Use of paid Crowdsourcing for the Collection of Geodata

Volker Walter, Dominik Laupheimer, Dieter Fritsch

Institute for Photogrammetry
University of Stuttgart
Geschwister-Scholl-Str. 24D
70174 Stuttgart
Contact: Volker.Walter@ifp.uni-stuttgart.de

Extended Abstract

Introduction

Crowdsourcing is a new technology and a new business model that will change the way in which we work in many fields in the future. Employers divide and source out their work to a huge number of anonymous workers on the Internet. The division and outsourcing is not a trivial process but requires the definition of complete new workflows – from the definition of subtasks, to the execution and quality control. A popular crowdsourcing project in the field of collection of geodata is OpenStreetMap, which is based on the work of unpaid volunteers. Crowdsourcing projects that are based on the work of unpaid volunteers need an active community whose members are convinced about the importance of the project and who have fun to collaborate. This can only be realized for some tasks. In the field of geodata collection many other tasks exist which can in principle be solved with crowdsourcing but where it is difficult to find a sufficient large number of volunteers. Other incentives must be provided in these cases, which can be monetary payments.

ICA European Symposium on CARTOGRAPHY
VIENNA, 10–12 NOVEMBER 2015

Published in "Proceedings of the 1st ICA European Symposium on Cartography", edited by Georg Gartner and Haosheng Huang, EuroCarto 2015, 10-12 November 2015, Vienna, Austria

Project

The majority of projects in the field of crowd-based geodata collection are based on the work of volunteers (VGI – Volunteered Geographic Information). In contrast to these kind of approaches, the project described in this paper has been realized with paid crowdsourcing, which means that the crowd-workers are paid for their work. We want to identify and quantify the parameters (e.g. amount of salary, size of working tiles or object types) which influence the quality of the results (especially: correctness, geometric accuracy and collection time).

We developed a web-based program for the collection of geodata and integrated it into the commercial crowdsourcing platform microWorkers (www.microWorkers.com), which takes over the recruitment and the payment. The platform has access to more than 600.000 registered crowd-workers. The workers are informed automatically when a new job is offered by an employer on the platform. The employers can restrict the jobs to specific groups of workers. For example, it is possible to offer the jobs only to workers that are living in a specific country or to workers that have already successfully worked on a particular number of other jobs. Further qualifications are possible with own developed tests, which must be solved before the job. After the job has been completed, the results are submitted to the employer who checks the quality of the results. The final payment is handled by the crowdsourcing platform.

Graphical User Interface

Figure 1 shows the Graphical User Interface of the program. The program was developed with JavaScript. The central element is an orthophoto with the size of 500 * 450m² from which the data has to be collected. The object classes that have to be collected are: forests (with polygons), streets (with lines) and buildings (with points). Wrong collected objects can be deleted. The workers have the possibility to make additional comments on their work and submit them together with the collected data.

First results

A RGB orthophoto with a ground sampling distance (GSD) of 1m and a size of approximately 5 * 4km² was subdivided into 88 patches with the size of 500 * 450m². Several campaigns with different parameters were launched on the microWorkers platform to evaluate the quality of the crowd-based data collection. The results are very promising and the quality of the data was even outperforming our expectations in many cases. However, the tests

showed also that the quality of the collected data varies significantly. Some crowd-workers collected the data with very high quality (see *Figure 2a*) whereas other crowd-workers collected completely wrong data (see *Figure 2b*). Interestingly, we observed that there is no direct connection between the amount of salary and the quality of the results. An increase of the salaries did not lead to a better quality but only to a faster completion of the campaigns.

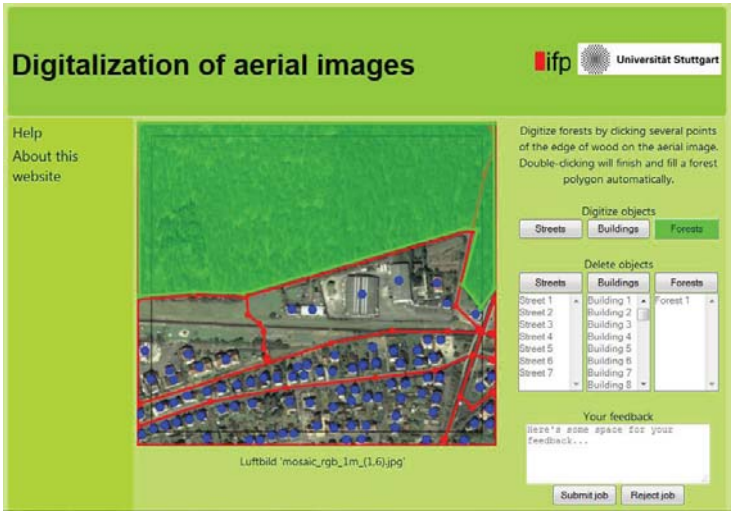
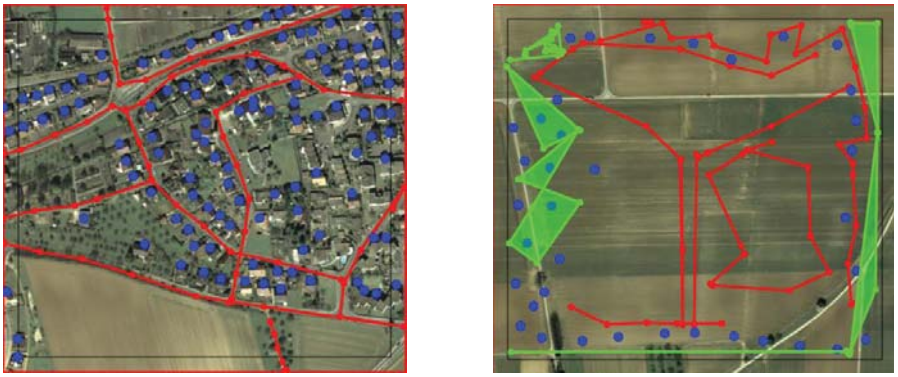


Figure 1. Graphical User Interface of the program for crowd-based data collection.



a b

Figure 2. Two examples of crowd-based spatial data collection: a) data with high quality b) useless data

Outlook

The test showed that in principle it is possible to produce high quality spatial datasets with paid crowdsourcing. The main problem is that the quality of the data is extremely heterogeneous. Therefore, it is necessary to find control mechanisms that evaluate the quality of the data. This either must be done automatically or again sourced out to the crowd. Furthermore, selection procedures are needed, which can automatically select crowdworkers who collect data with high quality. This can be realized for example with user profiles. Finally, algorithms are needed which integrate the individual results into an overall result. Spatial inconsistent datasets, which overlap multiply, have to be integrated into a consistent, uniform dataset. All these aspects will be investigated in our ongoing research.