Applied Geography 97 (2018) 25–34

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Combining automatic and manual image analysis in a web-mapping application for collaborative conflict damage assessment

Christian Knoth*, Sofian Slimani, Marius Appel, Edzer Pebesma

Institute for Geoinformatics, University of Münster, Heisenbergstraße 2, 48149 Münster, Germany

ARTICLE INFO

Keywords:
Collaborative mapping
Conflict damage assessment
Geographic object-based image analysis
Crowdsourcing

ABSTRACT

Remote sensing is increasingly being used by non-profit organizations and international initiatives to localize and document combat impacts such as conflict damage. Most of the practical applications rely on labor-intensive and time-consuming manual image analysis. Even when using crowdsourcing or volunteer networks, the workload can quickly become challenging when larger areas have to be monitored over longer time periods. In this paper, we propose an approach that combines automatic change detection methods with collaborative mapping in a web application for conflict damage assessment in Darfur, Sudan. Settlement areas are automatically detected and searched for destructed dwelling structures by geographic object-based image analysis (GEOBIA). The web application prioritizes these areas based on the detected degree of destruction to guide human analysts to the most important locations. In a user experiment with 30 participants we evaluated the performance of volunteers with and without the automatic prioritization and investigated their mapping sequences. Participants who were guided by the prioritization detected 70.7% more target objects than participants mapping without guidance, who invested parts of their mapping time in examining locations that show little to no destruction.

1. Introduction

1.1. Motivation

The use of satellite-based remote sensing for the monitoring and documentation of violent conflicts has strongly increased in recent years. The growing availability of satellite imagery enables visual access to areas that are hard to reach or too insecure to be covered by ground-based monitoring (Witmer, 2015; Wolfinbarger & Wyndham, 2011). Non-profit organizations and international initiatives are increasingly using these opportunities, e.g., to localize and document conflict damage, to corroborate on-the-ground reports on atrocities or even to report signs of likely upcoming hostilities (see, for example, American Association for the Advancement of Science, 2014b; Amnesty International, 2016; Harvard Humanitarian Initiative, 2012; United Nations Institute for Training and Research, 2011). The main goal often is to put pressure on involved actors through increasing public awareness (Livingston, 2015) and to eventually influence their behavior. In addition, remote sensing has also received attention as an investigatory tool by international judicial bodies such as the International Court of Justice (ICJ) or the International Criminal Court (ICC, see (Wolfinbarger, 2016). Another important application of remote sensing in conflicts is the monitoring of cultural heritage. The destruction of heritage sites due to, e.g., deliberate attacks or looting is a recurring issue during violent conflicts or phases of political instability (Kila, 2016). Satellite imagery has been used to investigate and document the loss of cultural heritage or to assess threats to archaeological sites in a number of countries (see, e.g., American Association for the Advancement of Science, 2014a; Banks, Fradley, Schiettecatte, & Zerbini, 2017; Bewley et al., 2015).

Most of the current practical applications of remote sensing in conflicts rely on manual image analysis by trained experts. This analysis is labor-intensive and time-consuming, especially when analyst resources are limited (Meier, 2011b). One strategy to cope with the immense workload is to distribute it among a larger number of volunteers in crowdsourcing applications, some of which are described in the following section.

However, the manual analysis of remote sensing images remains a labor-intensive task, especially when larger areas have to be monitored over longer time periods or when analysis tasks are more complex (e.g., requiring analysts to compare images from two dates). Even for volunteer networks, the workload can be challenging, not least because only a certain percentage of volunteers is available at any given time (Meier, 2011b).

* Corresponding author.
E-mail address: christian.knoth@uni-muenster.de (C. Knoth).

https://doi.org/10.1016/j.apgeog.2018.05.016
Received 27 July 2017; Received in revised form 17 May 2018; Accepted 18 May 2018

Another approach is to use automatic image processing on remote sensing data to analyze conflict areas. Automatic classification and change-detection methods, for example, have been developed for finding and analyzing combat impacts such as damaged and destroyed building structures (Pagot & Pesaresi, 2008; Sulik & Edwards, 2010), direct and indirect environmental effects of conflicts (Abuelgasim, Ross, Gopal, & Woodcock, 1999; Nackoney et al., 2014), or population displacement and refugee camp evolution (Lang, Tiede, Hölbling, Füreder, & Zeil, 2010).

An important concern in this context is how to effectively and responsibly integrate such methods into workflows of conflict monitoring and analysis. The documentation of conflict impacts, and especially of possible human rights violations, is a precarious field. There is broad consent that reported results in this field should be of highest possible accuracy (Orentlichter, 2016), not least because of the severe allegations they might lead to and the possible consequence of raising false charges (Blitt, 2004; Groome, 2011). Due to the uncertainty of automatic image analysis, the results should not be used as unmediated evidence in this context. Nonetheless, they can support processes of human rights reporting especially with regard to workload and efficiency in monitoring of larger areas. Several authors have argued for developing approaches to combine automatic image analysis with collaborative mapping for the investigation of conflict areas (see, for example, Meier, 2011b; Witmer, 2015).

In this paper we present an approach for such a hybrid detection method to increase the efficiency of manual image interpretation in conflict damage assessment. The proposed framework combines automatic image analysis with a microtasking approach for collaborative mapping that is implemented as a web application. We apply geographic object-based image analysis (GEOBIA) to automatically determine the degree of destruction in a conflict area. Based on this estimation, the web application coordinates the analyses of subsets by guiding analysts to those areas that contain most important information. We conduct a user experiment to investigate the mapping sequence and performance of volunteers and to compare the investigation with and without guidance through automatic prioritization of subsets.

1.2. Background

The advent of new information and communication technologies has had a huge impact on the humanitarian and human rights sectors. The workflows of NGOs have been changed in several facets, e.g. the collection and processing of information, the way humanitarian response is organized, and the array of involved actors (Meier, 2011b). For example, volunteers can support relief operations after disasters by issuing situation reports and help requests via SMS, social media, email and other, or by mapping affected areas based on remote sensing images. Prominent examples include the mapping campaign in the aftermath of the Haiti earthquake in 2010 (Heinzelman & Waters, 2010; Hester, Shaw, & Biewald, 2010), the efforts by the Humanitarian OpenStreetMap Team (HOT) after the earthquake in Nepal in 2015 (Poiani, dos Santos Rocha, Degrossi, & de Albuquerque, 2016) and many others. The Missing Maps project even engages volunteers in the preventative mapping of vulnerable, sparsely mapped areas in order to have high quality spatial data in place immediately when a disaster strikes (Herfort, Eckle, & de Albuquerque, 2016).

This increasing involvement of volunteers can also be observed in the field of fact-finding for human rights and international law. Traditionally the practice of gathering testimony through interviews of witnesses and victims is the main tool to gather information on human rights issues (see, for example, UN Commission of Inquiry on the Syrian Arab Republic, 2016), but new techniques exploiting, e.g., social media, Geographic Information Systems (GIS) and remote sensing are increasingly being used (Alston & Knuckey, 2016; Aronson, 2016).

One of the earliest examples is the Ushahidi platform, which was developed in 2008 to document human rights violations during the post-election violence in Kenya. It allowed witnesses and victims to report incidents via web-form, email or SMS. The reports were combined with additional information to build a crisis map that was updated in near-real time (Okolloh, 2009). The Ushahidi platform has since then been used in a large number of deployments in various scenarios such as the Libya Crisis Map 2011 (Burns, 2014).

An important notion of this development is that it is not merely about enlarging the workforce through the recruitment of volunteers or increasing efficiency by use of new technologies. It also opens up the documentation process to ordinary individuals including those directly impacted by conflicts, thus turning subjects into agents (Land, 2016).

A common strategy to involve volunteers in conflict documentation is collaborative mapping in satellite images using a microtasking approach. Here, the images are divided into a regular grid of smaller subsets that can be investigated individually (Barrington et al., 2011). Different conflict mapping projects have engaged in using volunteers for mapping conflict related issues such as emerging of refugee shelters in Somalia (Meier, 2011a). Amnesty International applied a crowdsourcing approach to map remote settlements in Darfur (Amnesty International, 2017a) and, in a follow-up project, monitor changes in those settlements (Amnesty International, 2017b).

The volunteer projects and communities are usually not focused only on conflict but engage in all kinds of crises (e.g., natural disasters). They differ in terms of size and their organizational form, ranging from networks of trained volunteers with organizing principles such as task sharing, a code of conduct or specific activation criteria (e.g., the Standby Task Force[1] or the GISCorps[2]) to open microtasking and crowdsourcing platforms where everyone can contribute on an ad hoc basis (e.g., Tomnod[3] or the HOT Tasking Manager[4]).

Volunteer networks have also partnered with microtasking platforms and satellite imagery vendors. In these examples, the former mobilized the volunteer workforce while the ladder provided the technological platform for the collaborative mapping and the satellite imagery to conduct the analysis on (Meier, 2013).

An important question for such platforms is how to divide, select, and prioritize tasks such that individual volunteers can efficiently contribute to the overall goal. Previous research on prioritizing tasks for collaborative mapping has been conducted in a natural disaster scenario (Hu, Janowicz, & Couclelis, 2016). The prioritization in this case relied on additional information on road networks. The concept applied information value theory to determine priorities of areas with regard to road connectivity for relief trips into disaster affected areas.

2. Analysis framework

The approach proposed in this paper is to leverage the capabilities of automatic image analysis to support the manual conflict damage assessment in remote sensing images. For this purpose, two aspects are considered. First, automatic algorithms can be used to *reduce* the areas of interest for a specific investigation. This is specifically important in large, rural regions where the areas of interest, i.e. the areas possibly affected by conflict, cannot be clearly defined a priori. This is also the case in our study area, Darfur, where small settlements are spread over wide areas and the locations of these settlements are sometimes very difficult to determine (American Association for the Advancement of Science, 2007).

Second, information derived from automatic methods can help *coordinating* the engagement of volunteers. This includes the prioritization of areas guiding the assignment of analysts and the sequence in which different areas are to be mapped. There is only little research about the

[1] see http://www.standbytaskforce.org/(accessed 2017-07-17).
[2] see http://www.giscorps.org (accessed 2017-07-17).
[3] see http://www.tomnod.com/(accessed 2017-07-17).
[4] see http://tasks.hotosm.org/about (accessed 2017-07-17).

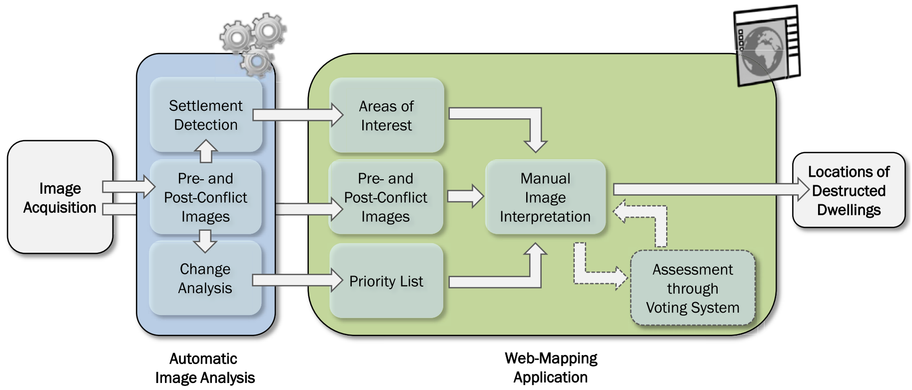

Fig. 1. Conceptual framework to combine automatic and manual crowd-based image analysis.

actual mapping sequence of volunteers in microtasking approaches without additional guidance. Hu and Janowicz (2016) analyzed the online mapping efforts by volunteers in the wake of three natural disasters. They compared the sequence in which tiles of the area were mapped to additional datasets representing potentially critical information within those tiles, namely population distribution and road infrastructure. Their results indicate that the mapping sequence was not strongly correlated with these additional datasets.

The following sections describe our approach to address the two abovementioned aspects of *reducing* the areas of interest and *prioritizing* important subsets in the context of conflict damage assessment (see Fig. 1). This includes the automatic image analysis for delineating settlement areas and detecting changes in these areas (see Section 2.1), and the concept for deriving the priorities of image subsets (see Section 2.2) within the web application (see Section 2.3).

In contrast to the studies mentioned above, our approach does not use any additional data to determine the location and priority of important subsets, but relies only on the analysis of the bi-temporal image data. The overall goal of is to guide analysts to tiles where a lot of destructed buildings are to be found.

2.1. Automatic settlement detection and change analysis

The algorithm automatically detecting destruction in the bi-temporal images is a process of geographic object-based image analysis (GEOBIA) and change detection. GEOBIA is specifically well suited for the analysis task in our framework. The incorporation of features such as shape, size or hierarchical relations in addition to spectral information allows to selectively extract specific objects (in our case settlement areas and destructed dwelling structures). In addition, the object-based approach naturally integrates raster and vector-based analysis (Blaschke et al., 2014), which facilitates the extraction of results in a form that can easily be integrated into larger geospatial analysis workflows (see Section 2.2).

In a first step, the images are searched for areas where settlement structures exist. This is done on a coarse object level created by chessboard segmentation with segments having an edge length of 25 m. The process performs an edge detection after Lee (1983) and detects anthropogenic structures based on the texture (using grey-level co-occurrence matrix measures, see Haralick, Shanmugam, & Dinstein, 1973) and the intensity of the edges per segment. In addition, the proximity of possible settlement segments to each other and the size of resulting settlement areas is taken into consideration.

In a second step, the areas identified as containing settlements are segmented on a finer scale using a morphological closing operator and subsequent multiresolution segmentation (Benz, Hofmann, Willhauck, Lingenfelder, & Heynen, 2004) to delineate objects such as single dwellings. Those objects are subject to the object-based change analysis. Changed structures are identified by their difference in change to the corresponding settlement areas regarding spectral and structural (edges) features. Target objects (i.e., dwelling structures) are then distinguished from similarly changed, but non-relevant objects (e.g., disappeared fences) using features such as shape, size or topological relations (for details on the algorithm, see Knoth & Pebesma, 2017).

The process returns two results, a shapefile of the settlement areas determined in the first step (including a buffer), and a shapefile of the destructed dwelling objects identified in the second step.

2.2. Microtasking and collaborative mapping

Both results of the abovementioned image analysis are used in the framework for organizing and facilitating the collaborative mapping. The settlement objects are used to define areas of interest whereas the result of the destruction detection is the basis for the prioritization of these areas.

Initially, the study area is divided into equally sized subsets that can be investigated individually by volunteers. The areas should be small enough to be quickly analyzed by single volunteers but large enough to make a substantial contribution (Barrington et al., 2011). Since the analysis tasks in our example are relatively complex and the target settlements consist of large numbers of very small dwelling structures, we chose a small subset size of 100 m by 100 m. Subsets (tiles) located in areas that have been detected as containing settlement structures are highlighted on the map. For organizing the collaborative mapping tasks, users are provided with a list of important tiles in descending order of priority. This priority is initially determined based on the number of huts detected as disappeared by the algorithm. When tiles are investigated by volunteers, the prioritization is updated accordingly, i.e., the number of previous visitors influences the priority of a tile. For example, areas with a higher number of possibly destructed dwellings and with the number of manual investigations lower than a certain threshold get higher priorities.

Through this dynamic prioritization, it is possible to guide volunteers according to different preferences and requirements of a specific use case. If the main goal is to distribute volunteers among all possibly affected areas, this threshold can be set to 1. In this case, each tile that has been investigated by one volunteer will be given lower priority. If the goal is to have the strongly affected areas mapped by several volunteers before proceeding to less affected tiles, a higher threshold can be chosen. In this case, tiles with a large number of possibly destructed dwellings will retain a high priority until they have been investigated by a higher number of volunteers. Alternatively, other rules can be used to assign priorities, e.g., the degree of disagreement between previous investigations (variability of the results of different users).

2.3. Implementation as a web application

The described tool for manual conflict damage assessment is

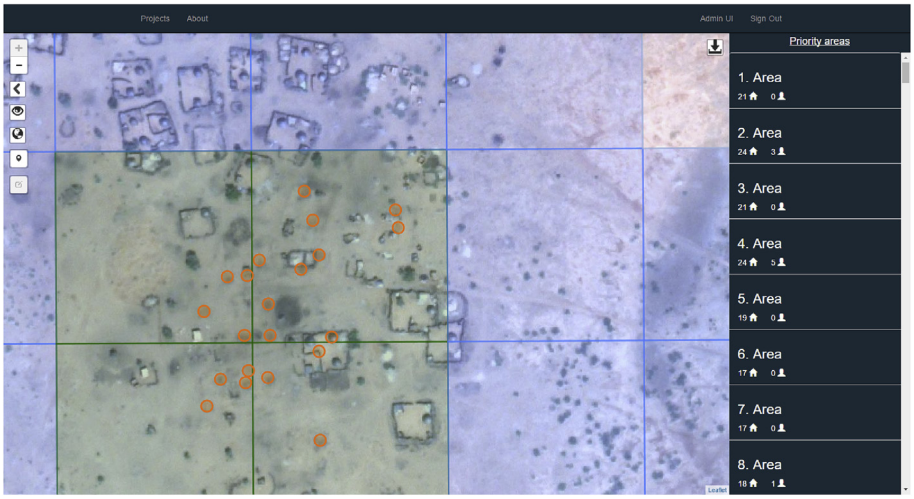

Fig. 2. Screenshot of the graphical user interface. It shows the post-conflict image overlaid by tiles in areas of interest (blue), previously investigated tiles (green) and tags indicating destroyed dwelling structures. The list on the right shows the priority areas. Image © 2017 DigitalGlobe. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

developed as a web application to lower the barriers for users to engage in the collaborative mapping (Fechner, Wilhelm, & Kray, 2015; Morgan, Gilbert, McDonald, & Zachry, 2014). The graphical user interface (GUI) provides standard map interaction (zooming, panning, jump to full extent) as well as functionality to switch between the pre- and post-conflict images, to tag destructed dwellings, and to save edits (see Fig. 2).

The main viewer shows the imagery, which is overlaid by a semi-transparent layer showing the grid of tiles. Those tiles that are located in settlement areas (see Section 2.1) are highlighted in blue. Tiles that have been mapped before, are highlighted in green. Users can click on any of the tiles to start mapping. The web-mapping application also provides a voting system that allows to assess the uncertainty of results and can be used to combine results of single volunteers based on consensus. If a tile has been mapped before, users can vote on each object previously identified as destructed dwelling to confirm or object the decision.

In addition, the GUI shows a list of tiles containing dwelling structures that have been identified as destroyed by the automatic method. These tiles are ordered in descending order of priority (see Section 2.2). Each entry of the list includes the number of dwellings identified as destructed, and the number of volunteers who have already examined the corresponding tile. By clicking on an entry in the list, the current view is directly zoomed to the corresponding tile. However, it is also possible to zoom and pan freely.

Users with administrator rights (e.g., project leaders or initiators of a mapping campaign) can download the resulting objects as GeoJSON or as a shapefile. For all objects, the number of positive and negative votes and the timestamps of edits are stored as attributes in the downloaded file.

The web application was implemented in JavaScript and uses well established web technology frameworks and libraries. On the server side, the NodeJS runtime environment and the web application frameworks ExpressJS and KeystoneJS are used whereas Pug is used for rendering. The interactive map was realized using the JavaScript library Leaflet. The data is stored in the document-oriented database MongoDB using the GeoJSON format, which is widely adopted in the field of open-source geospatial web applications.

3. User experiment

3.1. Study area and data

The data for the user experiment consisted of two very high-resolution satellite images of the Negeha region in Darfur, Sudan. The 'post-conflict' image was acquired by the WorldView-2 satellite in December 2010, after a reported attack in the area. The 'pre-conflict' image was acquired by the GeoEye-1 satellite and dates back to January 2010, which is before the alleged attack. Both images have a spatial resolution of 0.5 m. The area covered is about 105 km^2 and contains several small villages spread over the whole territory. The villages show different extents of destruction (see Fig. 6). Both images were visualized as true color RGB composites for the collaborative mapping. The imagery for our research was provided by the Geospatial Technologies and Human Rights Project of the American Association for the Advancement of Science. The automatic image analysis method was evaluated by comparing its results to reference data collected by visual inspection. This resulted in a user's accuracy of approximately 73% and a producer's accuracy of approximately 75% in this study area (for details on the evaluation, see Knoth & Pebesma, 2017). Therefore, the total amount of destructed dwellings that is used for the prioritization can be expected to be slightly overestimated.

3.2. Experimental setup

In total, 30 volunteers participated in the user experiment. Eighteen of them were male and 12 female. Their age ranged from 21 to 61 (mean 28.8, standard deviation 7.7). We did not make the experiment openly available but conducted the study in a controlled test environment to minimize the influence of external factors such as technical equipment or internet connectivity. Most of the participants were recruited from students and staff of the Department of Geosciences of the University of Münster. Seven participants stated to have previous experience in the visual analysis of remote sensing data. The main characteristics of our test group of volunteers correspond quite well with the patterns found by the Humanitarian OpenStreetMap Team (see Section 1.2) in their study (Humanitarian OpenStreetMap Team, 2017) on the state of their volunteer community conducted in 2017: regarding age structure, the largest group in this survey was formed by participants between 25 and 34 years with just over 43% (about 47% in our study), followed by the group between 18 and 24 years with about 20% (about 33% in our study) and the group between 35 and 44 years with about 19% (about 17% in our study). They also found an imbalance regarding gender with about 69% male and 29% female respondents (60% male and 40% female in our study). In addition, the report mentions a tendency for members to come from GIS or technical backgrounds (Humanitarian OpenStreetMap Team, 2017).

The main goal of the user experiment was to evaluate the possible gain in efficiency of the mapping processes by providing the automatic prioritization. The second goal was to qualitatively analyze the mapping sequence of volunteers in conflict damage assessment with and without guidance by the priority list.

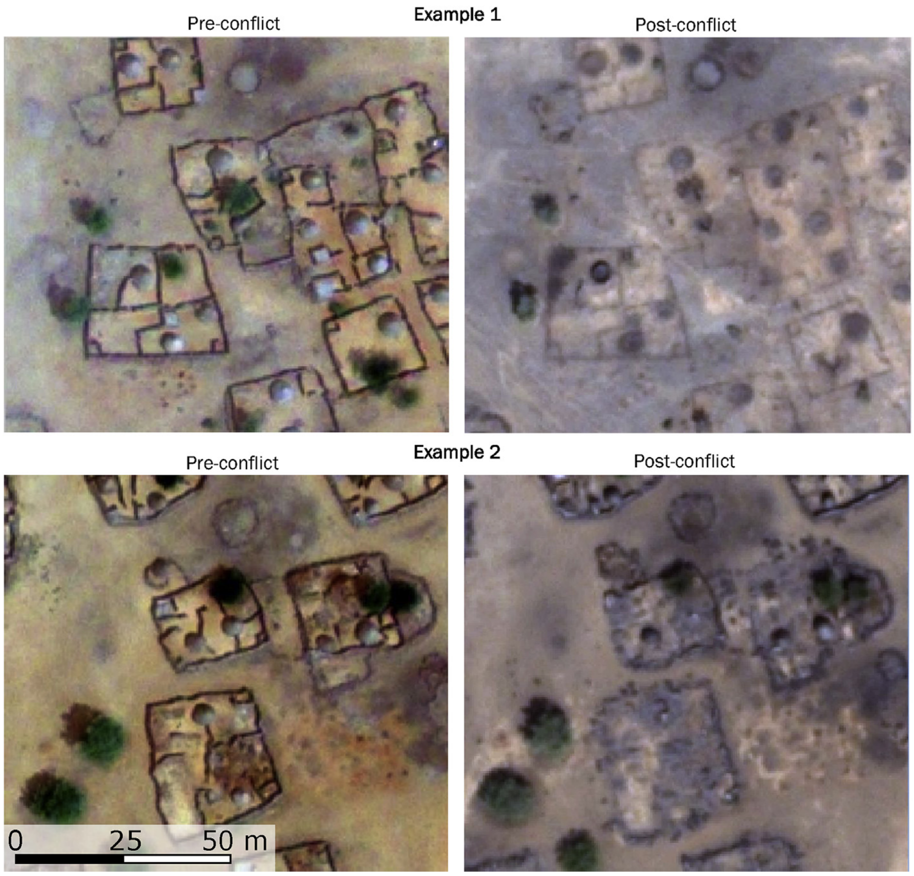

Fig. 3. Pre- and post-conflict images of two subsets of the study area showing typical examples of destroyed dwelling structures. Images © 2017 DigitalGlobe.

The experimental setup faces several challenges. The behavior and efficiency of volunteers should be evaluated in a campaign scenario, i.e., taking the possible influence of the results from other participants into account. At the same time several independent repetitions of the experiment need to be achieved in order to reduce the influence of chance or unusual user behavior.

To reconcile these requirements, the participants were randomly split into groups simulating small mapping campaigns. We created 10 campaigns with three participants working on each campaign. All campaigns took place in the same study area described above. Five campaigns were assigned to the scenario with and five to the scenario without prioritization. The participants stating previous experience in remote sensing (see above) were randomly distributed among the groups. Three of them worked in the scenario with priority list and 4 in the scenario without priority list.

All participants worked on the same machine for the experiment. Before starting the actual mapping process, each participant received a short instruction including example image subsets specifying the target objects of the campaign (two examples are shown in Fig. 3). Afterwards they were presented with the web application and given 2 min time for getting acquainted with the user interface outside the actual study area. In the subsequent mapping phase, participants were asked to search for destroyed dwellings for exactly 10 min.

The first participant of each campaign started mapping an empty project without any previously investigated tiles. The second and third participants in each campaign started working in the same project with the tiles mapped by previous volunteers of that campaign being visible. For this experiment, we chose a prioritization scheme in which each tile was removed from the priority list after being investigated by one participant.

3.3. Evaluation

To evaluate the results of the user experiment, we first analyzed the overall performance of the participants. We counted the number of objects that were identified during the fixed amount of time (10 min per participant) for all users and grouped them depending on whether they were guided by a priority list or not. We only counted correctly detected objects, i.e., objects that we could confirm as having been destroyed. We then applied statistical tests to evaluate whether performance differences can be assumed to be non-random. Due to the small sample size and potentially unequal variances we applied Welch's *t*-test on the square root transformed number of correct detections, a nonparametric Mann-Whitney-*U*-test, and a simple permutation test. To evaluate the prioritization, the number of correct detections was also analyzed per tile and in comparison to the automatically assigned priority. In addition to the overall performance, we compared the mapping behavior of the participants with and without guidance in a qualitative analysis. We examined the sequence in which the tiles were mapped by each participant and how often the different parts of the study area were investigated throughout the two groups.

4. Results

4.1. Overall performance

Table 1 and Fig. 4 summarize the performance of participants with and without prioritization.

In total, the participants mapping with guidance by the priority list investigated 100 tiles while the participants without prioritization examined 128 tiles. The number of correctly mapped target objects, i.e., destructed or disappeared dwelling structures, was 812 for the participants with and 476 for the participants without priority list. The mean

Table 1
Basic summary statistics of the user experiment results for groups with and without prioritization.

Priority list	Number of correctly detected destructions per user						Total number of investigated tiles
	Min	Median	Mean	Max	Stdev	Sum	
No	2	25	31.73	103	26.61	476	128
Yes	25	49	54.13	103	19.58	812	100

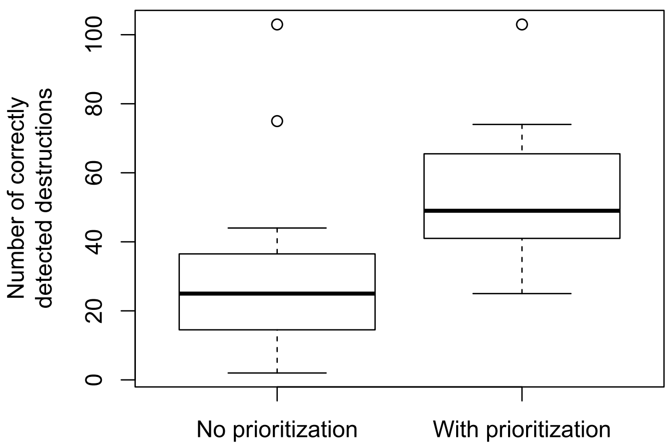

Fig. 4. Summary of the number of correctly detected destructions for groups with and without prioritization guidance.

number of correctly mapped target objects per participant was 54.1 for the group with and 31.7 for the group without the priority list. Thus, the prioritization resulted in a gain of 70.7% in the mean number of correctly identified target objects.

Because of the more extreme upward outliers in the unguided user group (see Fig. 4), the difference is even bigger when comparing the medians (49 with and 25 without prioritization). Accordingly, the variability in performance was higher for the participants without priority list (standard deviation of 26.6) than for the group with prioritization (standard deviation of 19.6).

For the participants with priority list, the average number of identified target objects per user decreased with the position of that user within the order of the corresponding campaign. The users who came first in the order of their campaign detected on average 62.2 objects, the users who came second detected 52.6, and the users who came third detected 47.6 objects. No such trend was observed for the participants without priority list (first: 31.2, second: 43.2, third: 20.8).

Despite the limited sample size ($n = 15$ for both groups), the applied statistical tests suggest significant differences between the overall mapping performance of the groups. The one-sided two sample Welch's t-test on the square root transformed number of correct detections rejects the hypothesis that the performance of users without prioritization is greater than or equal to the performance of users with prioritization at $p = 0.0027$. The non-parametric Mann-Whitney-U-test yields $p = 0.0013$. Both results support the finding that the distributions differ significantly, without assuming equal standard deviations. We furthermore drew 100000 random permutations by reassigning observations to the groups and derived empirical mean and median differences to compare with the original grouping. As a result, only 0.45% of the empirical median and 1.33% of the empirical mean values had larger absolute group differences than the original grouping, further supporting our finding.

4.2. Mapping sequence

The qualitative analysis of the results of our study revealed certain patterns regarding the mapping sequence of the volunteers. For the participants with prioritization, two different strategies became apparent. Some participants followed the priority list very closely, always mapping tiles that at this point had high priorities. Those participants examined several different settlement areas distributed among the study area. Most participants, however, used the priority list mainly to

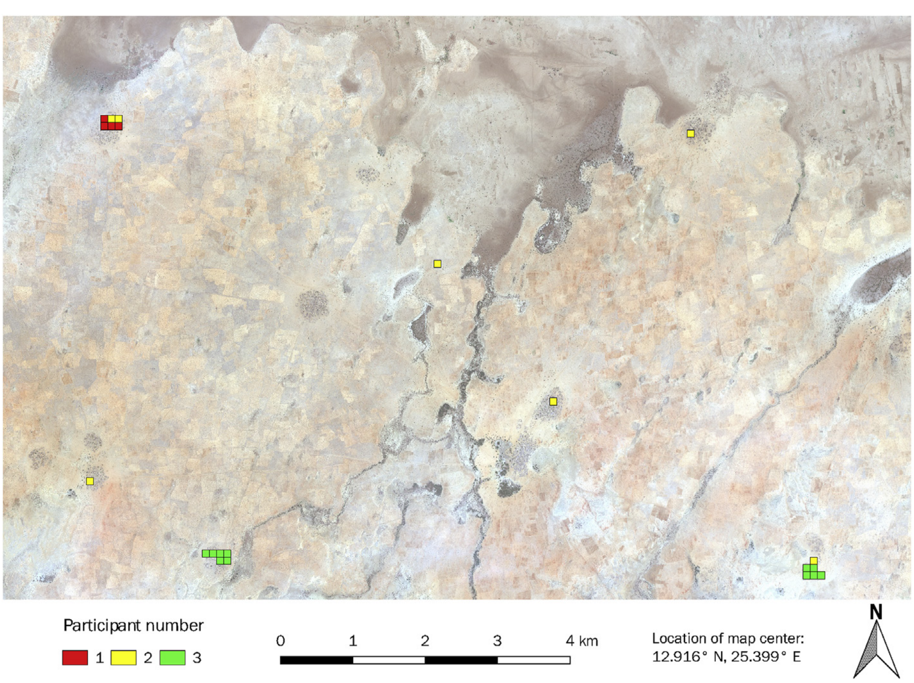

Fig. 5. Overview of the mapping sequence of participants in one group of the user study (mapping with priority list), overlaid on a semi-transparent image of the study area for orientation. Tiles mapped by the first participant in the group are shown in red, tiles mapped by the second and third participant are shown in yellow and green, respectively. Background Image © 2017 DigitalGlobe. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

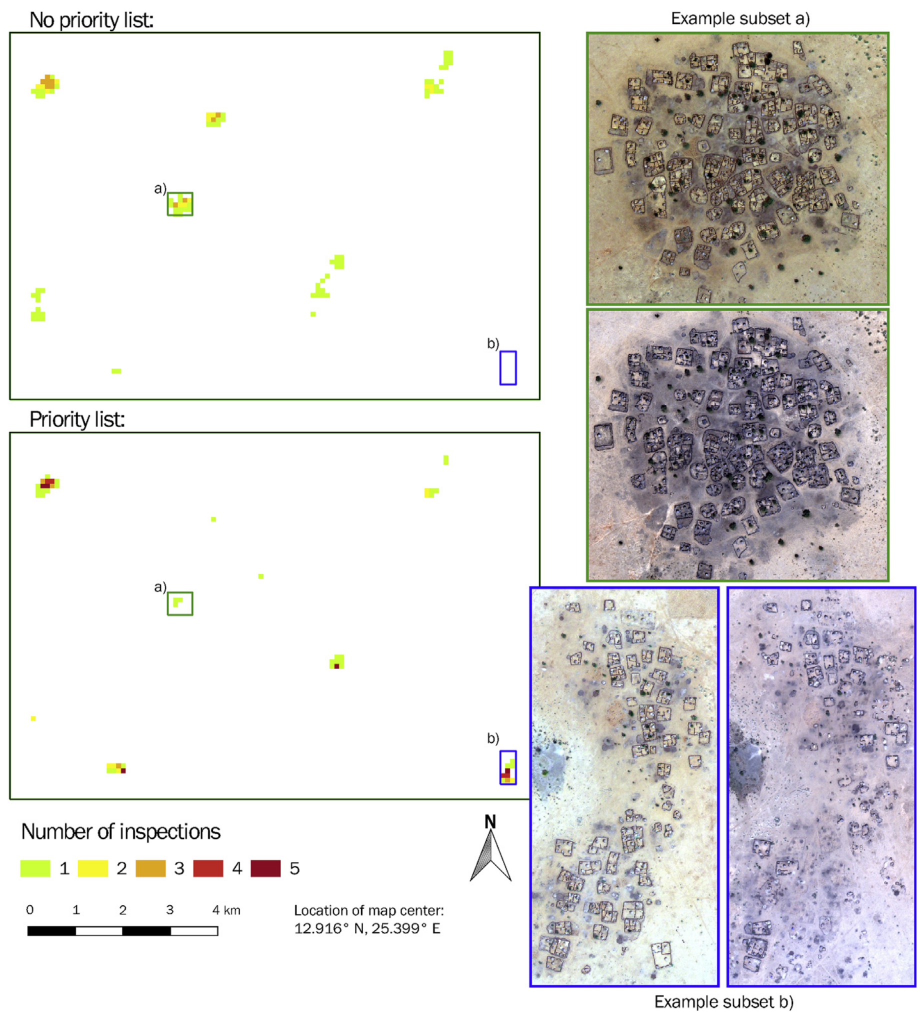

Fig. 6. Comparison of the number of investigations per tile for the participants without (upper left) and with (lower left) priority list. The example subsets a) and b) show pre- and post-conflict villages with low (upper right) and high (lower right) degree of destruction. Images © 2017 DigitalGlobe.

zoom to a highly affected area. After finishing the first tile, they continued to map the destruction in one or more neighboring tiles of the same settlement before continuing with the next tile on the priority list.

Fig. 5 shows the usage pattern of a group with priority list, which exemplifies the different strategies. Here, the second participant of that campaign strictly followed the prioritization and thus changed the location several times. The third participant continued to investigate the neighborhood of the tile that was first suggested by the prioritization before proceeding to the next tile in the priority list. The first participant of that campaign remained in the village containing the first tile on the priority list for the whole mapping time.

The participants without priority list, as expected, tended to remain longer in a specific village with none of them changing the settlement area more than once during their mapping task. On average, the participants with priority list visited ≈ 2.5 different settlements while those without priority list worked in ≈ 1.4 different villages.

Regarding the locations and spatial distribution of investigated tiles, the overall pattern of the participants with priority list, as expected, is strongly steered by the prioritization. For the participants without priority list, no clear pattern is visible when comparing the five campaigns. Fig. 6 depicts the number of investigations per tile, aggregated over all participants with and without priority list. It shows that the investigations by participants without priority list were more evenly distributed than those of the group with prioritization. Only a slight concentration of investigations can be observed in villages located in the north-western part of the study area (see Fig. 6). The participants with priority list concentrated more on specific tiles. These tiles in turn, where distributed over different locations within the study area.

The spatial distribution of investigated tiles also reveals that in the group of users without prioritization, more attention was paid on settlements showing a relatively low degree of destruction (see Fig. 6, Example a). At the same time, two villages with a relatively large amount of destruction were given little to no attention (see, Fig. 6, Example b).

In addition, we found the upward outliers in the group without prioritization, i.e. the participants who had identified a lot more target objects than the rest of that group, to be those participants who had mapped in an area with many high priority tiles. Overall, there is a strong correlation between the automatically assigned priority of tiles and the number of correctly identified target objects in these tiles for all participants (see Fig. 7).

5. Discussion

The results of our study indicate that the proposed approach can increase the efficiency of manual conflict damage assessment. The prioritization successfully guided the participants to areas with high degrees of conflict damage. The number of target objects correctly identified was considerably higher for the group with than for the group without the prioritization. Although the sample size is relatively small and the variance especially in the group without prioritization is large, the tests conducted in Section 4.1 suggest the difference to be

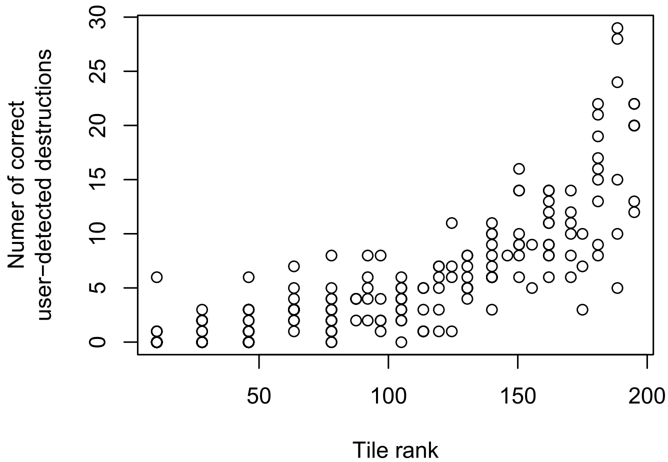

Fig. 7. Comparison of correct user-detected destructions in a tile and the tile's rank according to the automatic image analysis (with higher numbers indicating higher ranks, i.e., higher priorities). Tiles with the same number of automatically detected destructions are given the same rank. The corresponding rank correlation is 0.863.

significant. In addition, there is a strong correlation between the automatically assigned priority of tiles and the number of correctly identified target objects in those tiles.

Our interpretation of the overall performance is backed by the qualitative analysis of the mapping approaches of the participants. It indicates that the mapping strategies can be improved by guidance through automatic prioritization because users in the group without priority list spent more time examining settlement areas with only little conflict impact.

Since our study focused on the performance of volunteers in terms of how many objects they correctly identified in a fixed amount of time, we did not systematically compare the accuracies of the participants with and without prioritization. The producer's accuracy cannot be inferred from our results because the participants were not strictly instructed to conduct an exhaustive search in each tile before proceeding to another one. In addition, reference data on all destroyed target objects in all investigated subsets would be necessary to assess the number of false negatives. The user's accuracy for each participant can be inferred from the proportion of those user-detected targets that we could confirm as being correctly tagged. Here, the average accuracy was 93.96% (standard deviation 5.62%) for the participants with prioritization and 78.98% (standard deviation 19.23%) for the participants without prioritization. The differences can, in part, be explained by the participants with prioritization being guided to subsets that often showed not only a higher numbers of destructions but also more obvious changes (e.g. parts of villages that were completely destroyed). This may have reduced the risk of mistaken interpretations by those participants.

Regarding the method of prioritization, different strategies could be used depending on the specific goal of a mapping campaign (e.g., aiming for coverage of large areas or for a high number of repeated investigations of tiles, see Section 2.2). In large organizations employing manual image analysis, these strategies can be used to organize analysis tasks more efficiently. In our user study, we aimed at tiles being mapped only once per campaign. This was done to focus on the influence of the prioritization on the spatial distribution of investigated tiles, and to achieve comparable results for the groups with and without priority list. This means that the effect of the voting system and different strategies for the prioritization could not be evaluated.

In addition, it might be favorable to change the spatial entity to base the prioritization on. In our user study, the prioritization was based on the number of possibly destructed dwellings *per tile*. Section 4.2 has shown that users who closely follow this prioritization, often switch between different settlement areas. In some cases, it might be desirable to investigate highly affected settlements as a whole (including possibly less affected tiles) in order to grasp the full situation in those settlements before proceeding to another area. Our approach could accommodate this strategy. As mentioned in Section 2.1, the automatic image analysis detects coherent settlement areas, which can be delivered as polygon shapefiles. Therefore, it is possible to assign the same priority to all tiles within a specific settlement with regard to the degree of destruction throughout *the whole settlement*.

For the automatic change analysis, a method yielding probabilities of change instead of a crisp change vs. no-change distinction could be used. These probability values could be included in the prioritization of image tiles instead of or in addition to the number of changed objects. Depending on the focus of a monitoring campaign and the number of available analysts, it might also be favorable to increase the sensitivity of the automatic change detection method. In this study, the producer's and user's accuracy of the applied algorithm were well balanced, i.e., the total number of destroyed buildings was not highly over- or underestimated. The algorithm's sensitivity to change could be increased in order to achieve a higher producer's accuracy while in turn allowing for more false change alerts. As a result, more subsets of the images would be recommended for manual image analysis, increasing the manual workload but decreasing the risk of missing areas of destruction.

With regard to monitoring cultural heritage during conflicts, the potential gain in efficiency depends on the task at hand. When a specific heritage site under threat is known in advance and the area under investigation is of limited size, the direct manual monitoring without prior automatic analysis is probably more efficient. However, there are applications where a screening of larger spaces or continuous observations over longer periods are necessary. One example is the intensified looting of archaeological sites in Egypt in the wake of the Arab Spring in 2011. Parcak, Gathings, Childs, Mumford, & Cline, 2016 identified the remote investigation, potentially including crowdsourcing, to be a promising tool for monitoring cultural heritage loss in this context. Here, the combination of automatic and manual image analysis could make a significant contribution towards reducing workload and time consumption of the investigations.

An important open question that should be addressed in future research is how the gain in efficiency and the effort needed for the automatic image analysis scale up with larger monitoring areas. In a small study area, the gain in efficiency is probably outweighed by the effort needed to apply the automatic image analysis and the corresponding processing. However, if a similar approach can be deployed on a large study region with only little user input required for the automatic processing, we expect the overall possible gain in efficiency to be considerable. Especially when larger areas at risk need to mapped over longer periods, an automatic method alerting human analysts and guiding them to affected areas could be of great use.

To promote the application of the presented approach in practical conflict analysis, further developments facilitating the reuse and practical adoption of GEOBIA are needed. Currently, the majority of object-based approaches (including the change detection method used for this study) rely on proprietary software, which may hamper their practical adoption. However, first steps to create open, reproducible and easily applicable GEOBIA workflows based on open-source software and containerization technology have shown promising results (Knoth & Nüst, 2017). Another important issue is the transferability of the automatic change detection method. The algorithm applied in this study was developed to be robust against changes in image and sensor properties, but aims at changes of objects of a certain shape, size and spatial configuration. To detect changed objects that strongly differ from those targets, the method would need to be adapted (Knoth & Pebesma, 2017). However, GEOBIA has significant potential in this context because considerable progress towards transferability has been made in this field, e.g., through automatic determination of segmentation and classification parameters, and fuzzy classification workflows (Drăguţ, Csillik, Eisank, & Tiede, 2014; Martha, Kerle, & van Westen, 2011; Hofmann, 2016).

6. Conclusions

We showed that combining automatic with manual image analysis improves the efficiency of conflict damage assessment in remote sensing images. The integration of results from automatic methods allows different strategies for coordinating analysis tasks within potentially large groups of volunteers. However, more research is needed to investigate the possible gain in efficiency in analyses on larger geographical scales, and to address the possible effect of different prioritization strategies on the efficient coordination of large numbers of analysts.

The proposed approach is also applicable for smaller teams of trained experts. This can be important in conflict monitoring because of the specifically high requirements regarding reliability in such a politically precarious field and the important role of expert testimonies. Here, the improvement of efficiency can have an even greater impact because of the more limited analyst resources available.

Acknowledgements

We thank the Geospatial Technologies and Human Rights Project of the American Association for the Advancement of Science (AAAS) for providing the image data for our research.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apgeog.2018.05.016.

References

Abuelgasim, A., Ross, W., Gopal, S., & Woodcock, C. (1999). Change detection using adaptive fuzzy neural Networks: Environmental damage assessment after the Gulf war. *Remote Sensing of Environment, 70*(2), 208–223. http://dx.doi.org/10.1016/S0034-4257(99)00039-5.

Alston, P., & Knuckey, S. (2016). The transformation of human rights fact- Finding: Challenges and opportunities. In P. Alston, & S. Knuckey (Eds.). *The transformation of human rights fact-finding* (pp. 3–24). Oxford: Oxford University Press.

American Association for the Advancement of Science (2007). *High-resolution satellite imagery and the conflict in Chad and Sudan.* http://www.aaas.org/content/high-resolution -satellite-imagery-and-conflict-chad-and-sudan, Accessed date: 30 May 2017.

American Association for the Advancement of Science (2014a). *Ancient history, modern Destruction: Assessing the current status of Syria's world heritage sites using high-resolution satellite imagery.* Washington, DC: American Association for the Advancement of Science.

American Association for the Advancement of Science (2014b). *Human rights applications of remote sensing. Case studies from the geospatial technologies and human rights project.* Washington, DC: American Association for the Advancement of Sciencehttp://dx.doi.org/10.1007/s13398-014-0173-7.2.

Amnesty International (2016). *Scorched earth, poisoned air. Sudanese government forces ravage jebel marra, Darfur.* London: Amnesty International.

Amnesty International (2017a). *Decode darfur.* https://decoders.amnesty.org/projects/decode-darfur, Accessed date: 24 May 2017.

Amnesty International (2017b). *Decode the difference.* https://decoders.amnesty.org/projects/decode -the-difference, Accessed date: 24 May 2017.

Aronson, J. D. (2016). Mobile phones, social media and big data in human rights fact-finding: Possibilities, challenges, and limitations. In P. Alston, & S. Knuckey (Eds.). *The transformation of human rights fact-finding* (pp. 441–461). Oxford: Oxford University Press.

Banks, R., Fradley, M., Schiettecatte, J., & Zerbini, A. (2017). An integrated approach to surveying the archaeological landscapes of Yemen. *Proceedings of the Seminar for Arabian Studies, 74*, 9–24.

Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., ... Huyck, C. (2011). Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics, 54*(6), 680–687. http://dx.doi.org/10.4401/ag-5324.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing, 58*(3–4), 239–258. http://dx.doi.org/10.1016/j.isprsjprs.2003.10.002.

Bewley, R., Wilson, A. I., Kennedy, D., Mattingly, D., Banks, R., Bishop, M., ... Zerbini, A. (2015). March/april). Endangered archaeology in the Middle East and North Africa: Introducing the EAMENA project. *Caa2015. keep the revolution going: Proceedings of the 43rd annual conference on computer applications and quantitative methods in archaeology. Siena*.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., ... Tiede, D. (2014). Geographic object-based image analysis - towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing, 87*, 180–191. http://dx.doi.org/10.1016/j.isprsjprs.2013.09.014.

Blitt, R. C. (2004). Who will watch the Watchdogs? Human rights nongovernmental organizations and the case for regulation. *Buffalo Human Rights Law Review, 10*, 261–398.

Burns, R. (2014). Moments of closure in the knowledge politics of digital humanitarianism. *Geoforum, 53*, 51–62. http://dx.doi.org/10.1016/j.geoforum.2014.02 .002.

Drăguţ, L., Csillik, O., Eisank, C., & Tiede, D. (2014). Automated parame- terisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing, 88*, 119–127.

Fechner, T., Wilhelm, D., & Kray, C. (2015, April). Ethermap. *Proceedings of the 33rd annual acm conference on human factors in computing systems - chi '15. Seoul*http://dx.doi.org/10.1145/2702123.2702536.

Groome, D. (2011). *The handbook of human rights investigation. a comprehensive guide to the investigation and documentation of violent human rights abuses* (2nd ed.). North Charleston, SC: CreateSpace Independent Publishing Platform.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 3*(6), 610–621. http://dx.doi.org/10.1109/TSMC.1973.4309314.

Harvard Humanitarian Initiative (2012). *Satellite senstinel project. Making the world a witness. Report on the pilot phase*Cambridge: Harvard Humanitarian Initiative.

Heinzelman, J., & Waters, C. (2010). Crowdsourcing crisis information in disaster. *Physical Review Letters, 96*(25), http://dx.doi.org/10.1103/PhysRevLett.96 .258102.

Herfort, B., Eckle, M., & de Albuquerque, J. P. (2016, May). Being specific about geographic information crowdsourcing : A typology and analysis of the missing maps project in South Kivu. *Proceedings of the iscram 2016 conference. Rio de Janeiro*.

Hester, V., Shaw, A., & Biewald, L. (2010, December). Scalable crisis relief: Crowdsourced sms translation and categorization with mission 4636. *Proceedings of the first acm symposium on computing for development. London*.

Hofmann, P. (2016). Defuzzification strategies for fuzzy classifications of remote sensing data. *Remote Sensing, 8*(6), http://dx.doi.org/10.3390/rs8060467.

Hu, Y., Janowicz, K., & Couclelis, H. (2016). Prioritizing disaster mapping tasks for online volunteers based on information value theory. *Geographical Analysis, 49*(2), 175–198. Retrieved from https://doi.org/10.11117,%202Fgean.12117%20http://dx.doi.org/%2010.1111/gean.12117.

Hu, Y., & Janowicz, K. (2016). Understanding the mapping sequence of online volunteers in disaster response. *International Conference on GIScience Short Paper Proceedings, 1*(1), http://dx.doi.org/10.21433/B3118n57h6zv.

Humanitarian OpenStreetMap Team (2017). *State of the community report 2017.* Washington, DC: Humanitarian OpenStreetMap Team.

Kila, J. (2016). Heritage destruction in the Mediterranean region. *Iemed mediterranean yearbook 2016* (pp. 326–329). IEMed.

Knoth, C., & Nüst, D. (2017). Reproducibility and practical adoption of GEO- BIA with open-source software in docker containers. *Remote Sensing, 9*(3), http://dx.doi.org/10.3390/rs9030290.

Knoth, C., & Pebesma, E. (2017). Detecting dwelling destruction in dar- fur through object-based change analysis of very high-resolution imagery. *International Journal of Remote Sensing, 38*(1), 273–295. http://dx.doi.org/10.1080/01431161.2016.1266105.

Land, M. K. (2016). Democratizing human rights fact-finding. In P. Alston, & S. Knuckey (Eds.). *The transformation of human rights fact-finding.* Oxford: Oxford University Press.

Lang, S., Tiede, D., Hölbling, D., Füreder, P., & Zeil, P. (2010). Earth observation (EO)-based ex post assessment of internally displaced person (IDP) camp evolution and population dynamics in Zam Zam, Dar- fur. *International Journal of Remote Sensing, 31*(21), 5709–5731. http://dx.doi.org/10.1080/01431161.2010.496803.

Lee, J.-S. (1983). Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing, 24*(2), 255–269. http://dx.doi.org/10.1016/0734-189X(83)90047-6.

Livingston, S. (2015). *Commercial remote sensing satellites and the regulation of violence in areas of limited statehood (No. 5).* Philadelphia, PA: Center for Global Communication Studies.

Martha, T., Kerle, N., & van Westen, C. (2011). Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *International Journal of Geographical Information Science, 49*(12), 4928–4943. http://dx.doi.org/10.1109/TGRS.2011.2151866.

Meier, P. (2011a). *Crowdsourcing satellite imagery analysis for Somalia: Results of trial run.* https://irevolutions.org/2011/08/31/results-crowdsourcing-sat-imagery-somalia/, Accessed date: 24 May 2017.

Meier, P. (2011b). New information technologies and their impact on the humanitarian sector. *International Review of the Red Cross, 93*(884), 1239–1263. http://dx.doi.org/10.1017/S1816383112000318.

Meier, P. (2013). Human computation for disaster response. In P. Michelucci (Ed.). *Handbook of human computation* (pp. 95–104). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-8806-4_11.

Morgan, J. T., Gilbert, M., McDonald, D. W., & Zachry, M. (2014, February). Editing beyond articles. *Proceedings of the 17th acm conference on computer supported co-operative work & social computing - cscw '14* (pp. 550–563). . http://dx.doi.org/10.1145/2531602.2531654 Baltimore, MD.

Nackoney, J., Molinario, G., Potapov, P., Turubanova, S., Hansen, M. C., & Furuichi, T. (2014). Impacts of civil conflict on primary forest habitat in northern Democratic

Republic of the Congo, 1990-2010. *Biological Conservation, 170*, 321–328. http://dx.doi.org/10.1016/j.biocon.2013.12.033.

Okolloh, O. (2009). *Ushahidi or 'testimony': Web 2.0 tools for crowdsourcing crisis information.* Participatory Learning and Action(59).

Orentlichter, D. (2016). Does human rights fact-finding need international guidelines? In P. Alston, & S. Knuckey (Eds.). *The transformation of human rights fact-finding* (pp. 499–523). Oxford: Oxford University Press.

Pagot, E., & Pesaresi, M. (2008). Systematic study of the urban postconflict change classification performance using spectral and structural features in a support vector machine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1*(2), 120–128. http://dx.doi.org/10.1109/JSTARS.2008.2001154.

Parcak, S., Gathings, D., Childs, C., Mumford, G., & Cline, E. (2016). Satellite evidence of archaeological site looting in Egypt: 2002-2013. *Antiquity, 90*(349), 188–205. http://dx.doi.org/10.15184/aqy.2016.1.

Poiani, T. H., dos Santos Rocha, R., Degrossi, L. C., & de Albuquerque, J. P. (2016, January). Potential of collaborative mapping for disaster relief: A case study of OpenStreetMap in the Nepal earthquake 2015. *Proceedings of the 49th Hawaii international conference on system sciences (HICSS)*Kauai, HI: IEEE. http://dx.doi.org/10.1109/hicss.2016.31.

Sulik, J. J., & Edwards, S. (2010). Feature extraction for Darfur: Geospa- tial applications in the documentation of human rights abuses. *International Journal of Remote Sensing, 31*(10), 2521–2533. http://dx.doi.org/10.1080/01431161003698369.

UN Commission of Inquiry on the Syrian Arab Republic (2016). *Out of sight, out of mind: Deaths in detention in the Syrian Arab Republic.* Geneva: Office of the United Nations High Commissioner for Human Rights.

United Nations Institute for Training and Research (2011). *UNOSAT brief 2011 - satellite applications for human security.* Geneva: United Nations Institute for Training and Research.

Witmer, F. D. W. (2015). Remote sensing of violent conflict: Eyes from above. *International Journal of Remote Sensing, 36*(9), 2326–2352. http://dx.doi.org/10.1080/01431161.2015.1035412.

Wolfinbarger, S. (2016). Remote sensing as a tool for human rights fact- finding. In P. Alston, & S. Knuckey (Eds.). *The transformation of human rights fact-finding* (pp. 463–477). Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof.

Wolfinbarger, S., & Wyndham, J. (2011). Remote visual evidence of displacement. *Forced Migration Review, 38*, 20–21 The Technology Issue.