

Use of paid Crowdsourcing for the Collection of Geodata

Volker Walter, Dominik Laupheimer, Dieter Fritsch

Institute for Photogrammetry

Stuttgart University

70174 Stuttgart, Germany

Contact: Volker.Walter@ifp.uni-stuttgart.de

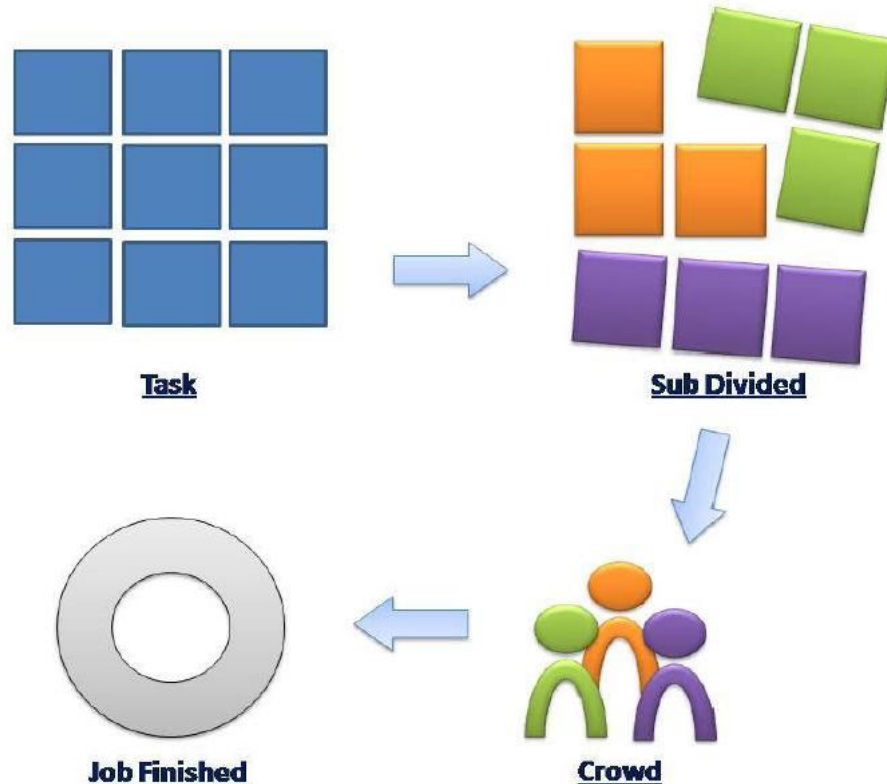
Contents



- Motivation
- Paid Crowdsourcing
- Our Approach
- First Results
- Summary and Future Work

What is Crowdsourcing?

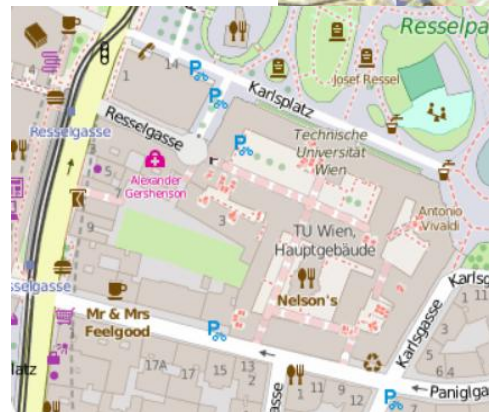
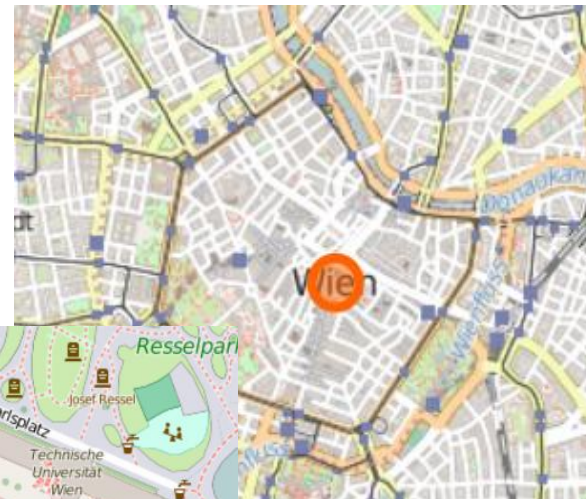
- Crowdsourcing is a technology and a business model that will change the way in which we work in many fields in the future.
- Employers divide and source out their work to a huge number of anonymous workers in the internet (the Crowd).



Crowdsourcing for the Collection of Geodata



- A popular crowdsourcing project in the field of geodata collection is OpenStreetMap



- A huge amount of unpaid volunteers collect world-wide geodata with high quality



Motivation of our Project



- 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 have fun to collaborate
- This can only be realized for some tasks. In the field of geodata collection exist many other tasks 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.
- In this project we investigate, how the collection of geodata can be realized with paid crowdsourcing.



Platforms for paid Crowdsourcing



- Crowdsourcing platforms have access to many registered crowdworkers
- Employers can publish new jobs:
 - Job description
 - Height of payment
 - Link to a web-interface where the job can be executed
- The workers are informed when a job is offered by an employer
- After the job has been completed, the results are submitted to the employer who checks the quality of the results.
- The payment (if job is accomplished successful) is handled by the crowdsourcing platform.
- **-> Crowdsourcing platforms are responsible for the recruitment, management and payment of the workers**



Platforms for paid Crowdsourcing



- Amazon Mechanical Turk
 - One of the first Crowdsourcing platforms
 - 500,000 registered workers
 - Employers must have an postal address in USA



- microWorkers
 - 500,000 registered workers
 - Used in this project



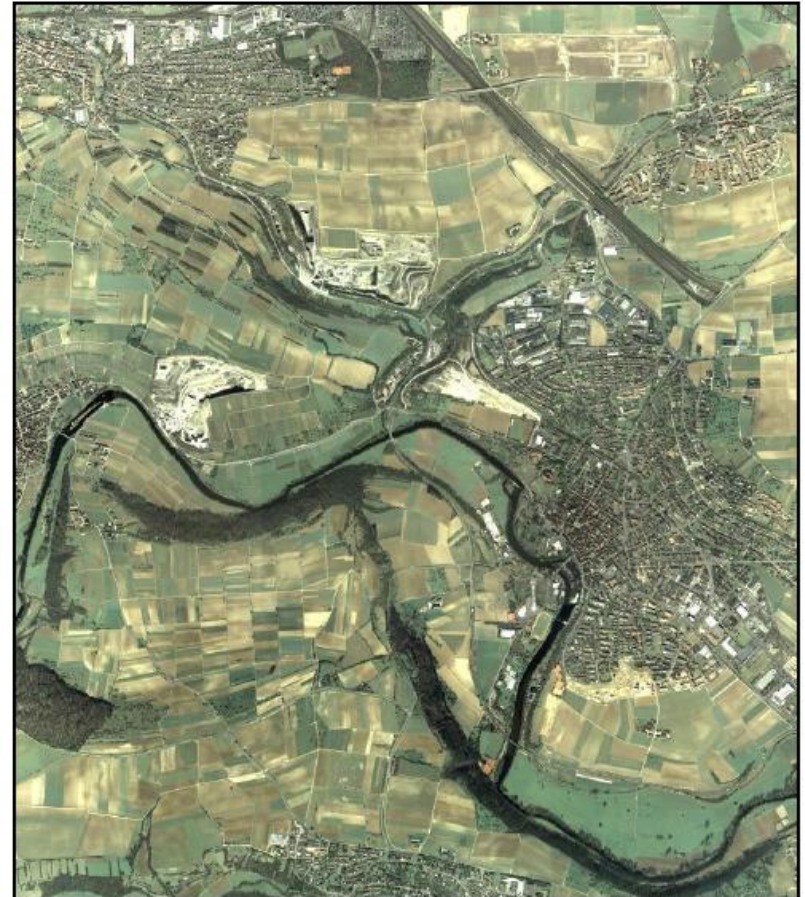
- Other platforms: ShortTask, ClickWorker, Minute Worker, RapidWorker, CloudCrowd, CrowdFlower, ...



Data



- A RGB orthophoto with a ground sampling distance (GSD) of 1m and a size of approximately $5 * 4\text{km}^2$ was subdivided into 88 patches with the size of $500 * 450\text{m}^2$
- Objects that had to be collected:
 - Forests (Polygons)
 - Streets (Lines)
 - Houses (Points)



Our Approach



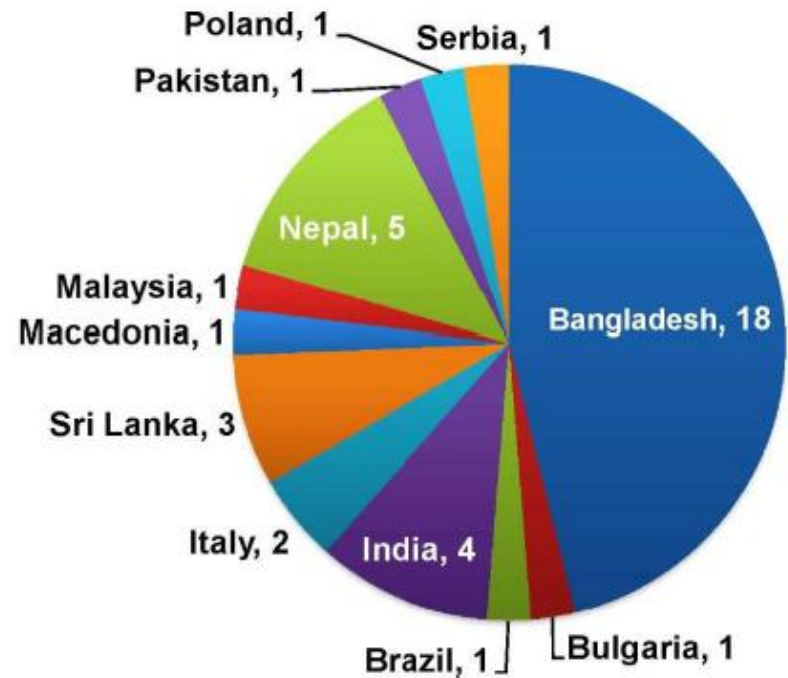
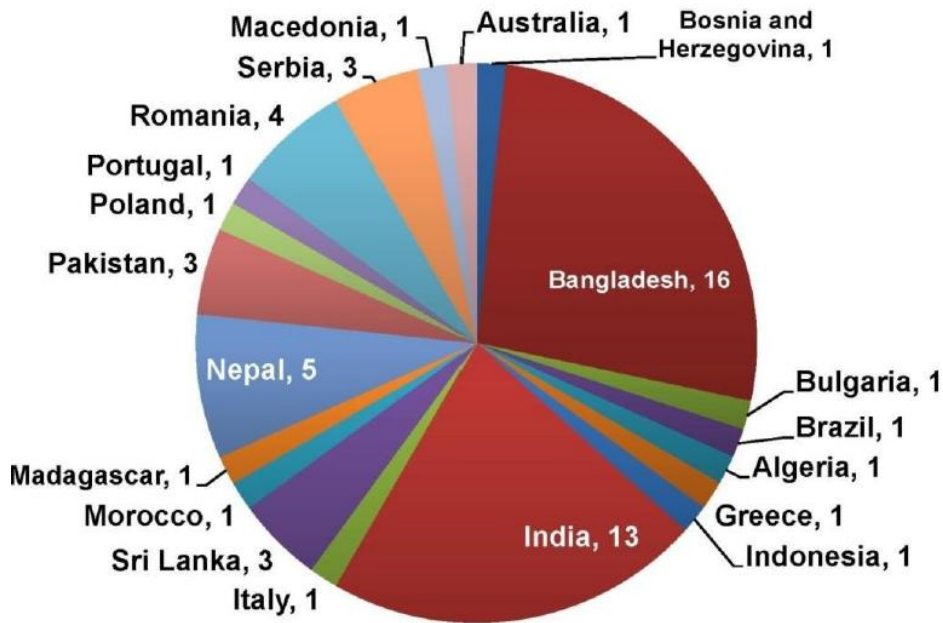
- Web-based data collection program:



- Six campaigns with different parameters were launched on the microWorkers platform to evaluate the quality of the crowd-based data collection.



Spatial Distribution of the Crowdworker (Campaign 1 and 2)



Data Examples: Category 1 (Very Good)



- Some houses are missing in the middle of the image, but most of the area is accurately digitized



Data Examples: Category 2 (Good)



- Some houses and streets are missing in the upper left area, but overall the work has high accuracy



Data Examples: Category 3 (Partly Good)



- The lower right part of the image and the forest at the bottom is not digitized. The remaining work is accomplished with accuracy



Data Examples: Category 4 (Poor)



- Most of the streets are ok. Some houses are missing – some are incorrect. The geometry of the forest is not precise



Data Examples: Category 5 (Unsatisfactory)



- The major area is not digitized



Data Examples: Category 5 (Unsatisfactory)



- Senseless Data



Data Examples: Category 5 (Unsatisfactory)



- No Data

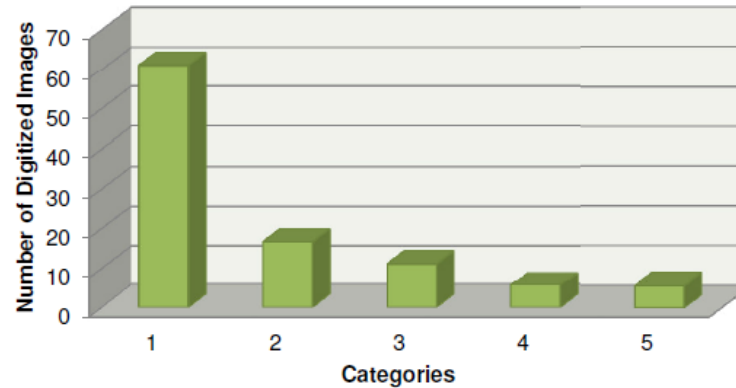


Quality of the Results



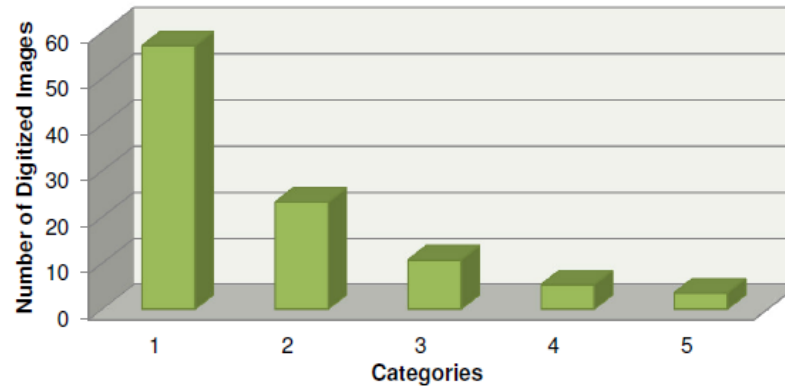
0,10 \$ per Job

Number of Digitized Images for Campaign 1



0,45 \$ per Job

Number of Digitized Images for Campaign 2



We observed that (in our tests) there is no direct connection between the amount of salary and the quality of the results.



Time to accomplish a campaign



- An increase of the salaries leads to a faster completion of the campaigns.

Campaign	Payment / job in \$	Total No. of Jobs	No. of Days
01	0.10	56	12
02	0.45	56	03



Summary and Future Work



- 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.
 - 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.
 - Selection procedures are needed, which can automatically select crowd-workers who collect data with high quality. This can be realized for example with user profiles.
- 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 future research.

