Crowd Controls Crowd: Quality Improvement of Polygon Integration in Paid Crowdsourcing

David Collmar[1], Volker Walter[1], Uwe Soergel[1]

[1] Institute for Photogrammetry (ifp), University of Stuttgart, Germany
{david.collmar, volker.walter, uwe.soergel}@ifp.uni-stuttgart.de

Keywords: Crowdsourcing, Quality Control, Data Enhancement, Cost Optimization, Polygon Integration, Wisdom of the Crowd

Abstract

In this study, we introduce a crowd-driven data enhancement strategy for the integration of polygons in paid crowdsourcing. First, we capture redundant polygons with a web-based tool using one set of crowdworkers. Then, we present the acquired polygons to other crowdworkers in a polygon editing tool, with instructions to validate and improve those acquisitions. This procedure is repeated by showing a third set of crowdworkers the already edited polygon geometries. Furthermore, a polygon integration procedure is performed for every step, i.e., the unedited polygons, those edited once and those edited twice, allowing for a comparative qualitative analysis. This analysis is conducted with a focus on both quality and cost control, aiming for small sample sizes in order to optimize costs. Additionally, we conduct further investigations to assess the effectiveness of our approach in different use cases, and explore potential adaptions for further enhancement.

1. Introduction

Crowdsourcing, a neologism that combines “crowd” and “outsourcing” (Howe, 2006), has been established as a standard technique in the acquisition of training data for machine learning applications (Jin et al., 2020). This includes obtaining of training data for image classification (Saralioglu and Gungor, 2020), which is vital for CNNs and other machine learning architectures, and therefore in high demand (Stonebraker and Rezig, 2019).

In general, crowdsourcing can be categorized into two types: voluntary and paid. Voluntary crowdsourcing is driven by intrinsic motivation, where participants engage because tasks might align with their interests (Hossain, 2012), thereby often resulting in high-quality output (Mason and Watts, 2009). However, it can be challenging to generate this interest to leverage intrinsic motivation, so paid crowdsourcing is an alternative. Paid crowdsourcing involves compensating workers, typically through designated platforms such as microWorkers.com, which facilitate the recruitment and payment of workers (Hirth et al., 2011). Tasks listed on these platforms are typically characterized by short durations, modest compensation, and low complexity (Hirth et al., 2011). However, the rather extrinsic motivation in paid crowdsourcing can influence the quality of the output due to various factors such as lack of motivation (Chandler et al., 2013), which is why data quality in paid crowdsourcing remains a constant topic of discussion (Rea et al., 2020).

As a result of these problems, a wide range of strategies have been developed to improve data quality (Zheng et al., 2017). Typical approaches are „quality control on task design” and “quality improvement after data collection” (Zhang et al., 2016). In the first approach, the focus lies on defining crowd tasks in such a way that they can be solved quickly and easily (Hirth et al., 2011), in order to ensure initial quality during a worker’s task performance. The second approach involves collecting redundant data of multiple different workers to leverage the principle of “wisdom of the crowd” (Jin et al., 2020), effectively countering the effects of poor-quality submissions (Zhang et al., 2016) that can otherwise be hard to identify (Li et al., 2016).

The aforementioned low task complexity, a typical characteristic of crowdsourcing, enables the majority of crowdworkers to solve the respective job effectively. Consequently, the integration of their results through majority voting is an effective method for improving quality (Zhang et al., 2016; Jin et al., 2020). Majority voting is renowned for its robustness and simplicity in implementation (Zheng et al., 2017), making it one of the most efficient aggregation methods (Tu et al., 2018). Additionally, the effectiveness of majority voting is not limited to its simple implementation; it can be applied to different kinds of data: While the use of majority voting for classifications seems rather obvious, majority voting even allows for usage in more complex problems such as the raster-based integration of polygons, where geometries are integrated on a pixel-level (Collmar et al., 2023a).

Integrating geometric data via such a pixel-level majority vote leads to the beneficial effect of an overall improvement in quality, but is also dependent on the number of polygons used as input, i.e., the amount of redundantly collected acquisitions: A large number of acquisitions leads to results of higher quality than a small number, making a large number of input data desirable for a quality-oriented approach (Collmar et al., 2023a). However, paid crowdsourcing is tightly coupled to the trade-off between cost and data quality (Li et al., 2017): Typically, smaller sample sizes are chosen to control costs, contradicting the need for larger samples to ensure data quality.

In order to address this challenge, i.e., ensuring high data quality even for small sample sizes, we propose an approach for enhancing the output quality of geometric data integration. This method is designed to be feasible even for small sample sizes, optimizing the mentioned trade-off in achieved quality and acquisitions cost. Fully embracing the “wisdom of the crowd” principle, we task crowdworkers with data refinement of previously collected polygons before performing the integration, asking them to review and optimize acquired geometries via editing. Subsequently, data integration is performed on the edited polygons. We then evaluate all output data by considering both total cost and data quality, allowing us to gain insights about the relationship between sample size and respective output. This allows us to maintain small sample sizes and thereby reduce costs while maximizing the output quality of crowd acquisitions.

2. Methodology

We explore the possibility of crowd-driven data refinement for polygon integration in a multi-step process by providing crowdworkers with the results of previous crowdsourcing campaigns, aiming for an improvement in data quality after integration.

The detailed methodology is shown in Figure 1. In the first step, depicted as **A** for acquisition, we ask crowdworkers to capture tree outlines via polygons from high-resolution orthophotos. The input data consists of *k* image sections, where each image section contains one single tree. Every image section is processed by *n* crowdworkers, leading to *n* polygons per section. These *n* polygons, in Figure 1 depicted as $P_{j,1}$ to $P_{j,n}$, where *j* is the number of the image section, are used as input for two different procedures: Firstly, a polygon integration that is used to calculate a single output polygon per image section *j*, which is called PI_j. Secondly, as input for the second crowd campaign, called **E₁,** which stands for (first) editing. Here, every previously acquired polygon is reviewed and edited by yet another crowdworker, leading to *n* once-edited polygons per image section, named $P'_{j,1}$ to $P'_{j,n}$. A polygon integration procedure is performed here as well, leading to an integrated polygon PI'_j for each image section. In the following step **E₂**, the second editing, these previously edited polygons are now reviewed and edited a second time by different crowdworkers, leading to *n* polygons that are named $P''_{j,1}$ to $P''_{j,n}$. Following the same procedures as before, an integration is performed again, leading to yet another integrated polygon, called PI''_j.

Overall, this leads to three different integrated polygons per image section *j*:

- The integrated polygons PI_j, which are calculated using the unedited polygons that were acquired via the acquisition procedure (A).
- The integrated polygons PI'_j, which are calculated using the results of the first editing procedure (E_1).
- The integrated polygons PI''_j, which are calculated using the results of the second editing procedure (E_2).

These integrated polygons are compared with ground truth data to assess the improvement achieved through the described processing methods.

However, there are different costs for the various approaches that must be taken into account: While the unedited polygons, acquired via the acquisition procedure (A), only require a single crowd step, those of both editing procedures require one or two additional crowd steps, resulting in higher total costs: In a case where n_A stands for the number of acquired polygons in the acquisition step (A), the same number of crowd interactions is required for the first editing step (E_1), resulting in a total number of $n_{E1} = n_A + n_A$ interactions, and thereby leading to twice the cost, since every task is paid the same. If the second editing step (E_2) is considered, even more crowd interactions are necessary, leading to thrice the cost if compared the acquisition step (A).

Therefore, the integration procedures for all three steps are performed not only for all available acquisitions but also by varying the number of polygons from 1 to *n*: If a total of *n* polygons is available per image section, then the integration was performed for all possible sample sizes, i.e., ranging from one polygon up to *n* polygons per section. This allows for a direct comparison of integrated results that require the same amount of crowd acquisitions and subsequently the same cost: If we consider $n_A = 6$, then six polygons are acquired through the acquisition step (A), leading to a total cost of $C_{A,total} = 6$. To reach the same total cost for step E_1, however, we need to set $C_{E1,total} = C_A + C_{E1} = 6$. Since same sample size is required, and therefore $C_A = C_{E1}$, we can calculate $C_{E1} = 3$. Similarly, $C_{E2} = 2$ also applies. From this we can derive sample sizes that lead to the same total cost, i.e., $6\ n_A = 3\ n_{E1} = 2\ n_{E2}$. This relationship enables a direct and meaningful comparison in regard of cost.

Even if not primarily driven by a cost-oriented perspective, a direct comparison of the same sample sizes, i.e., $C_A = C_{E1} = C_{E2}$, can be valuable in scenarios where the focus is on output quality, with total cost being a secondary consideration.

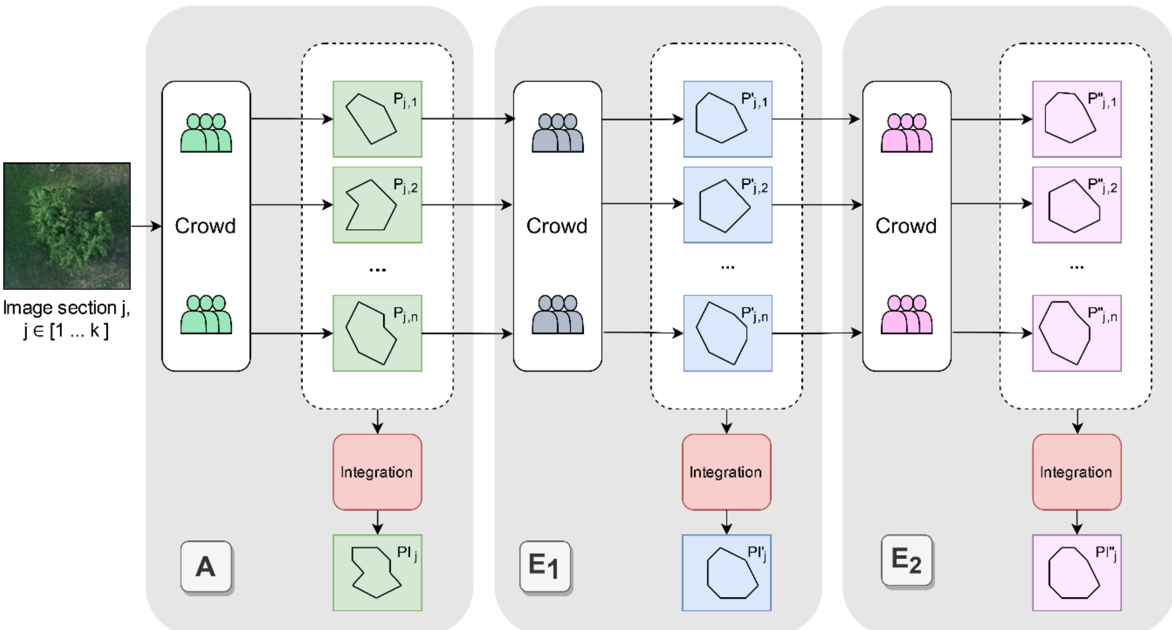

Figure 1. Methodology of the double editing process: Image sections are used for polygon acquisition (A), which are subsequently edited (E_1), followed by another round of editing (E_2). Integration is performed for all steps.

3. Dataset

We apply the presented methodology to a straightforward use case: The acquisition of tree outlines geometries in orthophotos, aiming for high-quality polygons that can later be used for the training of segmentation architectures. We utilize the same dataset as (Collmar et al., 2023a): The imagery was captured by a DJI FC6310R camera in combination with a DJI Phantom 4 RTK, and processed to a large-area orthomosaic of orchards. $k = 115$ image sections were extracted, each containing a single tree. The ground truth data for the quality evaluation in later parts of this work was collected by experts.

4. Webtools

All interactions with crowdworkers were conducted using web-based tools. Since the methodology includes two different principles, that is acquisition of geometries (step A) and editing of previously acquired geometries (steps E_1 and E_2), two different tools were implemented. Firstly, a tool that allows the acquisition of polygons (referred to as *polygon acquisition tool*), and secondly, an editor for already acquired polygons (referred to as *polygon editing tool*).

The graphical user interface (GUI) of the polygon acquisition tool is shown in Figure 2a, and the interface for the polygon editing tool in Figure 2b, respectively. In these figures, both GUIs contain letters to mark different sections: Located in the center of the interface is the data view, as is indicated by **A₁** and **A₂**. For the acquisition tool, workers are interacting with the provided imagery by clicking and therefore setting polygon vertices. In the editing tool, workers can move vertices via a drag and drop functionality. Additional functionalities are included via control buttons, i.e., undo and delete in the acquisition tool (**B₁**). For the editing tool, which works via highlighting vertices, functionalities like deleting are possible directly in the data view (**B₂**). Furthermore, letters **C₁** and **C₂** indicate the navigation through the respective webtool, since detailed instructions, preparatory task and a survey are also included in the acquisition process.

For the editing process, the symbol size of all vertices was increased, making it possible to click on individual vertices for selecting. This simplifies the before mentioned drag-and-drop functionality, allowing to intuitively relocate single vertices, as is demonstrated in Figure 3. Additionally, it is not only possible to select and move existing vertices, but also to delete vertices or to even add new ones.

(a) (b)

Figure 3. Relocation of a vertex in the polygon editing tool. (a) Position before relocation. (b) Position after relocation.

5. Campaigns

For a majority voting-based polygon integration, a larger sample size, i.e., a larger number of redundantly acquired polygons, typically leads to higher quality of the integration output up to a saturation point (Collmar et al., 2023a). Since we want to consider the factor of cost control, we aim for rather small sample sizes. Still, in order to make our results comparable to results of larger sample sizes, we decided to acquire each of the $k = 115$ image sections of the dataset by 20 different crowdworkers through our polygon acquisition tool, as per step A of the previously motivated methodology, resulting in a sample size of $n_A = 20$. Subsequently, the same sample size was utilized for steps E_1 and E_2, resulting in $n_A = n_{E1} = n_{E2} = 20$. This leads to $k \cdot n \cdot 3$ (steps) = 6,900 polygons in total. Each worker was paid \$0.15 for the processing of 5 datasets, resulting in total costs of \$207.

(a) (b)

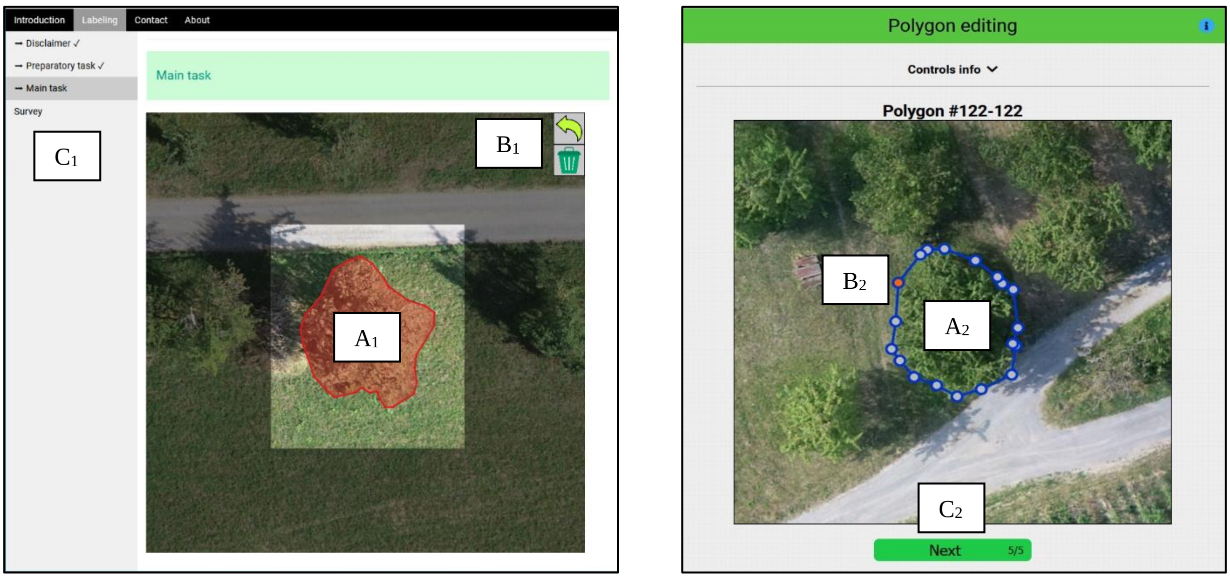

Figure 2. GUIs of the used webtools for crowd interactions. (a) Polygon acquisition tool. (b) Polygon editing tool.

6. Results

6.1 Polygon acquisition step (A)

The results of the acquisition step (A) are illustrated in Figure 4, which shows one of the image sections with $n_A = 20$ polygons plotted in yellow. The reference polygon is shown in red.

Figure 4. Image section with 20 geometries acquired through the polygon acquisition tool (yellow) and reference data (red).

It can be seen that only very few poor acquisitions were submitted, whereas most of the worker delivered results of good or acceptable quality. A potential reason for these rather positive results may lie in the previously addressed inclusion of detailed instructions and the preparatory task, effectively leading to “quality control on task designing” (Allahbakhsh et al., 2013). This inherently leads to the filtering of low-performing workers, as explained in (Jin et al., 2020). Overall, an average intersection over union (IoU) in comparison to the reference data of 0.692 was achieved for all 2,300 polygons, with approximately one third of polygons (756 or 32.87%) having an IoU value of 0.8 or larger. Only few polygons (335 or 14.57%) had an IoU value below 0.5, indicating acquisitions of low quality. Subsequently, all polygons for each of the 115 image sections were integrated using an integration threshold of 50% of the sample size n, as per (Collmar et al., 2023a), leading to 115 integrated polygons in total. The integration threshold of 50% indicates that a pixel is considered part of the integrated shape if it appears in at least half of all annotations, effectively resulting in a simple majority vote.

6.2 First editing step (E_1)

After the processing of the acquisition step (A), the unedited polygons were used as input for the first editing step (E_1), and also processed by a single crowdworker each, leading to 2,300 polygons that were edited once. The integration procedure as described in the previous step (A) was applied in similar fashion here, resulting again in 115 integrated polygons. In order to allow for a cost-oriented comparison, where $C_{A,total} = C_{E1,total}$, i.e., same total cost, the sample sizes need to be adjusted. Following the previous argumentation of Section 2, the ratio in sample sizes was set as follows: $2\, n_A = n_{E1}$. This means, in other words and for the example of $n_{E1} = 10$, that 20 unedited polygons from step (A) were integrated per image section, and their results compared to the integrated result of 10 edited polygons per image section. Edited polygons in that case refers to those that were processed in first editing step (E_1). IoU values were calculated for all integrated polygons and all different sample sizes, leading to $k = 115$ IoU values per sample size. Figure 5 presents the results in the form of boxplots: Box limits indicate upper and lower quartiles, the horizontal bar within visualizes the median. Dots represent outliers, which are calculated by 1.5 of the interquartile range, and whiskers show the non-outlier maximum and minimum, respectively.

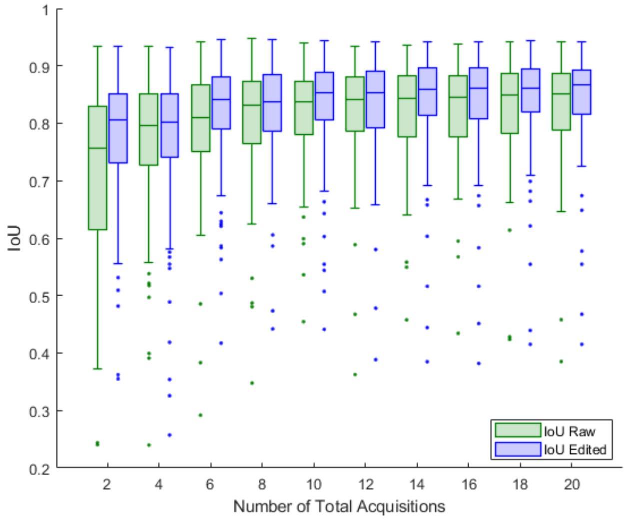

Figure 5. Boxplots of IoU values for steps (A) and (E_1) in comparison, adjusted for same cost.

As can be seen from Figure 5, the results of the first editing step lead to strictly better values not only in median, but also in box- and whisker limits, leading to a more compact distribution. Interestingly, a total of two acquisitions already lead to a large improvement. Two total acquisitions are achieved by a sample size of $n_A = 2$ for the acquisition step (A), meaning two polygons per image section are acquired by separate workers. For the first editing step (E_1), a sample size of $n_{E1} = 1$ means that a single polygon per image section is edited once by another worker, leading to two total acquisitions. While one could argue that the improvement in quality for such a low number of total acquisitions might be an outlier, the results for higher n_e show the systematic improvement: The edited results not only consistently outperform the unedited results for the same number of total acquisitions, but even for unedited results of a larger number: A total number of six edited acquisitions appears to surpass the unedited results even of a much higher number in total acquisitions.

Figure 6 further illustrates this point. Comparing the edited results for two acquisitions per image section (that is, as mentioned, a single acquired and edited polygon) to the unedited results in terms of median, as is done in Figure 6a, shows their gap in quality: Results for four unedited polygons are outperformed, although only a half of the number of total acquisitions is needed, resulting in reduced costs by factor two. Increasing the number of total acquisitions to six for the edited results confirms these findings, as is illustrated in Figure 6b, where the results of 11 unedited polygons are outperformed. Increasing the number of total acquisitions to 10 for the edited polygons even surpasses the results if all 20 available unedited polygons are used, both in terms of median values and compactness of the data distribution.

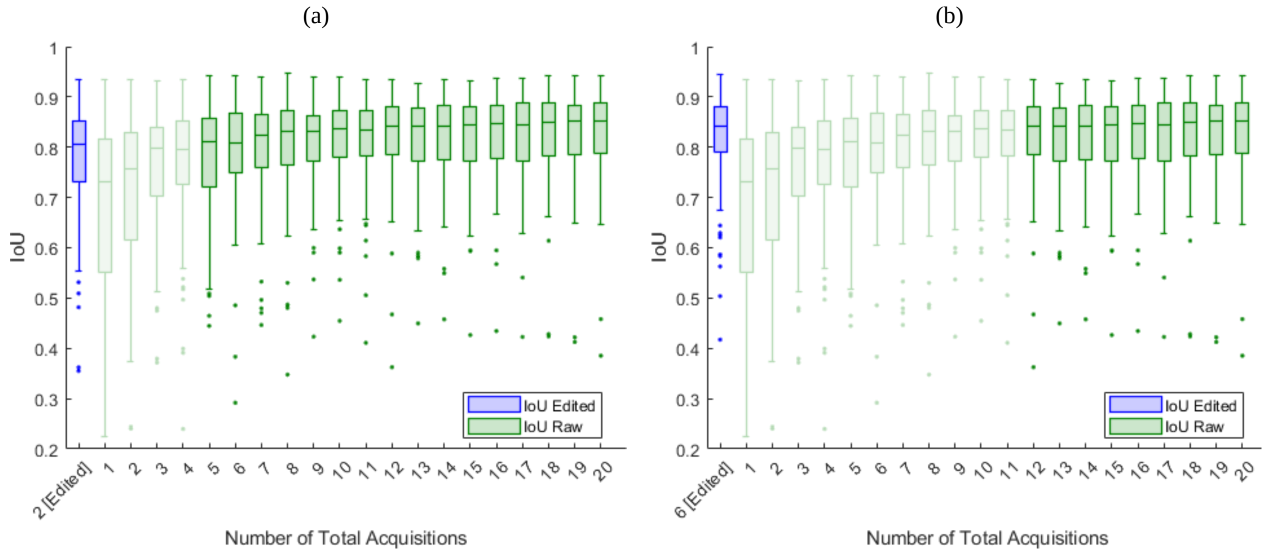

Figure 6. Comparison of boxplots for IoU values of integrated polygons after the editing step for selected *n* to those before. (a) For two polygons per section (one unedited, one edited). (b) For six polygons per section (three unedited, three edited). Faded boxplots indicate median IoU values lower than the one compared to.

A potential explanation could be formulated as follows: Initially, let’s consider a quote of $q = 15\%$ of crowdworkers that yield low-quality results. This number is based on the observation of the first acquisition process, as was described in the beginning of Section 6 on the previous page. While this assumption cannot be generalized, it serves as a preliminary basis for explanation, pending verification through later research. The assumed quote q, while appearing very specific, is interchangeable with any number less than 50%, leading to analogous conclusions, i.e., a chance of $(1-q)^2$ for poor acquisitions.

However, assuming that the number of 15% of crowdworkers delivering low-quality results can be generalized for all crowd interactions, including the second step, the data enhancement through polygon editing, this leads to the following calculations: If all crowd assignments are random, then 15% of all previously acquired data are assigned to low-quality crowdworkers during the editing step. Therefore, a mere 2.25% of acquisitions would be handled by low-quality workers in both steps. Conversely, a substantial 72.25% of crowd acquisitions would be processed by exclusively good workers in both phases. The remaining 25.5% of acquisitions have potential for lots of discussion: These are processed by both a low- and high-quality worker, in any order. If a low-quality worker precedes a high-quality one, it is plausible that the latter significantly improves the low-quality acquisition that was submitted in the first step, effectively cancelling out the low-quality submission and improving overall data quality. In the other case, where a high-quality worker precedes a low-quality worker, it is likely that the latter’s engagement with the editing tool is minimal, leaving the initially high-quality acquisition mostly unchanged, especially since it takes more time and effort to compromise an already good acquisition than to just apply minimal changes.

To summarize, one could infer that the impact of low-quality workers is substantially mitigated when their work is followed or preceded by a worker of high-quality. The observations in comparison to the unedited polygons appear to support these claims. However, it is imperative to underscore that these are assumptions, and further empirical research is needed in order to prove or disprove these hypotheses.

6.3 Second editing step (E_2)

As was shown in the previous section, performing a single editing step allows for a smaller sample size while leading to results of similar or higher quality. This raises the question if the edited results can be further enhanced by applying the secondary editing step. As per the argumentation in section 2, we only consider multiples of six for the number of total acquisitions. This allows us to align the results for all three methods, i.e., acquisition step (A), first editing step (E_1), and second editing step (E_2), enabling a direct comparison between these. The following Figure 7 visualizes and compares this alignment.

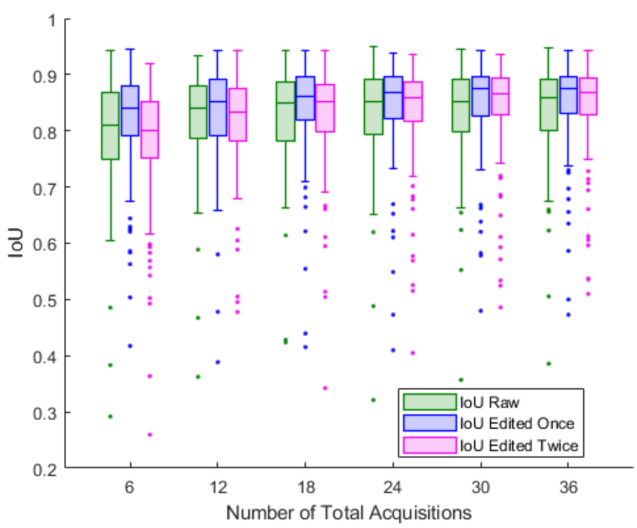

Figure 7. Boxplots in comparison for all three steps (A, E_1, E_2).

As Figure 7 shows, editing the same acquisition twice appears to not be feasible: Whereas the second editing step (E_2) leads to mostly similar but slightly worse results for a larger number of total acquisitions than the first editing step (E_1), the results of the second editing step (E_2) for smaller numbers are clearly outperformed by those of the first editing step (E_1). Since not even the unedited acquisitions from step (A) can be outperformed for a small number of acquisitions, it can be concluded that the second editing step does not yield any benefit in the considered case.

The reason behind this might become obvious when looking at the low number of acquisitions, for example for the value 12. Here, 12 unedited polygons per section are integrated and their quality represented through the boxplot in Figure 8. If a single editing step is applied, then 6 unedited polygons are edited and optimized once each, leading to a total number of 12 acquisitions, whereas this leads to only 6 polygons being integrated. If a second editing campaign is conducted, this number decreases further: In order to reach a number of 12 acquisitions in total, only 4 polygons can be enhanced through the two editing steps, therefore leading to a mere number of 4 polygons as input for the integration. Since it was shown in previous studies that the integration performs better for larger input sample sizes (Collmar et al., 2023a), it appears that the double editing process leads to sample sizes too small in relation to their respective costs. Also, no improvement is observable even for larger sample sizes, making this approach inferior in terms of quality for the considered, cost-oriented scenario.

Given a quality-oriented case instead of a cost-oriented case, where total cost is not a relevant factor, a direct comparison between all three steps with equal sample size and therefore equal cost, i.e., $C_A = C_{E1} = C_{E2}$, might be of interest. Figure 8 shows this case, and is not adjusted for the number of total acquisitions.

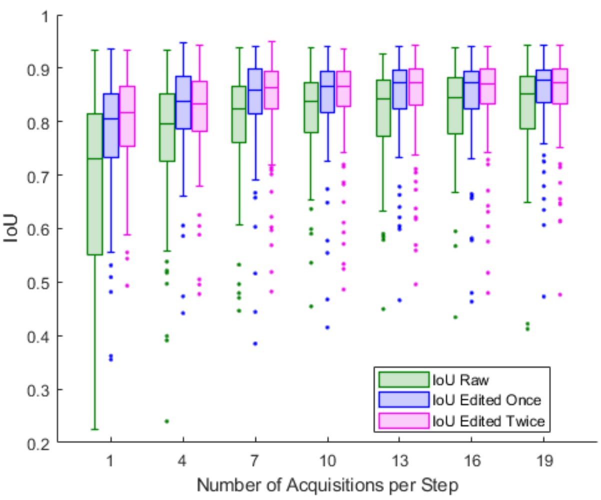

Figure 8. Boxplots in comparison for all three steps, not adjusted for number of total acquisitions.

Interestingly, Figure 8 only shows a clear improvement of quality for the first data point, which is a sample size of one for all steps: One single polygon per section was acquired in the acquisition step (A), and subsequently edited in the first step (E_1), followed by a second editing (E_2). For all other number of acquisitions that are shown in Figure 8, however, the results of the second editing step (E_2) appear to be approximately the same as those of the first editing step (E_1), only achieving slightly compacter distributions as is indicated by the whiskers. This observation is consistent with the argumentation of section 6.2, which argued that potentially only a low number of acquisitions are of obviously poor quality after a single editing step (e.g., 2.25%). Therefore, the potential for improvement by a second editing step is rather small, since only these 2.25% offer big potential for further improvement in quality. Overall, the analyzed data suggest that implementing a second editing step does not offer substantial benefits: Coming with higher cost in both time and money, as well as mostly similar results as the first editing step (E_1) in terms of quality, we deem the second editing step (E_2) not feasible for the highlighted cases.

7. Further investigations

7.1 Larger sample sizes and filtering

Another way for the enhancement of post-integration results for polygons is performing a filtering process before the integration, as was shown in (Collmar et al., 2023). This naturally leads to an important follow-up question: which approach is more effective; editing or filtering? Also, it was previously shown that results of the first editing step (E_1) outperform even larger sample sizes. This raises the question to what extent polygon editing can outperform the results of larger sample sizes.

To address both questions, we first raised the sample size for the polygon acquisition step (A) to 100 polygons per image section and then applied the same filtering as is described in (Collmar et al., 2023). Since the filtering process is performed before the integration, poor acquisitions are effectively eliminated, thereby enhancing the quality of input for the integration, making the process directly comparable to our editing pipeline. We used the filtering approach that led to best results in the previously referenced work, i.e., a combined moment filtering for a p value of 75%. This was applied to all acquired data, including those of large sample size. Following the filtering process, all remaining polygons were integrated for various sample sizes.

Figure 9 shows boxplots, comparing the IoU values after integration of the unedited results from the acquisition step (A), the results after the first editing step (E_1) as well as the filtered acquisitions. Again, the comparison is done on a cost-oriented base, ensuring that the total acquisitions and thereby total cost are equal. What can be seen from Figure 9 is that filtering before integration indeed leads to results of improved quality as is measured by the mean IoU values, independently of the number of acquisitions or sample size, respectively. Still, not only the IoU values are improved, but also the compactness of the data distribution, as is indicated by the box limits and whiskers.

However, it's crucial to note that the IoU values after editing consistently surpass those achieved through the filtering process. Not only is this the case for a total number of acquisitions from 10 to 40: A number of 20 unedited polygons, each edited a single time in the first editing step (E_1) and thereby resulting in 40 total acquisitions, outperform not only the results of 40 filtered acquisitions, but even score slightly higher median IoU values than even 100 filtered acquisitions. To illustrate this fact, a blue dashed line was included in Figure 9, indicating the median value of the largest available sample size after editing, i.e., 40.

While this is already interesting from a quality point of view, it is much more impressive from a cost-optimization perspective: Data filtering leads to better results than using raw data, however, still a rather large amount of crowd acquisitions is necessary to achieve high IoU values. On the other hand, this is not the case when applying a first editing step (E_1): 40 total acquisitions can outperform more than double their sample size in terms of median IoU values. Also, not only median values appear to be superior, but also the data compactness that is indicated by the whiskers in Figure 9.

https://doi.org/10.5194/isprs-annals-X-4-2024-83-2024 |

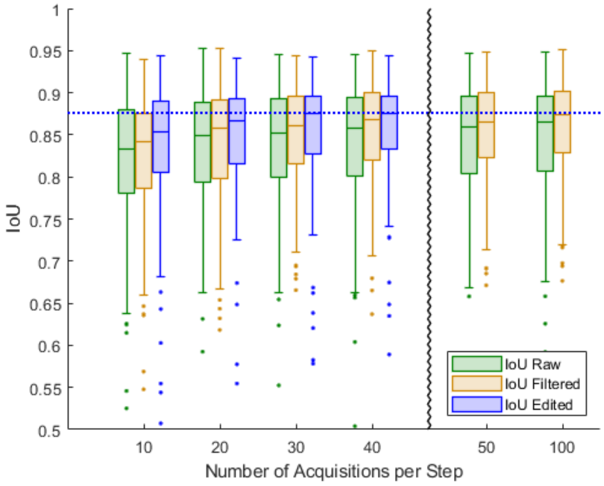

Figure 9. Boxplots in comparison for unedited results ("raw"), filtered unedited results, and edited results. The blue dashed line indicates the median value of the edited results for the largest sample size available, i.e., 40.

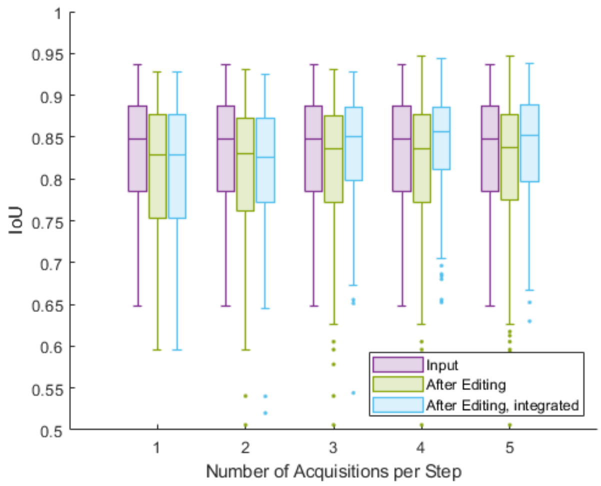

Figure 10. Boxplots in comparison for IoU values, if an editing step (E*) is performed after integration.

In conclusion, applying an editing step like it was done in (E_1) outperforms larger sample sizes not only when unedited polygons are used for the integration, but also after filtering was applied for input data, therefore allowing to cut acquisition numbers and subsequently necessary costs significantly. Furthermore, filtering might not be applicable for very low numbers of polygons (e.g., 2), whereas polygon editing still delivered impressive results. On the other hand, since filtering does not require an editing step, a second crowd campaign as well as the development of an editing platform, such as the polygon editing tool, is not necessary, making it a faster and easier alternative for cases where the optimization of cost and quality is not necessary.

7.2 Editing of integrated results

It has been established that a single editing step (E_1) can be used to enhance both cost-efficiency and data quality. However, in previous sections, this approach has only been applied to polygons before their integration. This leads to the question, if similar improvements can also be achieved if the crowd enhancement is applied after the integration. To address this question, we utilized the integrated results of the unedited polygons from the acquisition step (A) with the respective sample size of $n_A = 20$, consisting of 115 integrated polygons (i.e., one integrated polygon per image section) of high geometric quality. These polygons were then presented to crowdworkers via the polygon editing tool in a new editing campaign, referred to E*. For this new editing step (E*), we picked a small sample size in order to keep costs to a minimum, i.e., $n_{E*} = 5$.

In the previously observed cases, polygons acquired by crowdworkers were edited, mostly consisting of a rather small number of vertices. Since we used integrated polygons as input for the editing tool in this new editing step (E*), and given the fact that integrated polygons typically have a complex geometry and a large number of vertices (Collmar et al., 2023b), a reduction of vertices is needed for effective processing using the polygon editing tool. To achieve this, we employed the Douglas-Peucker algorithm for line smoothing (Douglas and Peucker, 1973). This led to only minimal deviations in IoU values (approx. 0.2% or 0.0002), while substantially reducing the number of vertices. After the smoothing process, the resulting polygons were then processed in the polygon editing tool. Subsequently, IoU values to the reference were calculated both before and after integration for all sample sizes $n_{E*} = [1…5]$. Their respective results are visualized in Figure 10, again in the form of boxplots, with the input data for comparison.

As can be seen from Figure 10, an editing step after the integration led to strictly inferior results independently of the number of acquisitions per step (i.e., sample size) in mean, median and standard deviation values, if no integration step is performed. It can thereby be concluded, that for the case where no integration is performed, editing integrated polygons does not yield a beneficial effect in regards of quality. This is not surprising if the nature of the input data is considered: Integrated polygons already include a large number of details (Collmar et al., 2023b), where an editing step following the integration could lead to a loss of details, especially, if no integration is performed.

Therefore, another integration process was applied after the new editing step (E*), leading to obviously better results as can be seen in Figure 10. Whereas the integrated results after the new editing step (E*) appear to be of worse quality, those of a higher number of acquisitions seem to be of similar quality to the input data.

While it is possible to achieve similar or marginally better IoU results, adopting this new workflow introduces significant complexity: Starting with data acquisition in step (A), followed by integration, as well as editing of the integrated results as per step (E*), and then performing another round of integration. This not only complicates the processing pipeline but also extends its duration, without yielding substantially improved outcomes. Furthermore, one output polygon in this new editing step (E*) requires already integrated polygons as input, which are calculated from the results of the acquisition step (A). These prerequisites appear to be the same as for the single editing step of the main methodology (E_1). However, the integration process with a subsequent polygon smoothing is required in between, effectively leading to a more complex version of the first editing step (E_1) and resulting in no better quality for higher complexity. This hinders the automation of the whole process, making the new approach not interesting for the considered case.

8. Conclusions

We adopted a new methodology where polygons are acquired through crowdsourcing in an acquisition step (A) and then presented to another set of crowdworkers in a polygon editor. This tool allows crowdworkers to edit and improve the previously acquired polygon geometries, resulting in the first crowdsourced editing step (E1). The editing step was repeated, presenting the already processed results of the first editing step (E1) to yet another set of crowdworkers, resulting in the second editing step (E2). We integrated all acquired and edited polygons after every step for a comparative qualitative analysis.

We demonstrated that the quality of the integrated data can be notably improved when applying the first editing step (E1). The results of this initial editing step were of such high quality that they surpassed the performance of even much larger samples of the acquisition step (A), which did not involve any editing. Conclusively, this allows for the same or even higher data quality when using considerably smaller sample sizes. Despite the extra costs for the editing step, this allows for a significant reduction of overall acquisition costs, making the approach attractive for any sample sizes. Additionally, robust results were achieved even for very small sample sizes.

Further investigations were conducted, highlighting the advantages of our presented methodology in comparison to the already existing method of polygon filtering, as well as showing its effectiveness even for large sample sizes, allowing for further cost-saving opportunities.

Conducting a second editing step (E_2), as well as processing already integrated polygons (E*), did not yield significant benefits. However, more research might be necessary to confirm our findings.

For future work we plan to perform a continued validation with different use cases and evaluation metrics. This could help to further demonstrate the efficiency of our proposed methodology.

Acknowledgements

Partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2120/1 – 390831618.

References

Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H. R., Bertino, E., Dustdar, S., 2013: Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, *17*(2), 76-81.

Chandler, J., Paolacci, G., Mueller, P., 2013: Risks and rewards of crowdsourcing marketplaces. *Handbook of human computation*, Springer Verlag, New York, 377-392.

Collmar, D., Walter, V., Kölle, M., Sörgel, U., 2023a: From Multiple Polygons to Single Geometry: Optimization of Polygon Integration for Crowdsourced Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *10*, 159-166.

Collmar, D., Walter, V., Kölle, M., Sörgel, U., 2023b: A Two-Step Approach for the Acquisition of Individual Tree Outlines using Paid Crowdsourcing. In *43. Wissenschaftlich-Technische Jahrestagung der DGPF, 22.-23. März 2023 in München*, 163-173.

Douglas, D. H., Peucker, T. K., 1973: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, *10*(2), 112-122.

Hirth, M., Hoßfeld, T., Tran-Gia, P., 2011: Anatomy of a crowdsourcing platform - using the example of microworkers.com. *2011 Fifth international conference on innovative mobile and internet services in ubiquitous computing*, 322-329.

Hossain, M., 2012: Users' motivation to participate in online crowdsourcing platforms. *2012 International Conference on Innovation Management and Technology Research*, 310-315.

Howe, J., 2006: The rise of crowdsourcing. *Wired magazine*, 14(6), 1-4.

Jin, Y., Carman, M., Zhu, Y., Xiang, Y., 2020: A technical survey on statistical modelling and design methods for crowdsourcing quality control. *Artificial Intelligence*, 287, 103351.

Li, C., Sheng, V. S., Jiang, L., Li, H., 2016: Noise filtering to improve data and model quality for crowdsourcing. *Knowledge-Based Systems*, 107, 96-103.

Li, G., Zheng, Y., Fan, J., Wang, J., Cheng, R., 2017: Crowdsourced data management: Overview and challenges. In *Proceedings of the 2017 ACM International Conference on Management of Data*, 1711-1716.

Mason, W., Watts, D. J., 2009: Financial incentives and the performance of crowds. In *Proceedings of the ACM SIGKDD workshop on human computation*, 77-85.

Rea, S. C., Kleeman, H., Zhu, Q., Gilbert, B., Yue, C., 2020: Crowdsourcing as a tool for research: Methodological, fair, and political considerations. *Bulletin of Science, Technology & Society*, 40(3-4), 40-53.

Saralioglu, E., Gungor, O., 2020: Crowdsourcing in remote sensing: A review of applications and future directions. *IEEE Geoscience and Remote Sensing Magazine*, 8(4), 89-110.

Stonebraker, M., Rezig, E., 2019: Machine learning and big data: What is important?. *IEEE Data Engineering Bulletin*, 42.4, 3-7.

Tu, J., Yu, G., Domeniconi, C., Wang, J., Xiao, G., Guo, M., 2018: Multi-label answer aggregation based on joint matrix factorization. In *2018 IEEE International Conference on Data Mining (ICDM)*, 517-526.

Zhang, J., Wu, X., Sheng, V. S., 2016: Learning from crowdsourced labeled data: a survey. *Artificial Intelligence Review*, 46(4), 543-576.

Zheng, Y., Li, G., Li, Y., Shan, C., Cheng, R., 2017: Truth inference in crowdsourcing: Is the problem solved?. *Proceedings of the VLDB Endowment*, *10*(5), 541-552.