of minimum bounding cylinders from 3D point clouds. We will demonstrate the used web-interface and compare the results with reference data. To improve the quality of the results, we collect the data not only once but multiple times. This enables us to implement a so-called "Wisdom of the Crowd" approach where we can identify automatically outliers and derive integrated cylinders. We will show in this paper that this approach increases significantly the quality of the results.

1. INTRODUCTION

Quality control and improvement is a major challenge of crowd-based data collection (Liu et al. 2018, Leibovici et al., 2017) because data provided by crowdworkers can be erroneous (Vaughan, 2017). The crowd is composed of people with unknown and very diverse abilities, skills, interests, personal objectives, and technological resources (Daniel et al., 2018). Most crowdworkers are unfamiliar with the standards on spatial data collection and they may feel no need to follow such rules (Hashemi and Abbaspous, 2015).

The more complex the crowdsourcing task, the more heterogeneous will be the results of different crowdworkers. Another problem can be dishonest workers who try to maximize their income by submitting as many tasks as possible, even they did not complete the tasks or did the tasks only sloppy (Hirth et al., 2013). Additionally, spam and adversarial workers may exist who could be very harmful to the quality of the collected data (Zhang et al., 2016). Therefore, one of the fundamental challenges in crowdsourcing is inferring the ground truth from noisy data collected by non-experts (Zhou et al., 2012). Quality control and quality improvement is a hot research topic in crowdsourcing (Zhang et al., 2016).

We can distinguish two different approaches for the collection of geospatial data by the crowd: Collection by volunteers without payments (Volunteered Geographical Information – VGI [Goodchild, 2007]) and collection by paid crowdworkers. Crowdsourcing projects that are based on the work of unpaid volunteers need an active community who has an intrinsic motivation for collaboration. The main factors, that users collect voluntarily geospatial data, are that their contributions are made freely available and that other users benefit from providing digital maps (Budhathoki and Haythornthwaite, 2012). If this is not the case, other incentives must be provided such as monetary payments.

The most important VGI project is OpenStreetMap (OSM). OSM quality has been a subject of a considerable amount of research [e.g. (Fonte et al., 2017) or (Degrossi et al., 2018)]. The basic quality control concept of OSM is that users verify the data of other users, which leads to an increasing quality over time (Barron et al., 2014). However, in particular areas it may also happen that the quality decreases (Fonte et al., 2017). OSM has no centralized quality control (Degrossi et al., 2018), whereas in paid crowdsourcing the quality control is under the responsibility of the employer. Fonte et al. (2017) proposed several quality indicators to handle the specific nature of VGI data, such as demographic and socio-economic indicators, which are important when users collect data in areas in which they live, which is mostly not the case in paid crowdsourcing. The data collection in OSM is an open-end process, whereas campaigns in paid crowdsourcing are usually limited in time.

Generally, there are two different approaches to control and improve the quality of paid crowdsourced data (Zhang et al. 2016): "Quality Control on Task Designing" and "Quality Improvement after Data Collection". The first approach tries to guide the crowdworkers to provide high quality data. Many methods exist for this, such as qualification tests, reputation systems, task assignment, task and workflow optimization, training, real-time quality assurance, quality checkpoints or incentive payment mechanisms. A comprehensive overview of these techniques can be found in (Daniel et al., 2018).

In the second approach, additional procedures are used to improve the quality after the data has been collected. A common idea is the repeated data collection by different crowdworkers. After data collection, mechanisms are used to filter out noisy data and to infer the truth.

The process of estimating the truth from multiple collected data is called "ground truth inference". A ground truth inference

---

[*] Corresponding author

algorithm uses multiple collected noisy data as input and generates as output the estimated truth. If for example labels are collected multiple by different crowdworkers, a straightforward approach is to use the most common label. The employer duplicates the task and *n* different workers complete the task. The result, which has the majority, is assumed correct (Hirth et al., 2013).

Salk et al. (2016) examined the use of majority classification for the identification of croplands in remote sensing images. They defined a binary crowdsourcing task, evaluated the accuracy of the results, and compared them with expert validations. Hecht et al. (2017) realized an intrinsic quality control of semantic data to evaluate the results of crowdsourced classifications of building footprint data. In order to reduce noise, they collected the data multiple and aggregated the results with majority voting. Zhou et al. (2015) proposed to use the information measure *minimax conditional entropy*. They assumed that labels are generated by a probability distribution. By maximizing the entropy of this distribution, they can estimate true labels from a set of noisy labels.

Whereas majority classification can be easily realized for labelling tasks, it is difficult to use it for spatial data collection tasks. The reason is that labels can be classified into a finite number of classes whereas the geometry of a spatial object can have any shape. However, Walter and Sörgel (2018) found out that high quality spatial data can be achieved with paid crowdsourcing by collecting the data not only once but multiple times by many crowdworkers and then integrating the different representations into one common result.

This follows the idea of the "Wisdom of the Crowd". Surowiecki (2005) has shown in his book "Wisdom of the Crowd - why many are smarter than the few and how collective wisdom shapes business, economics, societies and nations" on many examples from very different fields that averages of multiple guesses are often better than the best individual guess. Large groups of people are smarter and can solve complicated problems even better than specialists can. For this, we need multiple representations as an input (which can be easily realized with paid crowdsourcing) and an averaging process that integrates the multiple results.

In this paper, we want to demonstrate the "Wisdom of the Crowd" principle on the crowd-based collection of trees from 3D point clouds. The interpretation of 3D point clouds is a non-trivial task and can be challenging for non-experts. Most of the existing work in the field of crowd-based geospatial data collection is based on 2-dimensional image data. To the best of our knowledge, (Herfort et al., 2018) is the only work that uses 3D point clouds in the context of crowdsourcing. We will demonstrate in this paper that groups of crowdworkers are smarter than individual crowdworkers are and that averaging multiple collected instances into one integrated representation leads to higher quality data.

The rest of the paper is organized as follows. In section 2 we will show the datasets from which the crowdworkers had to collect the data. The web-based interface, which contains the tools for the data collection, is demonstrated in section 3. In section 4 we discuss the quality evaluation approach and in section 5 the integration process is presented. The results of the data collection and data integration are shown in section 6. A discussion of the results can be found in section 7.

2. DATA

For our tests we used two datasets: the ISPRS Vaihingen 3D Semantic Labelling dataset (Niemeyer et al., 2014) (V3D) and a dataset from the State Office for Spatial Information and Land Development, Baden-Württemberg (Landesamt für Geo-information und Landentwicklung LGL, Baden-Württemberg) (M3D). Both datasets were collected by airborne laserscanning and coloured using rgb orthophotos. We selected five circular sections with a radius of 50 m from the V3D dataset and 14 sections with the same extent from the M3D dataset. The crowdworker had to collect the trees only in an inner area of the circle with a radius of 30 m. We carefully collected the reference data for both datasets by ourselves.

The V3D dataset has a point density of 4 to 7 points/m² and contains mainly detached houses with surrounding gardens with trees and bushes (see Figure 1). The M3D dataset has a point density of 4 to 32 points/m² and contains the surrounding of the Mercedes Benz Museum, Stuttgart with single-line and double-line tree rows and groups of trees (see Figure 2).

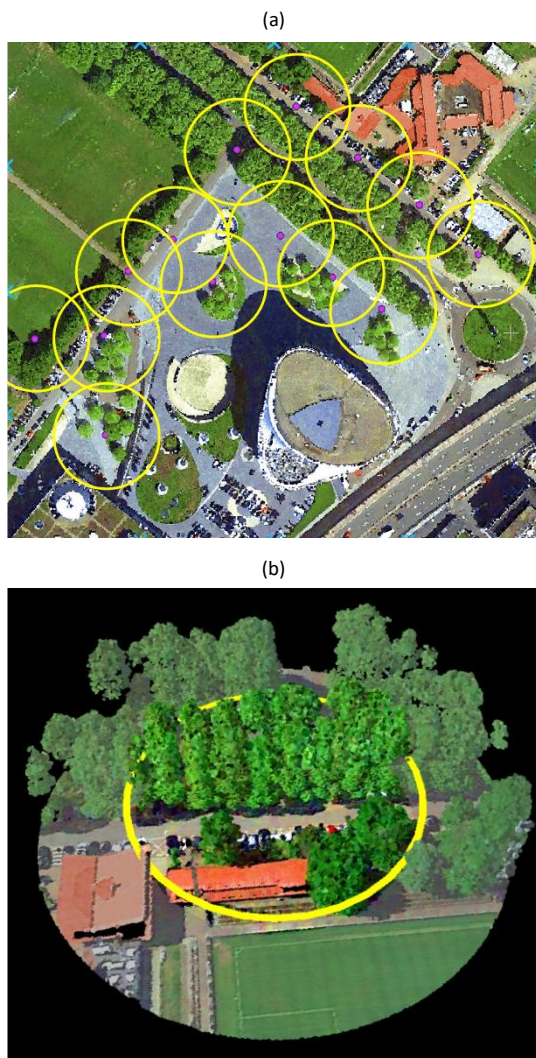Figure 1. Overview (a) and example of one section (b) of the V3D-dataset

(a)

(b)

Figure 2. Overview (a) and example of one section (b) of the M3D-dataset

3. CROWD-BASED DATA COLLECTION

Figure 3 shows the web-based Graphical User Interface of the program for the crowd-based collection of trees. The interface was developed with JavaScript and HTML. The 3D visualisation is realized with the JavaScript 3D library *Three.js* (Cabello, 2019). The functionality on the server is implemented with PHP. The program consists of three parts:

Visualisation of the point cloud: the users can rotate, move and zoom the point cloud with the mouse. It is possible to reset the view or to select between different standard views.

Functionality for the collection of the data: the users must click on *New* in order to add a new cylinder to the 3D scene. The position of the new cylinder can be changed by dragging the mouse. The radius and the height of the cylinder can be changed by clicking on the corresponding control buttons. The bottom of the cylinder is automatically adjusted to the terrain.

Management of the cylinders: all collected cylinders are shown in a list. Wrongly collected cylinders can be deleted from the list. An already collected cylinder can be activated by clicking on the corresponding element in the list. There is

always only one cylinder active. If a cylinder is active, the position, size, and radius of this cylinder can be changed. The crowdjob can be finished by clicking on the *Submit* button. The collected data will then be submitted to the server.

Figure 3. Web-tool for the crowd-based collection of trees from 3D point clouds

All crowdjobs were published on the commercial platform microWorkers (www.microworkers.com) which takes over the recruitment and the payment. According to their website, the platform has access to more than 1,500,000 registered crowdworkers (January 2019). The payment is \$0.10 for each job plus \$0.01 per collected tree.

4. QUALITY EVALUATION

A common approach for quality evaluation is to subdivide all collected data instances into the categories: *True Positive TP*, *False Negative FN*, *False Positive FP* and *True Negative TN* (compare: Heipke et al. 1997). From that we can calculate *Completeness* = TP / (TP + FN), *Correctness* = TP / (TP + FP) and *Quality* = TP / (TP + FP + FN) where TP, FP and FN are the numbers of cylinders of each category.

The problem is that not all crowd-based collected cylinders can be matched uniquely 1:1 to a reference cylinder. Figure 4 shows on three examples that also 1:n, n:1 and n:m relations are possible. Therefore, we evaluate the quality of the crowd-based collected data using an approach of Rottensteiner et al. (2005) which is based on the evaluation of mutual overlaps. In our application, we extend the approach to three dimensions.

The set of crowd-based collected cylinders is C_n and the set of reference cylinders is C_r. For each cylinder $c_n \in C_n$ and $c_r \in C_r$ we calculate the ratios q_{nr} and q_{rn} with $q_{nr} = V_{n \cap r}$ / V_n and $q_{rn} = V_{n \cap r}$ / V_r where $V_{n \cap r}$ is the intersecting volume of the crowd-based collected cylinder c_n and a reference cylinder c_r, V_n is the volume of cylinder c_n, and V_r the volume of cylinder c_r.

For the classification of the overlap we define $U(c_i, c_j)$ = *strong* **if**: $q_{ij} > 80\%$, *partial* **if**: $50\% < q_{ij} \leq 80\%$, *weak* **if**: $10\% < q_{ij} \leq 50\%$, *none* **if**: $q_{ij} \leq 10\%$. $U(c_i, c_j)$ is not symmetrical because $U(c_i, c_j)$ does not necessarily equal $U(c_j, c_i)$. If both $U(c_i, c_j)$ and $U(c_j, c_i)$ are in the categories *weak* or *none*, we expect that no correspondence between c_i, and c_j exist.

After the elimination of the cylinders without correspondences, for every crowd-based collected cylinder c_n we have a subset C_{nr} of corresponding reference cylinders and for every reference cylinder c_r a subset C_{rn} of corresponding crowd-based cylinders.
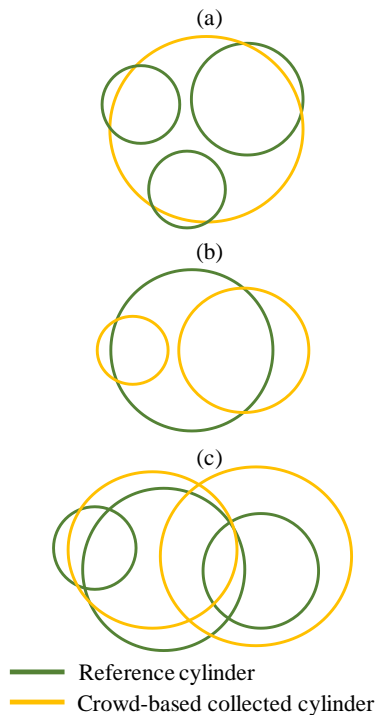
Figure 4. Examples of different relationships between reference data and crowd-based collected data: (a) *n*:1 relationship: three reference cylinders are collected with one cylinder by a crowdworker (b) 1:*n* relationship: one reference cylinder is collected with two cylinders by a crowdworker (c) *n:m* relationship: three reference cylinders are collected with two cylinders by a crowdworker

For each cylinder $c_n \in C_n$ and $c_r \in C_r$ we calculate now the ratios of total overlap $d_n = V_{Cnr \cap n} / V_n$ and $d_r = V_{Crn \cap r} / V_r$ with $V_{Cnr \cap n}$ is the intersecting volume of C_{nr} with c_n and $V_{Crn \cap n}$ is the intersecting volume of C_{rn} with c_r. Based on the ratios d_n and d_r, we subdivide all cylinders $c_n \in C_n$ and $c_r \in C_r$ into four categories according to table 1.

Criteria	crowd cyl. C_n	reference cyl. C_r
$d > 80\%$	*Fully correct*	*Fully complete*
$50\% < d \leq 80\%$	*Partially correct*	*Partially complete*
$10\% < d \leq 50\%$	*Hardly correct*	*Hardly complete*
$d \leq 10\%$	*Not correct*	*Not complete*

Table 1. Categories of cylinders based on an evaluation of the mutual overlap

Based on these categories we can refine the definition of *TP*, *FN* and *FP*:

- $TP_{Completeness}$: number of reference cylinders that are fully or partially collected by the crowdworkers
- $TP_{Correctness}$: number of crowd-based collected cylinders that are fully or partially correct
- *FN*: number of reference cylinders that are not fully or partially collected by the crowdworkers
- *FP*: number of crowd-based collected cylinders that are collected only hardly correct or not correct

Because of the new definition of $TP_{Completeness}$ and $TP_{Correctness}$ we must redefine the *Quality* (Rutzinger et al. 2009):

$$Quality = \frac{1}{1 + \dfrac{FN}{TP_{Completeness}} + \dfrac{FP}{TP_{Correctness}}} \quad (1.1)$$

Instead of defining *TP*, *FP* and *FN* as total number of cylinders of each category, we use the total volumes of the corresponding cylinders to weight trees with large volumes higher than trees with small volumes. This affects also the *Completeness* and *Correctness*.

5. INTEGRATION OF MULTIPLE COLLECTED CYLINDERS

The multiple collected cylinders are integrated in two steps. First, we use a DBSCAN clustering algorithm to detect clusters and remove outliers by evaluating the *x*- and *y*-coordinates of the cylinders. In a second step, we integrate all cylinders of each cluster.

5.1 Clustering with DBSCAN

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) is a density-based algorithm for the detection of clusters and outliers (Ester et al. 1996). The advantages of DBSCAN are that the number of clusters must not be specified in prior - like in K-means - and that it is robust to outliers.

DBSCAN requires two parameters: (1) *Epsilon* defines the maximum distance between two points to be considered as neighbours and (2) *MinPts* defines the minimum number of points in a cluster. Based on empirical tests we use *Epsilon* = 2.5 m and *MinPts* = 4 as parameter setting.

5.2 Integration

DBSCAN removes outliers by evaluating the positions of the cylinders. For the integration of the cylinders of each cluster, we use an iterative approach that removes additional outliers by evaluating also the radius and the heights of the cylinders.

As start for the iteration, we calculate an integrated cylinder $C_{Integrate}$ by averaging *x*, *y*, *radius* and *height* of the cylinders of each cluster. In the next step, we calculate the intersecting volumes $V_{Intersect\_k}$ of the integrated cylinder $C_{Integrate}$ with volume $V_{Integrate}$ and all other cylinders C_k of the same cluster with volumes V_k.

Then, we calculate the two ratios $q_{Integrate}$ and q_k, which describe the mutual overlap: $q_{Integrate} = V_{Intersect\_k}/V_{Integrate}$ and $q_k = V_{Intersect\_k}/V_k$. If $q_{Integrate}$ or q_k is smaller than 50 % the corresponding cylinder C_k is marked as outlier and removed from the cluster and a new iteration is started. Figure 5 shows examples of possible outliers in a bird view.

6. RESULTS

6.1 Results of the crowd-based data collection

The crowdworkers executed 50 jobs on the V3D dataset and 140 jobs on the M3D dataset. Table 2 shows the average working time per crowd job, the average number of collected trees per crowd job and the average number of trees in the corresponding reference.

$q_{Integrate} < 50\%,\ q_k > 50\%$

$q_{Integrate} > 50\%,\ q_k < 50\%$

$q_{Integrate} < 50\%,\ q_k < 50\%$

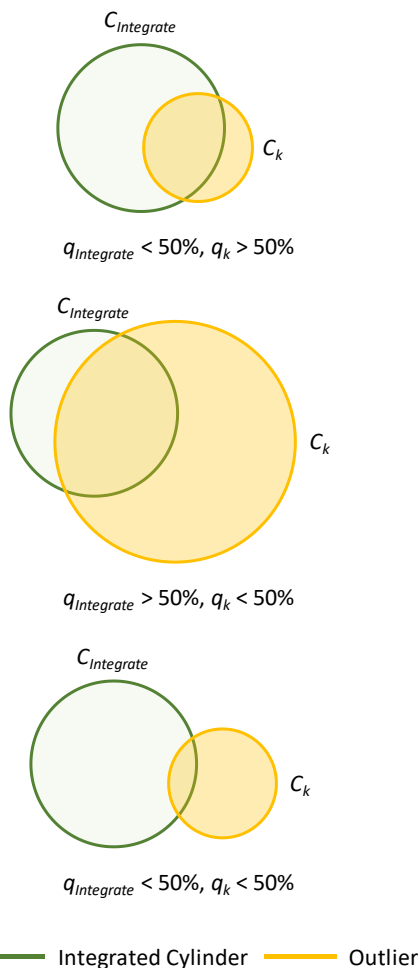—— Integrated Cylinder   —— Outlier

Figure 5. Detection of outlier based on an evaluation of the mutual overlap of the integrated cylinder with the other cylinders of each cluster (2D-view from above, green: integrated cylinder, orange: outlier)

Dataset	Average working time [s]	Average number of collected cylinder	Average number of trees in references
V3D	407	7	9
M3D	518	11	13

Table 2. Average working time per crowd job, average number of collected cylinders per crowd job and average number of trees in the corresponding reference for the V3D and M3D datasets

The working time needed to execute one crowd job is typically between 6 to 10 minutes. It can be seen that the average number of collected cylinders is smaller than the average number of trees in the reference that indicates that often not all trees were collected or that several closely neighbouring trees in the reference were collected with only one cylinder by the crowdworkers.

Table 3 shows the quality measures of all crowd jobs for both datasets. While the average *Correctness* is clearly higher than 90%, the *Completeness* is in both datasets lower. The reason is that falsely collected trees are more often in the class *False Negative* as in the class *False Positive*. This means that the trees

that were collected by the crowdworkers are in most cases correct but the crowdworkers were unable to identify all trees in the point cloud.

Dataset	Average Completeness	Average Correctness	Average Quality
V3D	85.19	92.64	79.37
M3D	92.42	96.82	89.58

Table 3. Average *Completeness* [%], *Correctness* [%] and *Quality* [%] of all crowd jobs of dataset V3D and M3D

The quality parameters of the M3D dataset are significantly higher as the quality parameters of the V3D dataset. The reason is that the trees in the V3D dataset are often surrounded by other vegetation (bushes and undergrowth) whereas this is not the case in the M3D dataset. Another reason could be the different point densities of both datasets. The point density of the V3D dataset is between 4 and 7 points/m² whereas the point density of the M3D dataset is between 4 and 32 points/m². A higher point density can help with the visual interpretation of the point clouds.

6.2 Results of the integration

Each crowdjob was processed by ten crowdworkers. Altogether, 346 cylinders were collected from the V3D dataset and 1537 cylinders from the M3D dataset. Figure 6 shows the *x*- and *y*-positions of all collected trees on an example of one of the sections of the V3D dataset. The outliers detected with DBSAN are marked with red colour. The remaining trees were subdivided into different cluster (Figure 7). The result of the outlier detection based on the evaluation of the heights and the radius is shown in Figure 8 and the final clusters in Figure 9. Table 4 summarizes the results of the clustering and outlier detection of all crowdjobs for both datasets.
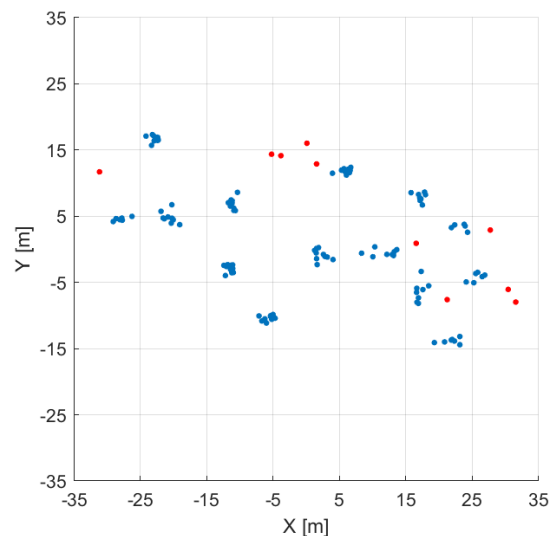

Figure 6. All collected cylinders (*x*- and *y*-position) of one section of the V3D dataset. Outlier detected with DBSCAN are marked with red colour

Dataset	Total number of cylinders	Outlier	Total number of clusters
V3D	346	56 (16.18%)	36
M3D	1537	185 (12.04%)	176

Table 4. Results of the clustering and outlier detection

Table 5 shows the quality measures of the integrated cylinders. The *Correctness* is 100 % in both datasets, which means that all integrated cylinders have a 1:1 correspondence in the reference data set. Therefore $FP = 0$, which leads to identical numerical values of the *Completeness* and the *Quality*. The *Quality* of the M3D dataset is still slightly higher than the *Quality* of the V3D dataset. Both measures are near 100 %.

Dataset	Completeness	Correctness	Quality
V3D	97.28	100.00	97.28
M3D	99.68	100.00	99.68

Table 5. *Completeness* [%], *Correctness* [%] and *Quality* [%] of all crowd jobs of the integrated cylinders of V3D and M3D

Figure 7. Remaining cylinders (*x*- and *y*-position) are subdivided with DBSCAN into clusters which are marked with different colours
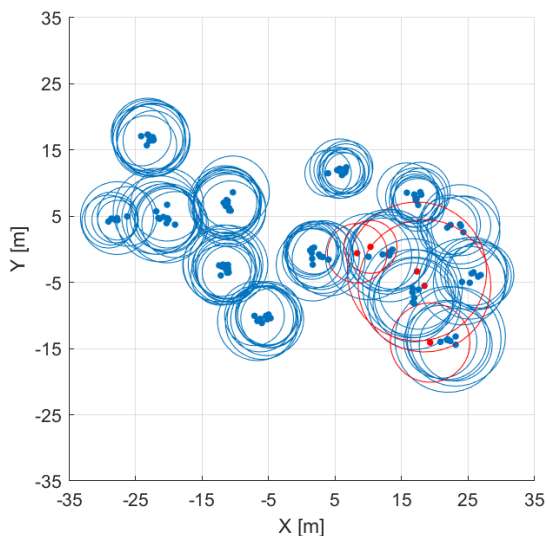

Figure 8. Further outlier detection based on the evaluation of the heights and the radius of the cylinders (cylinders in bird view)
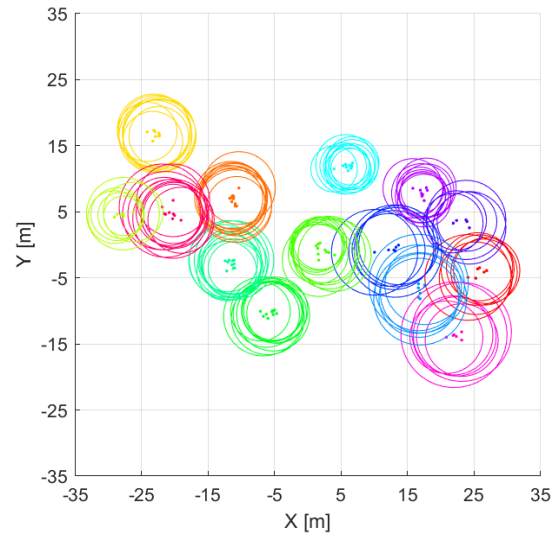

Figure 9. Final result of outlier detection and integration

Figure 10 shows visually the significant increase of the quality measures after the data integration for both datasets. Figure 11 shows the input data, the reference data and the result of the integration on one example data section of the V3D dataset.
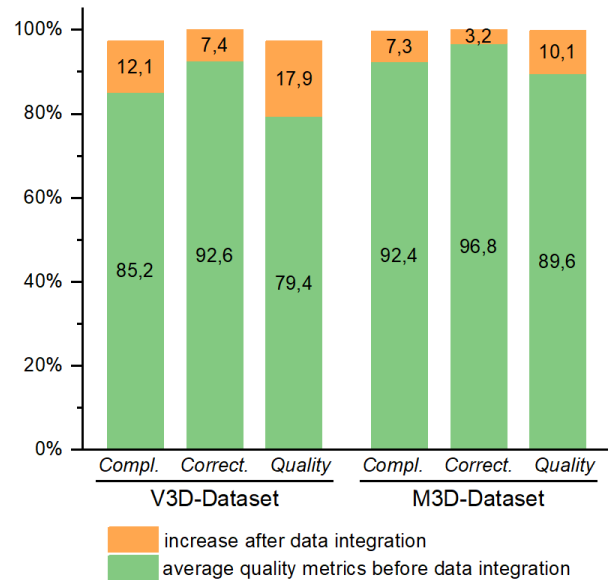

Figure 10. Increase of the *Completeness* [%], *Correctness* [%] and *Quality* [%] after data integration

7. DISCUSSION

Paid crowdsourcing can be a powerful tool to collect spatial data. However, the main problem is that the quality of the data can be very heterogeneous. This is not only because the objects are collected by untrained individuals but also because of the subjective nature of data collection (Walter, Sörgel, 2018).
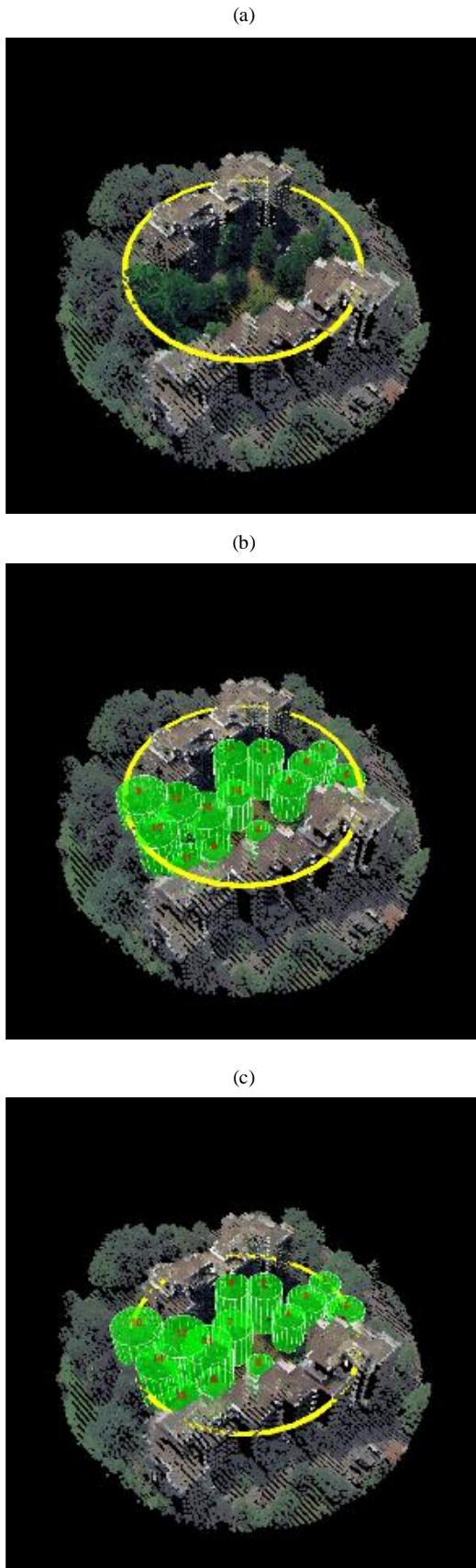
(a)

(b)

(c)

Figure 11. Input data (a) reference data (b) result of the integration (c) of one example data section of the V3D dataset

One method to increase the quality of crowd-based data collection is to collect the data not only once but multiple and integrate the multiple instances into one representation. This is the idea of the "Wisdom of the Crowd" that says that groups are smarter than individuals are, even if the individuals are experts.

We tested this idea on the crowd-based collection of trees from 3D point clouds. For this, we implemented a method for the integration of multiple cylinders with a simultaneous outlier detection and removal. Each crowdjob was duplicated 10 times. We compared the quality of the individual collected cylinders with the quality of the integrated cylinders. The quality of the integrated cylinders is significant higher as the quality of the individual cylinders. We were able to achieve a *Correctness* of 100 % and a *Completeness* and *Quality* of higher than 97 % for the V3D dataset and higher than 99 % for the M3D dataset.

Methods based on ergonomics and psychology are not part of this research. However, they can be combined with the proposed methods for a further improvement of the quality.

ACKNOWLEDGMENT

REFERENCES

Barron, C., Neis, P., Zipf, A., 2014. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Trans. GIS* 18, 877-895.

Budhathoki R., Haythornthwaite C., 2013. Motivation for Open Collaboration. Crowd and Community Models and the Case of OpenStreetMap. *American Behavioural Scientist* 57(5), 548-575.

Cabello, R., 2019. Three.js [WWW Document]. URL: https://threejs.org (accessed 15.1.20).

Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., Allahbakhsh, M., 2018. Quality Control in Crowdsourcing. A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions. *ACM Comput. Surv*. 51(1), Article 7, 40 pages.

Degrossi, L.C., Albuquerque, J.P., Rocha, R.S., Zipf, A., 2018. A taxonomy of quality assessment methods from volunteered and crowdsourced geographic information. *Trans. GIS* 22, 542–560.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering cluster in large spatial databases with noise. In. Evangelos Simoudis, Jiawei Han, Usama M. Fayyad (Eds.). Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press, 1996, 226-231

Fonte, C., Antoniou, V., Bastin, L., Estima, J., Jokar Arsanjani, J., Laso Bayas, J., See, L., Vatseva, R., 2017. Assessing VGI Data Quality. In. Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.M., Fonte, C.C., Antoniou, V. (Eds.). Mapping and the Citizen Sensor. Ubiquity Press Ltd., 137-163.

Goodchild, M.F., 2007. Citizens as sensors. the world of volunteered geography. *GeoJournal* 69(4), 211–221.

Hashemi, P., Abbaspour, R.A., 2015. Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept. OpenStreetMap in GIScience. Springer, 19–36.

Hecht, R., Kalla, M., Krüger, T., 2018. Crowd-sourced data collection to support automatic classification of building footprint data. Proceedings of the International Cartographic Association, 10.5194/ica-proc-1-54-2018.

Heipke, C., Mayer, H.M. Wiedemann, C., Jamet, O, 1997. Evaluation of Automatic Road Extraction. In: International Archives of Photogrammetry and Remote Sensing 32, 47-56.

Herfort, B., Höfle, B., Klonner, C., 2018. 3D micro-mapping. Towards assessing the quality of crowdsourcing to support 3D point cloud analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 137(3), 73-83.

Hirth, M., Hoßfeld, T., Tran-Gia, P., 2013. Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modelli*ng 57(11–12), 2918-2932.

Howe, J., (2006. The rise of crowdsourcing. *Wired Magazine* 14(6), 1–4.

Leibovici, D.G., Rosser, J.F., Hodges, C., Evans, B., Jackson, M.J., Higgins, C.I., 2017. On Data Quality Assurance and Conflation Entanglement in Crowdsourcing for Environmental Studies. *ISPRS International Journal for Geo-Information* 6(3), 78 pages.

Liu, Z., Shabani, S., Glassey Balet, N., Maria, S., Cretton, F., 2018. How to motivate participation and improve quality of crowdsourcing when building accessibility maps. In. 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), 1-6.

Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 152–165.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion* 6(4), 283–300.

Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2(1), 11–20.

Salk, C.F., Sturn, T., See, L., Fritz, S., Perger, C., 2016. Assessing quality of volunteer crowdsourcing contributions. Lessons from the Cropland Capture game. *International Journal of Digital Earth* 9(4), 410-426.

Surowiecki, J., 2005. The Wisdom of Crowd – why many are smarter than the few and how collective wisdom shapes business, economics, societies and nations. Doubleday, New York.

Vaughan, J.W., 2017. Making Better Use of the Crowd. How Crowdsourcing Can Advance Machine Learning Research. *Journal of Machine Learning Research* 18(193). 1-46.

Walter, V., Soergel, U., 2018. Implementation, Results and Problems of paid Crowd-based geospatial Data Collection. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 86(3-4), 87-197.

Zhang, J., Wu, X., Sheng, V., 2016. Learning from crowdsourced labelled data. a survey. *Artif. Intell. Rev.* 46(4), 543-576.

Zhou, D., Platt, J.C., Basu, S., Mao, Y., 2012. Learning from the Wisdom of Crowds by Minimax Entropy. *Advances in Neural Information Processing Systems (NIPS)* 25, 2204-2212.

Zhou, D., Liu, Q., Platt, J.C., Meek, C., Shah, N.B., 2015. Regularized minimax conditional entropy for crowdsourcing. Technical Report arXiv.1503.07240.