

# Crowd-sourced data collection to support automatic classification of building footprint data

Robert Hecht<sup>a</sup>, Matthias Kalla<sup>b</sup>, and Tobias Krüger<sup>a</sup>

<sup>a</sup> Leibniz Institute of Ecological Urban and Regional Development, Dresden, Germany; [r.hecht@ioer.de](mailto:r.hecht@ioer.de), [t.krueger@ioer.de](mailto:t.krueger@ioer.de)

<sup>b</sup> Faculty of Environmental Sciences, Technical University of Dresden, Germany; [matthias.kalla@mailbox.tu-dresden.de](mailto:matthias.kalla@mailbox.tu-dresden.de)

**Abstract:** Human settlements are mainly formed by buildings with their different characteristics and usage. Despite the importance of buildings for the economy and society, complete regional or even national figures of the entire building stock and its spatial distribution are still hardly available. Available digital topographic data sets created by National Mapping Agencies or mapped voluntarily through a crowd via Volunteered Geographic Information (VGI) platforms (e.g. OpenStreetMap) contain building footprint information but often lack additional information on building type, usage, age or number of floors. For this reason, predictive modeling is becoming increasingly important in this context. The capabilities of machine learning allow for the prediction of building types and other building characteristics and thus, the efficient classification and description of the entire building stock of cities and regions. However, such data-driven approaches always require a sufficient amount of ground truth (reference) information for training and validation. The collection of reference data is usually cost-intensive and time-consuming. Experiences from other disciplines have shown that crowdsourcing offers the possibility to support the process of obtaining ground truth data. Therefore, this paper presents the results of an experimental study aiming at assessing the accuracy of non-expert annotations on street view images collected from an internet crowd. The findings provide the basis for a future integration of a crowdsourcing component into the process of land use mapping, particularly the automatic building classification.

**Keywords:** Crowdsourcing, Building classification, Land Use Mapping, Sustainable Development

## 1. Introduction

Digital building models from National Mapping and Cadastral Agencies (NMCA) or Volunteered Geographic Information (VGI) platforms often lack attribute information, such as the building usage, housing type, number of floors, building height, and years of construction. However, this information is of particular importance for various research domains and applications such as spatial science, geography, urban planning, architecture, and disaster management. Supervised machine learning techniques help to classify the building footprints according to a predefined building typology and to semantically enrich the datasets with additional information. Such data-driven approaches provide promising results with high accuracies for single cities and regions (e.g. Römer and Plümer 2010; Henn et al. 2012, Hecht et al. 2015, Wurm et al. 2016). One of the main challenges is the limited transferability of the classifiers due to strong regional dependencies (Steiniger et al. 2008, Hecht et al. 2015). A trained machine learning classifier is only applicable for cities with a similar building structure and history. Changing the spatial and cultural context (e.g. other regions, countries, continents etc.) requires the collection of additional ground truth data in the specific area under investigation for model training and validating. To overcome these regional differences, an efficient strategy for ground truth data collection needs to be elaborated. In recent years, crowdsourcing has been proven suitable for collecting

training and validation data in a variety of research disciplines. In this study, we want to explore the potential of crowdsourcing in the context of mapping and monitoring urban land use, particularly the classification of building footprints in digital topographic databases.

## 2. Background and Related Work

Today citizens are becoming more and more important as a new source of geo-information. In the last few years, a number of different terms from different disciplines have emerged that describe the process of citizen-based sensing of geographic information, namely crowdsourcing, citizen science, collaborative mapping or the crowd-sourced information itself, such as Volunteered Geographic Information (VGI) or User-Generated Content (UGC). The form of data collection can be very different. According to See et al. (2016) crowdsourced geographic information can be either contributed actively as part of a crowdsourcing system/campaign (e.g. OpenStreetMap, Wikimapia) or contributed passively by mapping already existing crowdsourced data that has been collected for other purposes (e.g. mobile data, location-based social media content). Furthermore, the types of information (e.g. spatial vs. aspatial, labels vs. geometry etc.) or the forms of motivation strategies (gamification, paid crowd etc.) can vary.

In our context, we prefer using the term crowdsourcing defined as a type of participative online activity, particularly the process of a voluntary undertaking of specific tasks (Estellés-Arolas and Ladrón-de-Guevara

2012). Crowdsourcing appeared first in Howe (2006) describing the business practice of outsourcing activity to the crowd, which is today an attractive way of acquiring cheap and fast annotations from non-expert contributors over the Web that almost have the same quality as expert labels (Snow et al. 2008). The idea of using online users for the purpose to label images goes back to Luis von Ahn who designed the ESP game (von Ahn and Dabbish 2004) and further developed reCAPTCHA (von Ahn et al. 2008), a system to verify humanity and simultaneously assisting the digitization of books by solving complex OCR problems with crowdsourced labels. Today crowdsourcing is used in different research domains to collect large datasets that would otherwise not be possible using the researcher's own resources. On the other hand, it can be applied to solve computationally expensive and difficult problems. Annotations such as boxes, contours, correspondences or labels are of research interest, for example, in medical image processing (Maier-Hein et al. 2014) or autonomous driving (Donath and Kondermann 2013). In the context of land cover mapping and remote sensing, crowdsourcing is used to collect data (primarily labels) for validation and training, such as in the famous Geo-Wiki platform (Fritz et al. 2012, Laso Bayas et al. 2016). The Geo-Wiki developments go hand in hand with studies on data quality (See et al. 2013, Salk et al. 2016). In addition to the classification task (labeling), there are also conflation tasks and digitization tasks (Albuquerque et al. 2016). Hillen and Höfle (2015) have proposed a prototype implementation of a system for digitizing building footprints, namely Geo-reCAPTCHA. They adapted the reCAPTCHA idea to create geographic information via web-based micro-mapping tasks and assessed time and quality of the data. Further, in an EU project CAP4Access crowdsourcing was also tried out for the acquisition of sidewalk information that is necessary for routing and navigation services tailored to the needs of wheelchair users (Hahmann et al. 2016).

### 3. System for crowd-sourced data collection supporting building type recognition

In this section, we outline an integrated system for automatic classification of building footprints supported by crowdsourcing. The automatic classification of building footprints uses a supervised machine learning approach as described, for example, in Hecht et al. (2015). Crowd-sourced annotations on geo-coded street view images from the internet supports the training and validation of the classifier. The conceptual model for crowd-sourced collection of ground-truth data mainly consists of the following steps, also shown in Figure 1:

- 1) Definition of building types and visual characteristics
- 2) Construction and design of image annotation tasks
- 3) Collection of street view imagery
- 4) Perform image annotation
- 5) Post-processing, quality assessment and data filtering
- 6) Inference of building types based on the crowd-sourced data

At first, a target building classification scheme needs to be defined (1). This includes the definition of the building types that are subject to the subsequent classification process as well as their visual characteristics in the images intended to be used. Subsequently, the classification problem has to be decomposed into individual image annotation tasks (2). This is carried out by constructing and designing very simple tasks, which requires some a priori expert knowledge of the relevant visual characteristics for distinguishing different building types. A simple task would be the boolean query of whether a particular property is true or false. A more complex task offers more than two answer options in single-choice or multiple-choice mode. The annotation tasks can be implemented in different crowdsourcing/micro-task platforms (e.g. CrowdFlower, Amazon Mechanical Turk, Crowdcrafting) or in own applications and games (e.g. Cropland Capture, [www.cropland.geo-wiki.org](http://www.cropland.geo-wiki.org)) for different devices. Easy handling is one of the most important aspects for task designing. The task should be solvable in an acceptable amount of time. Further, limitations of the display size (desktop computer vs. smartphone) should also be considered to ensure readability.

The next step (3) is to collect the images to be annotated by the crowd. Generally, street view or bird's-eye perspective images are desired since these types of images allow for the recognition of several building properties. Potential image sources are Google Street View, Microsoft Bing Maps Bird's Eye Views, or the street-level imagery from the VGI platform Mapillary. In addition to these sources, any kind of geotagged imagery in social media (Flickr, Facebook, Instagram etc.) can be used as long as an automated access is given through a provided API by means of a spatial query. The basis for the selection of the images can be a random sample of addresses (address list), which was created in advance from a given spatial database. Once the images are collected, the image annotation can be performed (4). The task responses are usually recorded along with metadata (time, user name, country, etc.). In order to reduce noise and to allow for intrinsic quality controlling redundant labels are gathered by assigning a task to different annotators.

In a post processing step (5) the results are aggregated by majority voting. With the help of intrinsic measures, the quality of annotations can be assessed. Based on the measures bad annotations or unreliable annotators can be identified and excluded from further processing.

In a final step (6), the building types are derived based on the crowd-sourced building characteristics resulting in categorical data. This ground truth data can either be used for training and / or validation in the context of an automatic building classification.

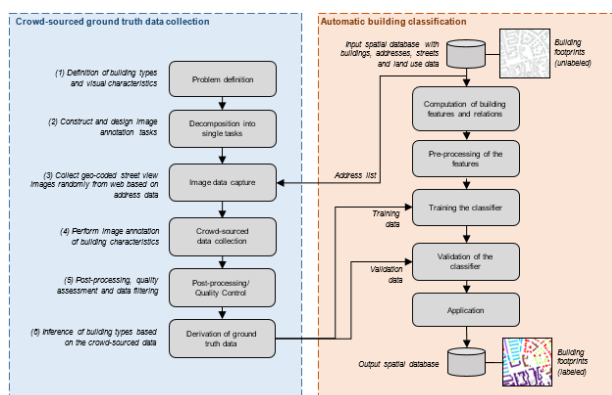


Fig. 1. Workflow of crowd-sourced data collection to support automatic building classification

#### 4. Implementation of the experimental study

The aim of the study is to validate the performance of crowdsourcing for obtaining ground truth information of specific building characteristics using street view imagery from online resources. According to the conceptual workflow, we further specify the design of the study including problem definition, the annotation tasks, implementation and the validation.

##### 4.1 Definition of building types

In our experiment we sought to qualify crowdsourcing annotations in the context of building type recognition. We focus on the classification of the residential building stock in Germany. Several different typologies for different purposes can be found in the literature. We use a hierarchically structured typology already used in Hecht et al. (2015) differentiating nine residential building types, particularly detached single-family house (SFH), semi-detached SFH, terraced SFH, multi-family house (MFH) in open structure, high-rise buildings as MFH, traditional MFH in row, industrialized MFH in row, block perimeter development and rural houses.

##### 4.2 Task Design

Since the tasks for the crowd members need to be as easy as possible, we chose very basic questions that anybody should be able to answer. Therefore, relevant building criteria are identified which are necessary to separate the individual building types. The identified criteria are: the morphological type, number of floors, housing type, roof type, and the building age. We defined six questions in a single selection mode, each requesting a different building criteria. The survey of the building age is carried out via the façade type separated for the SFH and the MFH. In this case, the annotators is asked to assign the most similar façade (out of a set of typical façades of a certain building period) to a building.

Task	Question	No. of options	Options
T1: Morphological type	Which type of building do you see?	3	Detached house, semi-detached house, row house
T2: Number of floors	How many storeys do	9	1, 2, 3, 4, 5, 6, 7, 8-15, 15 and more

T3: Housing type	you see (including ground storey)? Do you see a single-family house or a multi-family house?	2	single-family house (SFH), multi-family house (MFH)
T4: Roof type	Do you see a flat roof or a steep roof?	2	flat roof, steep roof
T5: Façade type (only MFH)	What type of façade is most similar to the building you see?	5	Wilhelminian Style (1870-1918), Traditional row houses (1918 - 1945), Traditional row houses (1945 - 1990), Industrial row houses (Precast concrete) (1970-1990), Modern construction (after 1990)
T6: Façade type (only SFH)	What type of façade is most similar to the building you see?	3	before 1870 (pre-industrial), 1870-1918, (Wilhelminian style), after 1919 (after first world war)

Table 1. Defined Tasks and the characteristics

##### 4.3 Image data capture

For the cities of Dresden and Hamburg, reference data were available that has been gained by experts through previous fieldwork. In addition to the building geometry, these include postal addresses as well as information about the building type (considering nine types), the building height (in m) and the period of construction. This building-based reference data is the basis for the drawing a random subset of 2,000 building addresses. The address list was used to create image requests using the Google Street View Image API. Using the API, static (non-interactive) views can be defined and embedded into web pages using URL parameters sent through a standard HTTP request. After the creation of the initial (default) URL list, street views were examined manually with regard to their usefulness and recognizability of the image content and, if necessary, URL parameters (e.g. size, fov und pitch) were revised. Approximately 46 % of the street views could not be used due to privacy concerns in Germany. In these cases, houses are blurred in Google Street View. The final data set containing 924 buildings (approx. 100 per building type) has been stored in a database including address data, ULR request, x, y coordinates of the buildings' centroid as well as the ground truth information on the building type, building height, roof type, etc.

#### 4.4 Implementation

We chose task implementation in an online gaming environment with support by Pallas Ludens GmbH (www.pallas-ludens.com), a company located in Heidelberg, Germany, specialized in these activities. In our study, tasks were embedded in computer games by replacing commercial ads with crowdsourcing tasks. We use online games such as Farmerama of the game publisher Bigpoint (www.bigpoint.net) attracting millions of desktop users in social networks around the world. The number of monthly active users available as a “crowd” is been estimated around 250,000 (Pallas Ludens 2014). The user interface for desktops consist of two components: a display field with the street view image of a building to be interpreted and a selection field for labeling. Interactive radio buttons with symbolic illustrations (showing category text on hover) support the annotation process. Users are asked to just click on one of the categories (see example in Figure 2).

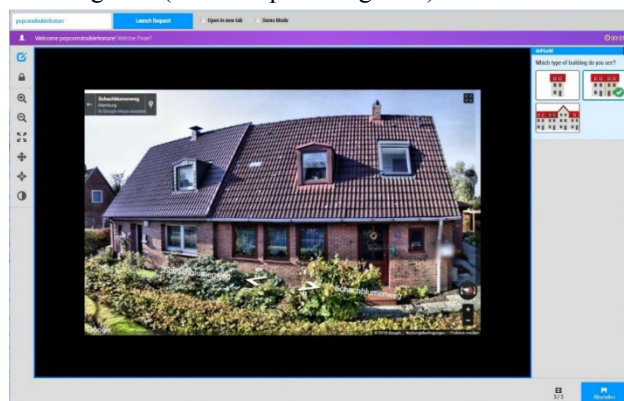


Fig. 2. Example of prototype interface (task 1) containing the display field with a street view of a semi-detached single-family house (left) and selection field with the three clickable answer options: detached, semi-detached and row house (right).

The results of the annotation process conducted and controlled by Pallas Ludens GmbH, lead to structured output files using JSON (JavaScript Object Notation). After conversion into a comma-separated values file (CSV), the following values are available for each annotation:

- annotation\_ID: annotation identifier (integer)
- task\_ID: task identifier number (integer)
- image\_ID: image identifier (integer)
- creator: user name as annotator identifier (text)
- result: label / category (text)

This data is the basis for the statistical analysis and validation. To enable a comparison with external reference data, the majority class is determined for each image from the multiple responses.

#### 4.5 Quality assessment

There are several ways of assessing quality of from crowd-sourced annotations. A common approach is to compare the data with external ground truth information and to calculate accuracy measures (external quality assessment). An introduction of measures of thematic classification accuracy give Congalton and Green (1998),

Foody (2002) and Liu et al. (2002). We used the overall accuracy that is calculated by dividing the total of correct annotations by the total number of annotations. Further, category-level accuracy measures such as the producer's accuracy (PA) and the user's accuracy (UA) representing individual accuracies for each category have been computed based on an error matrix (Congalton and Green, 1998). The error matrix reference data is represented in the columns and the classified data in the rows. The PA gives the ratio between correctly annotated objects and total number of reference objects of that category. The UA is the ratio of the correctly annotated objects of a certain category to the total number of all objects annotated belonging to the category.

		Error matrix taken from Congalton and Green (1998)			
Measures	Notation				
Overall Accuracy	$OA = \frac{\sum_{i=1}^k n_{ii}}{n}$	<div> <div>j = Columns (Reference)</div> <div>Row Total</div> <div> <div>1</div> <div>2</div> <div>k</div> <div><math>n_{i+}</math></div> </div> </div>			
Producer's accuracy	$PA = \frac{n_{jj}}{n_{+j}}$	<div> <div>i = Rows (Classification)</div> <div> <div>1</div> <div>2</div> <div>k</div> <div><math>n_{+i}</math></div> </div> </div>			
User's accuracy	$UA = \frac{n_{ii}}{n_{i+}}$	<div> <div>Column Total</div> <div> <div><math>n_{-1}</math></div> <div><math>n_{-2}</math></div> <div><math>n_{-k}</math></div> <div><math>n</math></div> </div> </div>			

Table 2. Measures for external validation

External validation always requires sufficient reference data, which is not always available. Therefore, researchers have developed approaches to evaluate the quality of a dataset with the aid of intrinsic indicators as a proxy (Senaratne et al. 2016). In our experiment, we focused on aspects of agreement and diversity on instance level (each image) using measures given in Table 3. The inter-annotator agreement (IAA) is a measure that reflects how reliable/confident a majority vote is by calculating the ratio of the number of annotations in the majority category and the total number of annotations per image. In other words, it is the agreement among annotators. In order to measure the diversity we use Shannon's Diversity Index (SHDI) and Shannon's Evenness Index (SHEI) known from the domain of landscape structure analysis (McGarigal and Marks 1995). SHDI is a quantitative measure reflecting the amount of information, in particular how many different classes occur per image, and simultaneously takes into account the occurrence of each class. Since SHDI is very sensitive to the number of possible categories  $k$ , the SHEI was



introduced, which is based on SHDI normalized by dividing by the maximum diversity present in case of equal class distribution (McGarigal and Marks 1995).

Intrinsic Measures	Notation
Inter-annotator agreement (IAA)	$IAA = \frac{a_{mc}}{a}$ <p>where <math>a_{mc}</math> is the number of annotations in the majority category and <math>a</math> the total number annotations per image</p>
Shannon's Diversity Index (SHDI)	$SHDI = - \sum_{i=1}^k P_i * \ln P_i$ <p><math>k</math> is the number of possible categories, and <math>P_i = a_i/a</math>, the proportion of annotations in the <math>i</math>th category (<math>i = 1, \dots, k</math>), where <math>a_i</math> is the number of annotations in category <math>i</math> and <math>a</math> is the total number of annotations.</p>
Shannon's Evenness Index (SHEI)	$SHEI = \frac{SHDI}{\ln k}$ <p>where <math>k</math> is the number of possible categories</p>

Table 3. Intrinsic measures

## 5. Experimental Results

In this section, we present the first results of our experimental study. After presenting a descriptive statistic of the output data, the results are evaluated by using the defined intrinsic and external quality measurements.

### 5.1 Descriptive Statistics

Table 4 gives an overview of the data in terms of the amount of images, categories, annotations, annotators and their relations. The latest column shows the number of annotations guaranteed for most of the images, which means that more than 95% of the images of each task have more than 14 annotations (see also histogram in Figure 3).

Task	images	possible categories	annotations	annotators	annotations per image (mean)	annotations per annotator (mean)	annotations guaranteed for 95% of images
T1: Morphological type	924	3	17644	2888	19,1	6,1	14
T2: Number of floors	924	9	13710	2097	14,8	6,5	15
T3: Housing type	924	2	13857	1001	15,0	13,8	14

T4: Roof type	639	2	13497	1047	21,1	12,9	20
T5: Façade type (only MFH)	519	5	18159	3525	35,0	5,2	20
T6: Façade type (only SHF)	405	3	14175	2896	35,0	4,9	35

Table 4. Overview of the resulting data and overall numbers

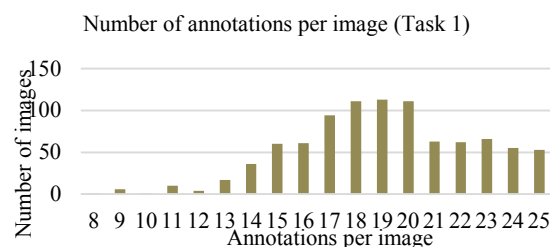


Fig. 3. Number of annotations per image for task 1 (Morphological type)

### 5.2 Intrinsic and external validation

In the following, the results of each task are described and evaluated using the defined internal and external measures (Table 5). The overall accuracy (OA) was determined by comparing the results of the majority vote with external reference data and computing the number of correct annotations and false annotations. The table shows the highest accuracy for task 2 (number of floors), task 3 (housing type) and task 4 (roof type) with OA values over 0.84. A similar picture is obtained by considering the intrinsic dimensions. The values of the inter-annotator agreement (IAA) are also high for tasks 3 and 4, which means that there is a high agreement between the annotators. The corresponding values for the diversity index SHDI are low, which suggests that many annotators have chosen the same class. However, the accuracy of the detection of facade types used for the reconstruction of the building age (task 5 und 6) is limited. Apparently, the assignment of the buildings to a certain type of façade may be too difficult, or only a few users are able to make these assignments correctly. Furthermore, the quality of the results of task 1 (morphological type) is at this stage unsatisfactory. Further investigations are needed in order to identify the causes for this misclassification. Initial checks indicate that there is a frequent confusion between the row houses and the semi-detached houses. The reason for this confusion is most likely a large number of street view images with an unfavorable view frames (image section) that do not allow the recognition of the neighboring buildings. Surprisingly, the accuracy of the recognition of the housing type is relatively high when looking at OA (0.86) and IAA (0.84). Here we had expected less accuracy.

Task	External measures			Intrinsic measures		
	Correct	False	OA	IAA	SHDI	SHEI
T1: Morphological type	529	395	0,57	0,61	0,56	0,51
T2: Number of floors*	803	121	0,87	0,51	0,82	0,38
T3: Housing type	792	132	0,86	0,84	0,25	0,36
T4: Roof type	538	101	0,84	0,76	0,35	0,51
T5: Façade type (only MFH)	247	272	0,48	0,42	0,92	0,57
T6: Façade type (only SHF)	262	142	0,65	0,59	0,61	0,56

Table 5. Overall accuracy (OA) and mean values of the inter-annotator agreement (IAA), Shannon's Diversity Index (SHDI) and Shannon's Evenness Index (SHEI)

## 6. Discussion and future research

In this paper, we propose an integrated system for automatic classification of building footprints that supports a crowd-sourced data collection component that can be used for training and validation. In an experimental study, the quality of crowd-sourced annotations on street view imagery is assessed. The annotations are related to a set of selected building characteristics relevant for distinguishing residential building types. These first results initially provide a rough overview of the quality. A deeper insight would be obtained by carrying out a more detailed analysis by having a look at the quality for different building types, calculating error matrices, and computing building-type-specific measures such as the producer's accuracy and the user's accuracy. Furthermore, the quality of the building types automatically derived from the building characteristics still needs to be evaluated.

For this experimental study, we chose online game environment for task implementation. However, open micro-task platforms such as Crowdcrafting can also be considered in future studies. The advantage of this platform is that it does not incur any costs in comparison to the use of commercial platforms. With regard to the image data used, the suitability of alternative data sources can be investigated such as Wikimapia or Mapillary. The VGI platform Wikimapia contains a large stock of geocoded images of buildings, while Mapillary offers street-level images. Another interesting data source might be the Bird's Eye Views from Microsoft Bing Maps offering multi-perspective views of buildings. The views can be provided to the crowd as an embedded interactive window using the provided API. A comparison of the different data sources could lead to a specific data set being particularly suitable for a certain tasks. For

example, the morphological type in the Birds Eye View is certainly better recognizable than in Google's Streetview images.

A further step will be to explore the relationship between the intrinsic measures and the data quality based on the external measurements. Thus, the question can be investigated whether the quality of an annotation can be estimated on the intrinsic measures solely. Furthermore, the data at annotator-level can be analyzed in order to estimate the annotator's credibility and to identify good and bad annotators. These findings would then form the basis for the development of suitable filters (selection criteria) in the post-processing/quality control step. By using only the high-quality annotations from the best annotators, the quality of the ground truth data can be improved. This, finally improves the accuracy of the whole system, particularly the machine learning classifier for predicting the building types based on the digital topographic data.

Even if further research is necessary, we believe that crowdsourcing in combination with geospatial web technologies have the potential to massively reduce time and costs in collecting ground truth data for training and validating all kind of predictive models. Especially the huge time savings can lead to a much faster mapping which is essential in disaster mapping.

## 7. Acknowledgements

We are grateful to Jörg Hennersdorf who supported us in reference data collection and Jenny Herbrich for data preparation. We would also thank Pallas Ludens GmbH for the technical support in integrating the annotation tasks. This project was supported by Pallas Ludens GmbH as one of three projects chosen in conjunction with the Pallas Ludens Academic Crowdsourcing Challenge in 2016. Further, parts of this work were supported by the German Federal Ministry of Education and Research under project ScaDS Dresden/Leipzig – Competence Center for Scalable Data Services and Solutions (BMBF 01IS14014B).

## 8. References

- Albuquerque, J.P., Herfort, B., Eckle, M. (2016). The Tasks of the Crowd: A Typology of Tasks in Geographic Information Crowdsourcing and a Case Study in Humanitarian Mapping. *Remote Sens.* 2016, 8, 859.
- Congalton, R.G., Green, K. (1998). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 1st ed. CRC Press.
- Donath, A., Kondermann, D. (2013) Is Crowdsourcing for Optical Flow Ground Truth Generation Feasible?. In: Chen M., Leibe B., Neumann B. (eds) *Computer Vision Systems. ICVS 2013. Lecture Notes in Computer Science*, Vol. 7963. Springer, Berlin, Heidelberg.
- Estellés-Arolas, E., González-Ladrón-de-Guevara, F. Towards an integrated crowdsourcing definition. *J. Inf. Sci.* 2012, 38, 189–200.

- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80 (1), 185–201.
- Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., Obersteiner, M. (2012). Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, 31(2012), 110–123.
- Hahmann, S., A. Zipf, A. Rousell, A. Mobasheri, M. Rylov, E., Steiger, L. Loos, J. Lauer (2016). GIS-Werkzeuge zur Verbesserung der barrierefreien Routenplanung aus dem Projekt CAP4Access. *AGIT - Journal für Angewandte Geoinformatik*, 2, 328–333.
- Hecht, R., Meinel, G., Buchroithner, M.F. (2015). Automatic identification of building types based on topographic databases – A comparison of different data sources. *International Journal of Cartography*, 2015(1), 18–31.
- Henn, A., Römer, C., Gröger, G., Plümer, L. (2012): Automatic classification of building types in 3D city models Using SVMs for semantic enrichment of low resolution building data. *GeoInformatica*, 16 (2), 281–306.
- Donath, A., Kondermann, D. (2013). Is crowdsourcing for optical flow ground truth generation feasible? In: *Computer Vision System*, Springer, 193–202.
- Laso Bayas, J.C., See, L., Fritz, S., Sturn, T., Perger, C., Dürauer, M., Karner, M., Moorthy, I., Schepaschenko, D., Domian, D., McCallum, I. (2016). Crowdsourcing In-Situ Data on Land Cover and Land Use Using Gamification and Mobile Technology. *Remote Sensing*, 8(11), 905.
- Liu, C., Frazier, P., Kumar, L. (2007): Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment*, 107(4) 606–616.
- McGarigal, K., Marks, B.J. (1995). FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure. Forest Science Department, Oregon State University, Corvallis.
- Maier-Hein, L., Mersmann, S., Kondermann, D., Stock, C., Kennigott, H.G., Sanchez, A., Wagner, M., Preukschas, A., Wekerle, A.L., Helfert, S., et al. (2014). Crowdsourcing for reference correspondence generation in endoscopic images. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, Springer, 349–356.
- Hillen, F., Höfle, B. (2015). Geo-reCAPTCHA: Crowdsourcing large amounts of geographic information from earth observation data. *International Journal of Applied Earth Observation and Geoinformation*, 2015, 40, 29–38.
- Howe, J. (2006). The rise of crowdsourcing. *Wired*, 14(6), 1–4.
- Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Business, 1 edition, August
- Pallas Ludens, (2014). We inject human intelligence precisely where automation fails. Presentation. [http://www.spectronet.de/portals/visqua/story\\_docs/vortraege\\_2014/141104\\_vision/141104\\_06\\_02\\_kondermann\\_pallas\\_ludens.pdf](http://www.spectronet.de/portals/visqua/story_docs/vortraege_2014/141104_vision/141104_06_02_kondermann_pallas_ludens.pdf). Accessed 25 June 2007.
- Römer, C., Plümer, L. (2010). Identifying architectural style in 3D city models with support vector machines. *Photogrammetrie – Fernerkundung – Geoinformation*, 5, 371–384.
- 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).
- See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., et al. (2013). Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. *PLoS ONE* 8(7), e69958.
- See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H-Y., Milčinski, G., Nikšič, M., Painho, M., Pödör, A., Olteanu-Raimond, A.-M., Rutzinger, M. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*. 2016; 5(5):55.
- Snow, R., O'Connor, B., Jurafsky, D. and Ng, A. J. (2008). Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, ACM Press, 254–263.
- Steiniger, S., Lange, T., Burghardt, D., & Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS*, 12(1), 31–59.
- von Ahn, L., Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Vienna, Austria*. ACM Press, New York, 2004, 319–326.
- von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, 321(5895), 1465–1468.
- Wurm, M., Schmitt, A., Taubenböck, H. (2016). Building types' classification using shape-based features and linear discriminant functions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 1901–1912.