Uncertainties of Human Perception in Visual Image Interpretation in Complex Urban Environments

Nicolas Johannes Kraff, Michael Wurm, and Hannes Taubenböck

***Abstract*—Today satellite images are mostly exploited automatically due to advances in image classification methods. Manual visual image interpretation (MVII), however, still plays a significant role e.g., to generate training data for machine-learning algorithms or for validation purposes. In certain urban environments, however, of e.g., highest densities and structural complexity, textural and spectral complications in overlapping roof-structures still demand the human interpreter if one aims to capture individual building structures. The cognitive perception and real-world experience are still inevitable. Against these backgrounds, this article aims at quantifying and interpreting the uncertainties of mapping rooftop footprints of such areas. We focus on the agreement among interpreters and which aspects of perception and elements of image interpretation affect mapping. Ten test persons digitized six complex built-up areas. Hereby, we receive quantitative information about spatial variables of buildings to systematically check the consistency and congruence of results. An additional questionnaire reveals qualitative information about obstacles. Generally, we find large differences among interpreters' mapping results and a high consistency of results for the same interpreter. We measure rising deviations correlate with a rising morphologic complexity. High degrees of individuality are expressed e.g., in time consumption, *in-situ*- or geographic information system (GIS)-precognition whereas data source mostly influences the mapping procedure. By this study, we aim to fill a gap as prior research using MVII often does not implement an uncertainty analysis or quantify mapping aberrations. We conclude that remote sensing studies should not only rely unquestioned on MVII for validation; furthermore, data and methods are needed to suspend uncertainty.**

***Index Terms*—Cognition, earth observation, elements of image interpretation, level of individual buildings (LoD-1), manual visual image interpretation (MVII), perception, uncertainty, urban morphology.**

Manuscript received January 16, 2020; revised June 19, 2020; accepted July 15, 2020. Date of publication July 23, 2020; date of current version August 5, 2020. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program under Grant 714087]- So2Sat which had no involvement in any occupations. *(Corresponding author: Nicolas Johannes Kraff.)*

Nicolas Johannes Kraff and Michael Wurm are with the Department of Geo-Risks and Civil Security, German Aerospace Center, Earth Observation Center, German Remote Sensing Data Center, 82234 Oberpfaffenhofen, Germany (e-mail: nicolas.kraff@dlr.de; michael.wurm@dlr.de).

Hannes Taubenböck is with the Department of Geo-Risks and Civil Security, German Aerospace Center, Earth Observation Center, German Remote Sensing Data Center, 82234 Oberpfaffenhofen, Germany, and also with the Department of Remote Sensing, Institute of Geography, University of Wuerzburg, 97074 Wuerzburg, Germany (e-mail: hannes.taubenboeck@dlr.de).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2020.3011543

I. Introduction

IN THE field of geoinformation, "uncertainty" can be understood as a human "intrinsic" feature that "exists in the whole process from geographical abstraction, data acquisition, and geo-processing to the use of data" [1]). Hence, uncertainty comprises the human perception as well as data source inaccuracies that influences geographic approaches, and therefore demands for an assessment.

In times of "big data" and new powerful methodologies for automatic image classification, the classic manual visual image interpretation (MVII) seems to become less important. However, for almost all quantitative performance evaluations, and especially to unveil the strength of machine learning algorithms, previous knowledge on spatial and thematic content of the data is of utmost importance. Unfortunately, this previous knowledge is often – if it even exists – not consistent, particularly for spatially highly resolved data and very specific types of land use. We find this gap specifically in complex urban environments, which are characterized by multifaceted built-up structures with changing land uses in very close vicinity. One example is the domain of "urban poverty mapping," which experiences increased attention from various scientific communities. Consistent labels related to built-up structures characteristic for poor urban areas in image data are only rarely available e.g., [2]. While from an initial perspective it seems obvious to visually localize characteristic slum structures in very high resolution (VHR) optical satellite images, we observed in previous studies significant challenges in detailed manual mapping of slums [3]. Therefore, in this article, we aim at quantifying data resulting from the subjective human perception. And we test whether perception is consistent among interpreters within a series of rooftop footprint mapping experiments. The setting is based in very complex urban structures characterizing poor urban areas, e.g., slums/informal settlements but also deprived formal areas.

In related studies of slum mapping using high resolution satellite images, latest image analysis methods from "artificial intelligence" (AI) have been adopted at very high mapping accuracies e.g., [4]. However, it is generally accepted in literature that MVII still offers highest classification accuracies [2], [5]–[10]. It is often applied for providing geo-information at a very high level of spatial and thematic detail [e.g., at the level of individual buildings (LoD-1)], which is an extremely challenging task in these complex urban landscapes [7]. It is argued that a human interpreter with professional knowledge is less susceptible to failures of building delineation, originating

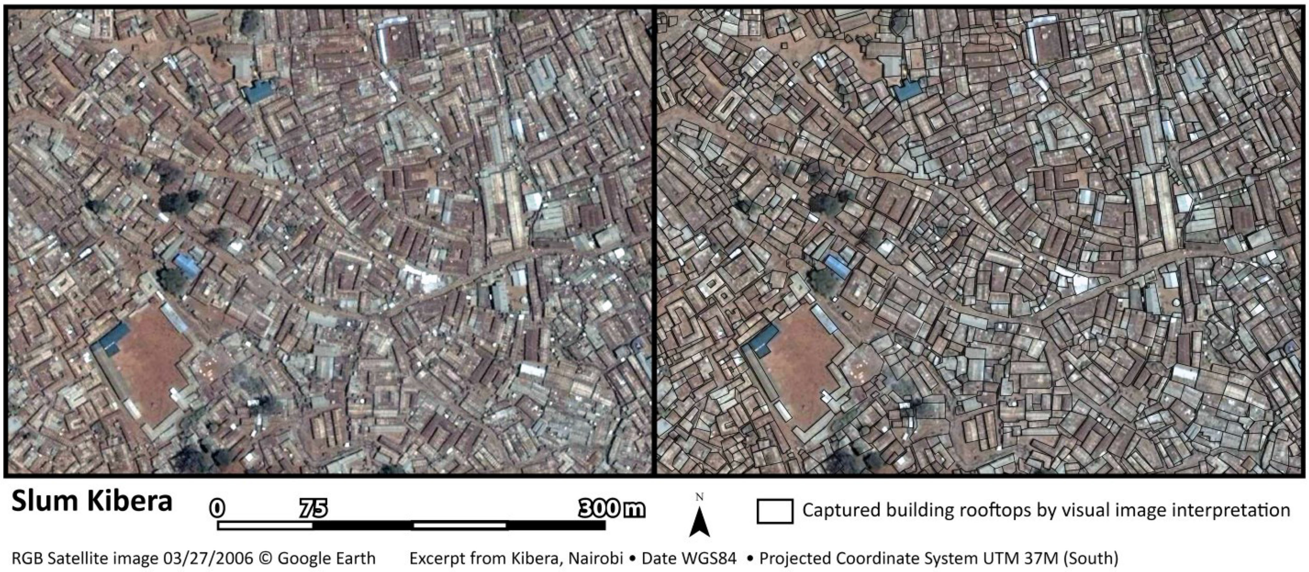

Fig. 1. Excerpt from Nairobi's largest slum Kibera. Captured building rooftops of a homogenous textural pattern by MVII, 03/27/2006; ©Google Earth 2019.

from e.g., homogenous colors and textures due to microstructures, than automated algorithms [8], [11]. Therefore, the classic method is still used extensively as reference to assess accuracies of automated processes e.g., [4], [12] or as labels for training classifiers [13]. At the same time, it is also argued that MVII is prone to inconsistent results among various interpreters due to for instance varying experience [14]. Fig. 1 demonstrates these challenges for a complex urban setting with roof structures of spectral and textural similarity in highest density.

Thus, poor urban areas often consist of very complex morphologic conditions and thus are comparatively difficult to map with regard to other, formal urban structures [10]. An interpreter's subjective way how to capture these structures have hardly been quantified yet. Uncertainty visualization has mainly been put into focus on an impact on decision making; yet, its application to constitute uncertain geodata is less explored [15]. Thus, recent methodologic developments on especially high detailed delineation and categorization of uncertain complex morphologic appearances of urban poverty e.g., [3], [16] need to be critically assessed by systematically documenting, quantifying, and visualizing uncertainties.

In this study, we set up two experiments to objectively analyze uncertainties in image interpretation: We test the consistency of and among interpreters. And, we test the influence of different complexities of urban morphologies on the mapping results. The research design is based on cross-sectional data: Six different complex urban settings (approximately 17 000 mapped individual buildings) are mapped by ten interpreters at LoD-1 based on a predefined ontology. Mapping outcomes for each of the interpreters are evaluated quantitatively based on five spatial variables. Following the quantitative mapping evaluation, we develop a structured survey with a questionnaire. It serves as qualitative analysis of the interpreters' subjective responses concerning the image interpretation process. Overall, we aim at answering the following research questions.

1) Do we observe uncertainties among the interpreters' mapping of complex urban areas?
2) Is the degree of uncertainty related to the complexity of the areas' morphologic structure?
3) Which aspects of perception and image interpretation elements specifically impact the MVII and thus, foster uncertainties?

The remainder of this article is organized as follows: Section II briefly reviews earth observation (EO) classification approaches followed by the state of the art of uncertainty applications in the geographical research field of complex urban areas. Additionally, the MVII process with its relevant key issues of perception is illuminated. Section III introduces the selected study areas (A), the technical aspects for mapping (B), as well as the spatial variables for analyzing and comparing digitized results (C). Furthermore, the questionnaire's structure (D) is presented. Section IV comprises the quantitative results among interpreters as well as the subjective opinions based on the questionnaire. In Section V, we discuss the results in the methodological context of EO and complex urban area research. Finally, Section VI concludes this article with an outlook.

II. State of the Art

Earth observation data have been largely deployed for mapping urban landscapes at various degrees of morphologic complexity. Especially morphologic appearances of poor urban areas, such as slums, are a severe challenge for image classification [12]. Contemporary automatic image classification approaches rely either on optical e.g., [17] or synthetic aperture radar (SAR) data e.g., [18], using either pixel based e.g., [19] or object-based-based methods e.g.,[20]. Other classifications rely e.g., on aggregated grids or streets blocks, e.g., [21], [22]. Recent advancements in AI and processing power allow for machine learning techniques such as "deep learning" for slum identification and delineation e.g., [4], [23], [24]. Most studies in current literature focus on classifications at district level [2], yet on single building level (LoD-1), there are only few studies - e.g., [16], [25]. MVII still plays a crucial role for the provision of reliable geodata, but these are currently rarely available, if they exist at all [10]. In 15 out of 87 related studies on slum mapping, MVII is still deployed [2]. However, only very few capture slums at the level of individual buildings, as e.g., [26] explicitly refer to the single building "object" and its environment. A morphologic categorization based on LoD-1 data is provided by [3] and in a follow-up, they describe the change over time [27]. However, although nowadays images reach resolutions in centimeter ranges, textural patterns, spectral differences and clarity of object transitions in complex urban environments remain challenging. Automated image analysis—even if as powerful as AI—are prone to uncertainties and classification errors, e.g., small paths between buildings, complex roof structures, or shadow casts that break the common patterns (cf. Fig. 1).

Due to a limited computerization of the real world, the knowledge delta between human, machine, and the real world automatically creates uncertainty. Thus, uncertainty analysis and -visualization in a geographic context is crucial. It affects MVII by cognitive entities (perception, memory, and thinking) as well as source-inaccuracy, -incompleteness, -inconsistency, and -imprecision [28]–[31]. However, analyses of uncertainty visualization rather focus on decision making than methods [15]. Thus, only few studies focus on the assessment of human-related uncertainties in geodata based on a quantitative methodology. Especially qualified feedback on human perception is mostly absent. According to [12] who measure significant variations between delineated slum boundaries among interpreters, variations of classification results are commonly more conducted in uncertainty studies of rural areas than in the urban context. Summarized appropriately [5]: *"There* are *difficulties in controlling quality over time and between interpreters*". And, as *"uncertainty may be of conceptual nature, and associated with the process of abstraction or generalization about real-world phenomena"* [1], especially uncertainty of the opposed precise reproduction of LoD-1 built-up environment is difficult to capture [9], especially for an entire slum.

A. Issues of the Image Interpretation Process

As MVII is generally influenced by subjective perception we briefly recapitulate established key issues for a better understanding of the relevant human-machine interaction. There exist different ways to design the interpretation process: Following [32], there is a hierarchical, *primary ordering* of analysis with basic elements as tone/color and complex elements as site and association. According to [33], the key issues of the interpretation process are the *criteria of classification* (e.g., residential areas) as well as the minimum mapping unit (MMU) (cp. also [34]). In comparison, [35] offers a very simplified scheme and uses *recognition* and *interpretation* as key issues that are hardly separable and affecting each other in an iterative process. This process is influenced by the interpreter's experience and precognition that leads to the final interpretation result.

TABLE I
ORDER, DATE OF RECORD, DIMENSIONS, AND COMPLEXITY OF SELECTED SUBSETS OF DOCUMENTED COMPLEX URBAN AREAS

Subsets of selected urban areas	Mapping order	Morphologic classification	Complexity	Date of sensor's record	Subset's dimensio n (ha)
Agios Panteleimonas, Athens, Europe	1	mixed neighborhood	less complex	14.03.2008	17
Imbabah, Cairo, Africa	2	slum-like morphology	less complex	29.04.2015	8
Tei Toboc, Bucharest, Europe	3	mixed neighborhood	less complesx	25.06.2005	4
Santosh Nagar (Goregaon East), Mumbai, Asia	4	morphologic slum	highly complex	12.10.2003	2
Kibera, Nairobi, Africa	5	morphologic slum	highly complex	27.03.2006	6
Makoko, Lagos, Africa	6	morphologic slum	highly complex	17.01.2015	1

Based on this, we emphasize the following selected issues as coherent for this study: *Recognition* includes the physical and psychological visual perception, which finds particular applicability in "morphologically complex areas" as the *criterion of classification.* Based on an interpreter's subjective perception, a "laminar classification" of contrasts, edges, lines, colors, and brightness is linked to a strong choice of impulses. With it, equal laminar areas (e.g., densely urban) are merged. Yet, the interpreter is able to differentiate and extract single elements out of an area with homogenous textures [35]. In our case these are LoD-1 buildings as *MMU*, in high density built-up environments. Thereby, visual perception offers central advantages as "tolerance" and "amodal completion" help to identify elements on the strength of habit and experience where incompleteness (e.g., treetop covering part of a rooftop) is psychologically removed [36].

Furthermore, the interpreter's cognitive perception allows using visual *image interpretation elements* such as size, height, shadows, patterns, etc. [33]. In combination with experience, this is fundamental to draw vertices correctly and thus delineate single buildings precisely. Uncertainties might originate during the procedure of setting vertices as well as from deviations in results among human interpreters. Hence, an exploration of these interpretation elements affecting mapping is necessary [37]. Thereby, uncertainties can reveal first steps for later adoptions for machine learning or for an evaluation when used for validation purposes.

III. DATA AND METHODOLOGY

A. Selection of Study Sites

We chose six urban areas with a different character of morphological complexity as test sites (Table I). The selection is based on the following criteria: Each site contains a minimum of 300 buildings for digitization. Depending on this, we chose comparable extents resulting in not more than a maximum dimension of 17 ha. With it, we estimate work duration of 8–12 h as frame for the mapping exercise. Aiming at a high variety of morphologic appearances, we selected areas from three continents. The sites are morphologically representative

TABLE II
DESCRIPTIVE STATISTICS ACROSS ALL AREAS FOR EACH INTERPRETER AND CV AMONG ALL INTERPRETERS

		Interpreter													
Spatial variable	**Approach**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**mean**	**median**	**stdv.**	**CV (%)**
1.No. of buildings	total of all areas	2656	2531	1293	1231	1475	1437	1715	1670	1461	1496	1696.50	1485.50	470.07	**27.71**
2.Building size (m²)	mean of all areas	86.84	109.92	72.05	130.52	132.05	122.51	102.79	85.61	140.49	130.53	111.33	116.22	22.41	**20.12**
3.Building orientation (°)		40.95	40.66	44.63	46.15	40.14	46.01	43.41	46.21	43.37	43.04	43.46	43.39	2.19	**5.04**
4.Building density (%)	related to all areas	70.62	47.49	55.90	49.23	59.67	53.94	54.01	43.54	62.89	59.83	55.71	54.95	7.55	**13.55**

"subsets" for the entire districts, containing their characteristic structures, as for instance dense organic structures, complex alignments, and different types of buildings or open spaces and organized street networks (Fig. 2).

The sites show a documented variety in structural complexity [3], (cf. Section III.B) and have been related to poverty in literature.

B. Technical Aspects for the Mapping Exercise

For the process of digitization, we offered a short guideline with instructions on the workflow with the building ontology to all interpreters (Appendix 1). Following the MVII, ten interpreters produced a classification based on VHR optical satellite data (Quickbird, WorldView with a geometric resolution of up to 0.46 m [pan sharpened)]. Every single building is represented by its rooftop, which is used as proxy information for the buildings' ground area. Hereby, one polygon consists of at least 4 vertices. The level of image scale was left as an individual choice to each interpreter.

The analysis is applied with an equally distributed level of difficulty where every interpreter starts with three comparatively simple areas and continuously progresses toward three more complex ones, as we do not know whether all interpreters worked with a GIS before. For this reason, we also handed out a digitization guideline (cf. Appendix 1). In this vein, we continuously progress the level of difficulty, similarly to the funnel questioning, that we also used in the questionnaire (cf. Section III.D), in order to avoid an early loss of concentration. We set difficulty and complexity in relation and define a high complexity of an area being equivalent to a high number of buildings and high densities, an uneven size distribution and irregular orientations of buildings (cf. Table I). Thereby, we rely on the analysis done by [3], who classify these complex morphologies, ranging from "morphologic slum" to "structured neighborhood". Thus, we evaluate the complexity of an area as "highly complex", if the classification is close to the category "morphologic slum" and define complexity "less complex" for mixed and structured neighborhoods. We expect less complex areas to be digitized easier, faster and with a more uniform result.

Finally, the interpreters also provide qualitative assessments by a questionnaire based on 33 questions. For an unbiased approach, we choose students of different study fields (geography, engineering, computer sciences, and didactics) with and without precognition in terms of these complex morphologic sites, *in-situ* inspections or GIS applications. The interpreters worked in a controlled environment. We did not limit time for the digitization.

C. Statistical Analysis of Geometric Uncertainties

To assess the deviation among interpreters, we apply descriptive statistics across all areas for each interpreter and present a table again with *mean*, *median*, *standard deviation,* and the *coefficient of variation (CV) among all interpreters* (cf. Table II). With it, we first respond to *research question 1* (uncertainties among interpreters?) at an aggregated level as an interpreter-oriented (A) approach. In a second step, we evaluate *research question 2* (uncertainties related to complexity?) by comparing classification results between areas in relation to the area's complexity - area-oriented (B): We retrieve values specifically for each area and present the results among interpreters. Furthermore, we use boxplots for visualization of deviations and agreements among interpreters. As geometric identifiers, we systematically apply the following *five spatial variables*:

1) *Number of buildings:* To assess recognition of a building as a whole entity, we count (quantity) the digitized polygons of each interpreter and each study site, in order to compare total differences.
2) *Building size:* We calculate (sum, mean, and CV of) the size (m²) of each interpreter's digitized polygons to assess the digitisation manner, that is, a tolerant verus tight-fitting way of setting vertices to form a polygon.
3) *Building orientation:* We calculate (sum, mean, and CV of) interpreter's digitized building orientations. It is based on a polygon's longitudinal side to the cardinal direction north, where an absolute value equals its turn ($-90° = +90°$). In this vein, we receive mean information about precision in placement/alignment of vertices.
4) *Building density:* As aggregation from *number* and *size* of buildings, we calculate the density as sum of each interpreter's buildings (m²) in relation to the entire area(s) as reference unit (%). Afterwards, we calculate mean and

Fig. 2. Overview of the selected study sites for the mapping exercises; ©Google Earth 2019.

TABLE III
+ LEGEND DEVIATION OF MAPPING AMONG INTERPRETERS FOR EACH AREA IN RELATION TO THEIR COMPLEXITY AND COMPLICACY TO MAP, SCALED FOR CV AND FLEISS-KAPPA INDEPENDENTLY

		Complicacy>	simple	simple	simple	difficult	difficult	difficult
Spatial variable	**Approach**		Athens	Cairo	Bucharest	Mumbai	Nairobi	Lagos
1.No. of buildings	CV of mean of all interpreters' means/area	100 - CV = agreement (%)	74.13	81.39	27.97	52.49	64.22	63.46
2.Building size			81.83	84.25	74.63	56.99	71.67	70.19
3.Building orientation			89.30	91.84	87.29	92.97	95.82	82.20
4.Building density			84.31	91.47	59.28	76.44	83.32	65.66
5.Geometric polygon-matching	Fleiss-Kappa [0;1]		0.59	0.42	0.49	0.34	0.43	0.36

Strength of agreement	according to Fleiss-Kappa [0:1]	using coefficient of variation (%)
slight	0.00 - 0.20	<20 - 20
fair	0.21 - 0.40	21 - 40
moderate	0.41 - 0.60	41 - 60
substantial	0.61 - 0.80	61 - 80
almost perfect	0.81 - 1.00	81 - 100

CV among interpreters. Especially density allows demonstrating secondary effects about individual working habits and digitization continuity.

5) *Georeferenced building matching* (only B): We focus on the "interrater reliability" for the accuracy assessment of the precise geometric matching (rooftop placement) of polygons rather than an exact reproduction of the reality. Additional to the previous mathematical approach, we add this to understand the exact topological diversity of mapping and visualize the LoD-1 spatial arrangement. Thus, a background-grid of 2×2 m cells is created for each area. We assume 4 m² per cell to be precise, because it fits with our MMU of 4 m² being the smallest mapped building. We spatially join all ten interpreter's datasets with the grid and whenever a polygon hits or intersects one cell we count it as a match. Afterwards, we compare the counted matches of all cells and apply the measurement of observer agreement for categorical data with the *Fleiss-Kappa* index [38] for all areas. The interval scale starts from value 0 that represents a "poor" agreement to value 1 as an "almost perfect" agreement among the interpreters (cf. Table III).

D. Questionnaire

To respond to *research question 3* (on the perceptional impact on MVII), we set up a questionnaire (cf. Appendix 2). It is constructed in a "structured" way with "closed questions" being succinct and easily comprehensive. Furthermore, we follow the guidelines for empiric surveys [39]. The questionnaire is clustered in 5 thematic blocks, starts with easy questions and progresses to more difficult ones to keep the attention and elicit more details (funnel questioning). It ends with personalities.

1) Difficulty, work duration, GIS knowledge and *in-situ* expertise.
2) Elements of image interpretation (e.g., data source: pattern, shadow, clouds, …).
3) Time-related interpreter's subjective self-evaluation for GIS functions, recognition, and fatigue.
4) Used scale during digitization.
5) Personal questions (name, gender, age).

Blocks 2–4 are used to explore the perceptional influence affecting digitization as we put the mentioned elements in relation: We want to know, whether the interpreter's perception (recognition, interpretation) during digitizing polygons via "parameters" *position, shape, size, orientation,* and *quantity* is affected by the *elements of image interpretation*. And we want to know if this is congruent among interpreters (cf. Appendix 3). Due to the low basic population of only 10 respondents, we omitted a pretest of the survey. Responses were qualitatively interpreted and summarized. The survey was conducted after the digitization to avoid test persons to be influenced ex-ante. In case of incomprehensibilities to single questions we provided explanations as study supervisors. Subsequently, we appealed to all test persons to provide open, additional comments.

IV. RESULTS

A. Quantitative Results

In the first analysis, we following:

1) **Interpreter-oriented approach:** We distinguish the individual interpreters' mapping results across all areas. For the classified *number of buildings*, we find a significant coefficient of variations (CV) of 27.7% and with respect to *building size* we also find a substantial CV of 20% among the interpreters (Table II). Thus, some interpreters map buildings in a larger or smaller manner than others do. Comparing all variables, the lowest deviation is illustrated by the building orientation. A CV of only 5% shows a high concordance among interpreters by means of placed vertices allocation.

Altogether, there are significant differences in digitization results among interpreters. The varying numbers of buildings and the differences in morphologic characteristics lead to error propagation. As a result, from these spatial variables' aberration, also *building density* reveals a significant difference of 13.5% among interpreters. Thus, distorting variances in geographic interpretations of the built environment could occur, depending on the interpreter. With it, we affirm with respect to research question 1 that there are large differences among interpreters.

2) **Area-oriented approach:** In such areas of the study, where we defined the physical morphology to be less complex (Table III) we find less deviations across the interpreters' mapping

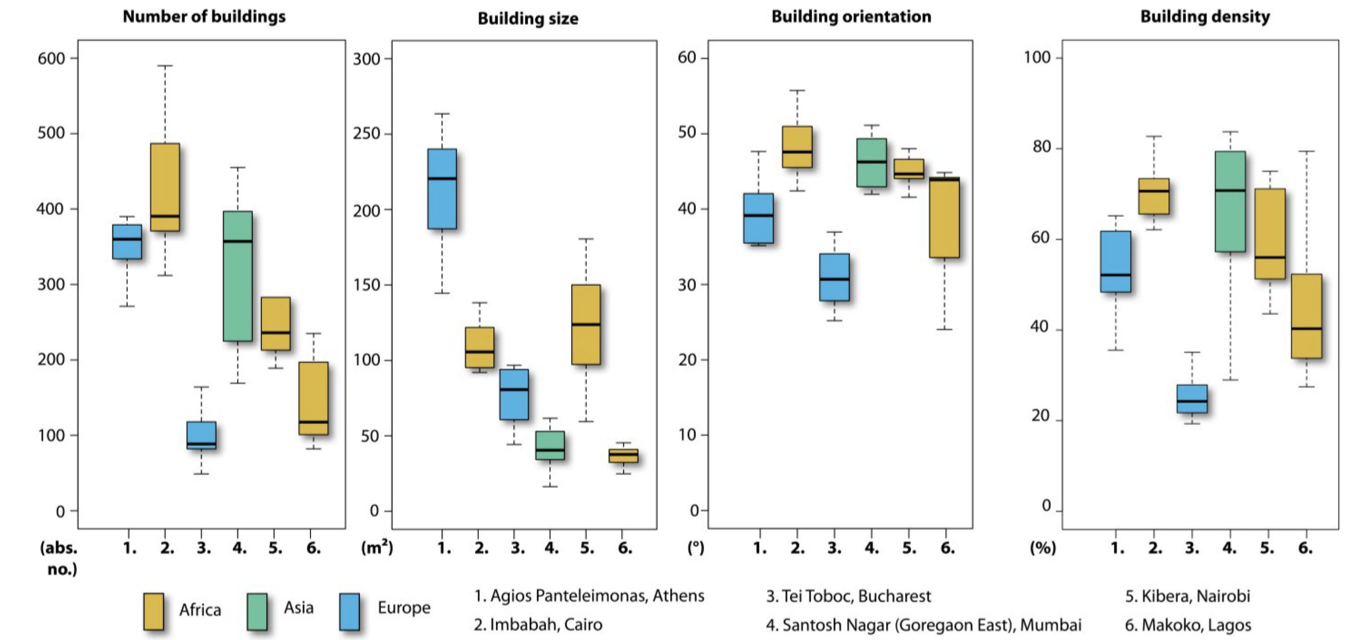

Fig. 3. Range of deviations among ten interpreters for each area.

results. Reciprocally, we find lower agreements in more complex areas. Examples are s Santosh Nagar (52.4% for *number of buildings)* or the palafitte settlement Makoko (82.2% for *orientations*). We assert a clearly aligned intraurban street network and rather larger buildings (Fig. 2) visually to be mapped easier.

Apart from this trend we discover two antithetic aspects: First, complex slum-like areas usually contain rather small *building sizes.* This is exemplified by the interpreters' work in the complex areas Mumbai and Lagos (cf. Fig. 3 *building size*). Due to high complexity in these areas, we assumed higher deviations among the interpreters, instead we find very high agreements. Second, our site in Bucharest shows comparably low agreements among interpreters (cf. Table III) even though registered as an area with a rather simple morphologic structure.

Generally, the range of variations among interpreters is high, especially for the *number of buildings* (Fig. 3). Compared to the aforementioned interpreter-oriented approach, also the variables *building orientation* and *density* reveal significantly higher variations. Furthermore, we find that variations are generally independent from the geographic location, even though we recognize certain geographic "tendencies" as the *number of buildings* varies most in the Asian areas or *building sizes* in the European ones. However, our limited basic population of only six areas does not allow drawing a valid representative continental wise conclusion.

With respect to the spatial variable *matching of geometries*, the Fleiss-Kappa index indicates poor agreements (Table III) among interpreters. The highest value reveals a "moderate" agreement (0.59) for the area of Athens. Thus, surprising or not, a near to perfect spatial match among interpreters is not given, not even for morphologically less complex areas. Fig. 4 illustrates the degree of cell matching and visualizes the spatial distribution of uncertainty: As for the mapped *building sizes*, we generally discover more spatial matches per cell where buildings adjoin each other. This is not only due to a high concentration of multiple building's edges on single cells. Thus, we find comparatively fewer matches at buildings' untouched margins, e.g., close to open spaces. At this location, we find a relatively higher disagreement among interpreters as for instance up to three interpreters only map one cell.

As aforementioned, we figured out that there are large deviations among interpreters on the variable *density* across the areas. It ranges from 59% agreement in Bucharest to 91% in Cairo (Table III). However, despite this variance, *density* reveals a relative consistency in the digitization results as well: We see interpreters usually digitize in a consistent manner across all areas (Fig. 5). For instance, interpreter 8 always maps lower densities for each area in comparison to other interpreters. We find similar results for *size* and *orientation*. Only variable *number of buildings* shows more inconsistent behavior. Hence, mapped data *within* one interpreter's dataset are observed consistent. Additionally, this consistency is independent from time, as the sequence of digitized areas does not seem to influence this result.

B. *Qualitative Results*

As introduced in Section III, the questionnaire follows four thematic blocks:

1) The (self-)evaluation plays an important role in the context of perception. Generally, six interpreters found the task, medium' difficult. Furthermore, most interpreters imagine the methodology "MVII" to be more accurate than automatized processes. Continuing with the topic precognition, half of the interpreters stated to have seen areas of urban poverty and have experience in these complex areas. Most

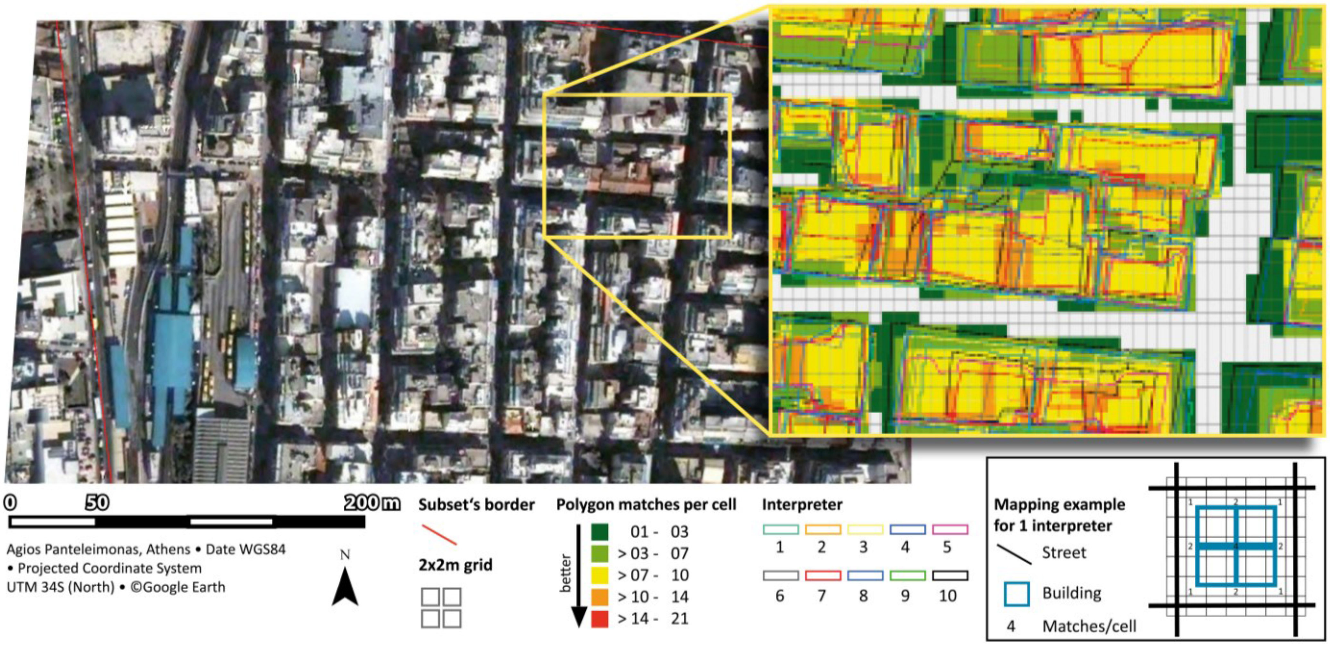

Fig. 4. Geometric matching of polygons per cell by interpreters' placed vertices at the example of Athens.

interpreters have worked with a GIS before. These findings play an important role especially in the context of the following aspects "*elements of image interpretation and time"*.

2) *Elements of image interpretation* are well known among the interpreters. With a high agreement we found interpreters' mapping being influenced by a couple of such different elements: Especially the "data source" resp. "image quality", as e.g., "atmospheric conditions", "brightness," "stereoscopic effects," "cast shadows," "contrast," "mixed pixels," and "geometric resolution" were stated to influence most, also being the major obstacle causing problems during digitization.
3) *Time and self-evaluation:* In contradiction to our estimated work duration of at least 8 h (cf. III.A), the analysis reveals the demanded time per interpreter with a maximum of 11 h, a minimum of 04:40 h, and an average of 7:47 h (Fig. 6). Thus, among the interpreters we find a remarkable disparity of claimed time. However, for each area the interpreters needed comparatively similar capacities of time. We found no relation between complexity of the area and the required time for digitization.

Obviously, GIS functions can improve working speed, depending on the tools, as snapping, auto complete, and cut polygons were used by most interpreters. Furthermore, interpreters 1, 5, 8, and 9 stated to be experienced, but this did not lead to less working duration. Plus, regarding speed, six interpreters confirmed to pay less attention for details after a couple of hours of digitizing. We assume that working duration and the rising difficulty both have an influence on fatigue. This explains the less spent time for the last area (Lagos) to be digitized. Fewer buildings were captured than we can find in reality. Thus, the time duration should have been higher for mapping Lagos. Furthermore, it was confirmed that classification took longer in case of precognition because it influenced their way of image interpretation. Also, two interpreters had seen some of the selected areas themselves before and were able to even remember buildings or vegetation. As another aspect, nearly all interpreters estimate their own level of details as "medium," stating this fact to also influence the mapping. In general, we find a large variety of factors influencing the mapping process on individual level. This reveals that even if the same introduction is given to the interpreters, the starting positions and individual experience and perception differs significantly.

4) *Scale:* All interpreters used a flexible scale with zoom function that also influenced the mapping, where more than half of them state to perceive the topological relativity between objects in a better way.

As response to the request for additional comments, the following remarks were given.

1) *Outlines:* Despite clearly announced boundaries, it was unclear whether industrial areas respectively the railway building should be mapped and whether a building's courtyard had to be part of a polygon or to be cut.
2) *LoD-1 delineation:* Interpreters had problems to identify balconies, sticking out from the buildings in Agios Panteleimonas due to a satellite nadir offset (cf. Fig. 7). The fundamental question arose, if balconies are part of the buildings ground footprint and whether to be mapped or not. Also, it was stated very often to be difficult to distinct a tiny footpath, roof ridge, or building shadow in highly dense areas.
3) *Image quality:* Some interpreters remarked a bad image quality. For Cairo there were difficulties to visually differentiate buildings from the ground/soil; similarly, delineation between buildings was said to be difficult in Santosh

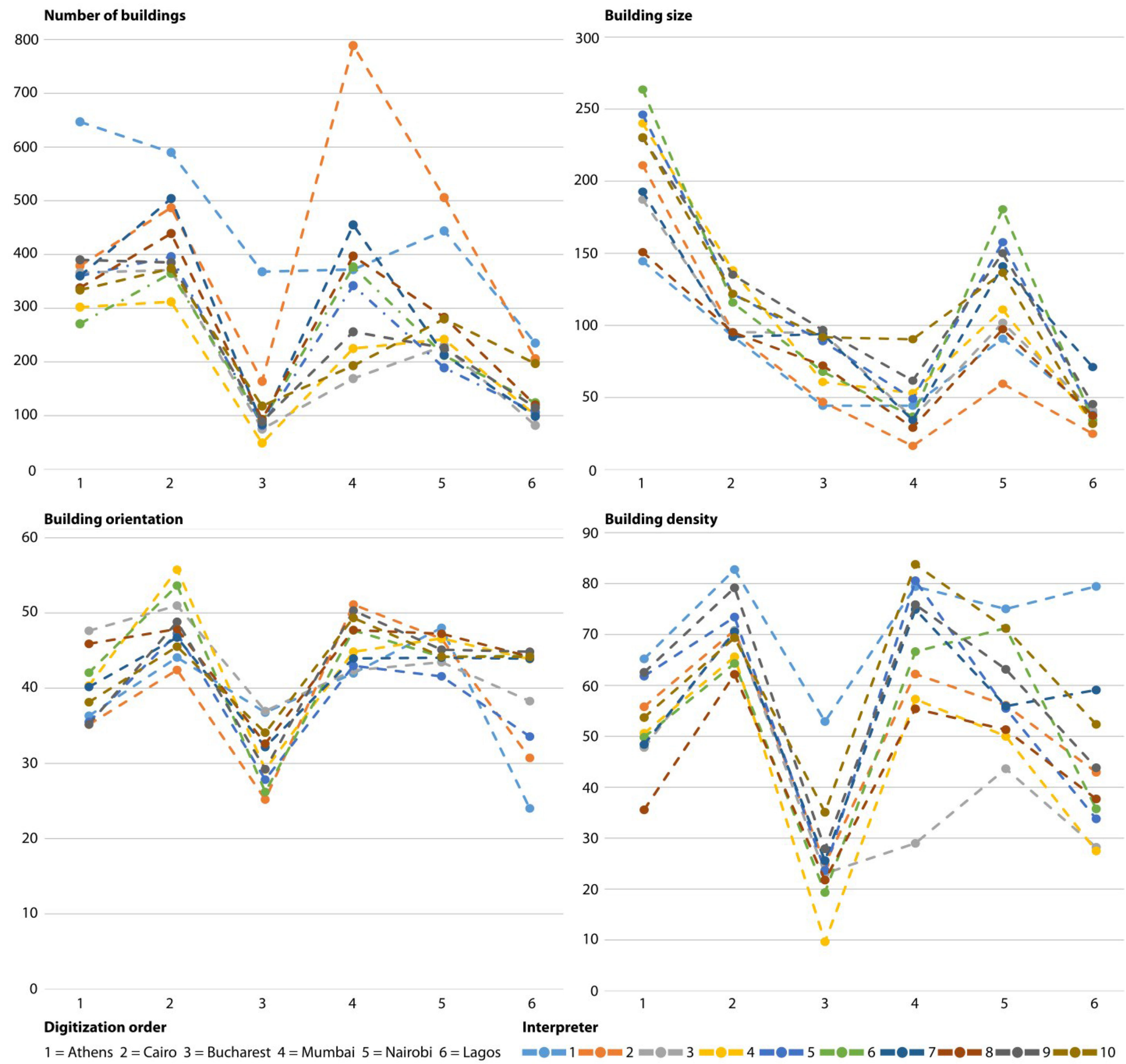

Fig. 5. Relativity of density values among interpreters per consecutively mapped area.

Nagar and images of Kibera and Tei Toboc contained white unnatural dots.

4) *Health:* It was reported that interpreters lost concentration over time.

With respect to research question 3, we conclude that interpreters highly agreed that digitization "parameters" *shape* and *size* being mostly affected by the *elements of image interpretation* such as particularly "data source" and "image quality". Furthermore, interpreter-related factors as individual *in-situ*- and GIS- knowledge, fatigue, and working manner - including duration and utilization of zoom - expressively affect the MVII.

V. Discussion and Interpretation

MVII is often used as a reference for automatic processes and for validation since it is still being seen as a methodology resulting in highest accuracies. First, we presented MVII's key issues, namely *classification*, *MMU* and *perception*. Second, we revealed that their application in the context of complex urban areas does not allow an easy recognition/interpretation that contains diverse but also smallest building units. Finally, we demonstrate deviations among human interpreters that demand for an uncertainty analysis. However, this empirical uncertainty study does not compare humans with machines; hence, we are not able to assess which method comprises less uncertainty. Instead, it is our aim to assess inter-rater results and with it sharpen the classic methodology. Our experiment relies on six areas across the globe and we evaluate results of ten interpreters. Yet, our sample is comparatively small and we cannot claim the results to represent all such complex areas. Furthermore, a higher quantity of test persons might reassure our findings. Yet, our dataset builds a reliable basic population for the spatial quantitative analysis.

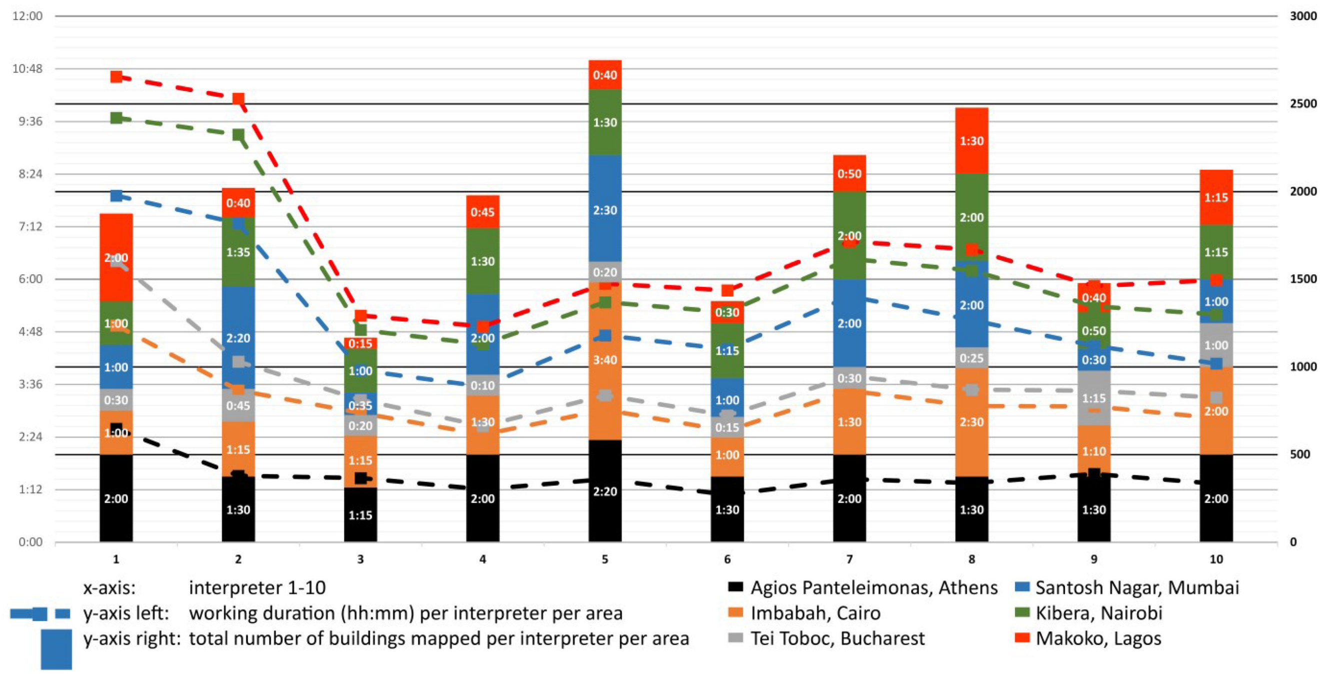

Fig. 6. Duration of mapping in relation to number of buildings, for each dataset and interpreter.

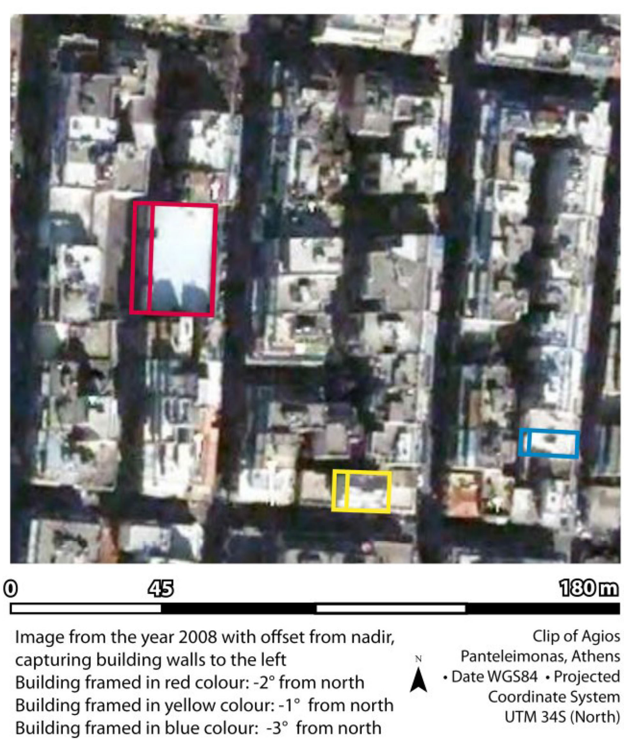

Fig. 7. Agios Panteleimonas, Athens: Satellite nadir offsets leading to different building angles and uncertainties handling balconies.

With it, this article clearly reveals substantial deviations among interpreters. We find rising variations in relation to a rising morphologic complexity of the areas. We find that also *image interpretation elements* as "key issue" of perception have an impact on the digitization. We illuminate that personal knowledge also affects the interpretation process and its duration. This study reveals that in literature mentioned "high accuracies of MVII" [2], [5]–[10] are not achieved in the sense that interpreters' agreements are low for our selected complex structural areas. As automatic classifications being able to capture individual buildings in slum-like structures in very high quality, we cannot evaluate whether MVII is still more accurate. However, we still see independent from the accuracy that interpreters can generally distinguish shacks from each other even in very complex areas of spectrally homogeneity.

However, interpreters clearly *recognize* or *interpret* built-up structures differently and what seems to be one rooftop for one person might be none or several rooftops for another person. It is rather questioned whether there is a building and how big it is, than how it is aligned. Questioning this, we also follow the uncertainty criteria for objects, as proclaimed by [31] who demonstrate "existential," "extensional" and "geometric" uncertainty. Most interpreters stated to have GIS knowledge and more than half of them indicated to have at least some *in-situ* experience. However, results reveal tremendous disparities, as for example for the digitized number of buildings (CV of 27.2%). To draw a more detailed picture of the spatial variables and to better visualize uncertainty, we set the interpreter's results in relation to the area's morphologic complexity, as categorized by [3]. This area-oriented approach reveals high agreements with small variations among interpreters in morphologic areas categorized as "simple". This fact significantly changes with a rising morphologic complexity to higher disagreements and variations; e.g. buildings with low distances to neighboring buildings in complex areas are more difficult to capture. Our findings are in line with established findings about digitization challenges:

The interpreters experience typical disturbances from image sources, such as radial offsets with building-lean effect (Fig. 7), misleading stereoscopic, mirroring, and error effects with displacement and overlapping of buildings [cf. 40]. Particularly the latter effect was experienced by the interpreters and also the aforementioned "amodal completion" found application, as we recognized polygons underneath mirroring white shapes for instance in Tei Toboc (Fig. 2,). This might be one reason for the low agreement among interpreters in that area. However, being the only area offering heterogeneous textures with a significant vegetation it should be easier to map.

We constitute many of the mentioned results to be caused by the areas' homogenous spectral signatures. It embodied the main landscape class with homogenously colored and textured elements and mixed pixels where single rooftops are very difficult to capture or, especially to delineate (Fig. 1). This is affirmed by e.g., [12] and [41] stating that *"there is an influence of ground sample distance (GSD) respectively the image quality on object identification"*. Interpreters collectively agree that *image interpretation elements* depending on the image source quality (like atmospheric conditions, brightness, contrasts, etc. and especially the geometric resolution) mostly influence the "digitization parameters" *shape* and *size*, followed by their *quantity*.

However, next to our findings about geometric deviations we also find a substantial consistency when comparing all interpreters by using the example of *building density* (Fig. 5). In fact, we discovered highly varying density values *among* interpreters (Table III). However, at the same time we find that one interpreter tends to be comparatively consistent across areas; i.e., if one interpreter digitizes an area with less density than others, it is highly likely that he/she will do so for other areas, too. *Size* and *orientation* also demonstrated consistency. This finding fundamentally extenuates the discovered uncertainties.

So far, we experienced complex morphologic situations challenging the human interpreter's cognition more than expected. We know there is a significant uncertainty among interpreters but what about the self-evaluation and personal condition? As human being, one interpreter might have a different perspective with his/her own *recognition* and *interpretation* during the MVII process than another. Interpreters used the zoom function in a flexible way to recognize the built-up structure and thereby confirmed an improved understanding of the topological context. With it, they also confirm prior findings who declared that there is no single perfect scale as many scales flexibly depend on specific image objects [42]. Considering precognition, we contrast experienced with unskilled user's results. We find that precognition rather delays working speed. This is in contrast to the findings from [12], where experienced interpreter's results do not deviate from others without experience. Abbreviated durations by reasons of fatigue ("vigilance") were only partially confirmed by the interpreters. However, the self-estimation differs to the actual produced results because more polygons were expected in the final, complex (Lagos) data set. It fits into the picture that self-evaluation for difficulty and level of detail was mostly quoted "medium". In the end, we discover very different working times (cf. Fig. 6) needed by the interpreters as well as trends of correlation between time consumption and polygon creation. Very similar results were proven by [14] in terms of accuracies (72 to 81%), vigilance, self-evaluation (medium), and duration.

Other studies, yet not always applied for LoD-1 or urban areas, also assessed the MVII and likewise measure inconsistencies among users, e.g., [43] manually delineating slum areas. As another example, [44] find a low accuracy (60%) reached among experts when extracting building plots for cadastral mapping. However, in their experiments human results were still better than those by machines. Another analysis was conducted in the frame of European agricultural subsidies by the Joint Research Centre (JRC) [45] to evaluate interpreter's digitization for EU's "land-parcel identification system". Classified agricultural areas fundamentally differ as the outcome revealed deviations *within* one interpreter only up to 1.7% and *among* all persons only up to 3.8%, which is significantly low in comparison to our complex urban areas but is in line with our findings with respect to consistency *within* interpreter's *density* data. The JRC study furthermore postulates that polygon-vertices, placed at the corner of an agricultural parcel/polygon, crucially influence boundaries and measured ranges. We experience similar findings with our analysis of the geometric matching, where corners are crucial. In the following we explain why: [41] remark a 2 m GSD to be necessary for capturing buildings in an urban environment, yet [46] state 0.5m to be appropriate. We chose the geometric matching of placed vertices on a 2×2m cell grid that fits with the MMU of a 4 m² building and is based on satellite images of $\sim$0.5m GSD. Applying Fleiss-Kappa to receive an exact interrater reliability, we find moderate matchings for simple morphologic areas and merely sufficient values for complex areas. Thus, even based on VHR images, interpreters still vary significantly when placing vertices into different cells and corners with sometimes only one hit per cell more than 2 m away from another (cf. Fig. 4). Under consideration of an even higher image quality (8 cm GSD), [14] measured a more precise deviation ($\sim$0.4–0.5m) among interpreters for urban road networks and lamp posts. Hence, in comparison to our study, a very accurate polygon-matching among interpreters (e.g., Fleiss-Kappa $>$0.8), at least for simple morphologic areas and under usage of VHR images, is absent and thus we find high uncertainty for MVII.

To conclude: Reference data are difficult to capture in VHR satellite data for such complex urban landscapes. Even MVII is not providing ground truth, but multiple subjective interpretations significantly deviating from each other. This surprising result brings doubts to many studies conducting accuracy assessment by MVII, for instance, initiatives aiming the global mapping of buildings in urban areas, e.g., [4], [24], [47], [48] that require training data.

VI. Conclusion and Outlook

In this applied empirical uncertainty study, we explore the similarities and deviations among human interpreters in classifying complex urban areas and find large variations among interpreters. Although, different EO approaches are widely applied, human performance assessment has been hardly investigated

[12], [14]. A missing *"comprehensive understanding of the parameters that influence successful uncertainty visualization"* is postulated by [30], whereas uncertainty is understood as an unknown inaccuracy. In this study, we use a methodological mixture of mapped cross-sectional data and multifaceted responses from interpreters. With it, we set a range of parameters to reveal and visualize uncharted and significant uncertainties by demonstrating inaccuracies related to the classic methodology. We find large deviations among interpreters and in relation to a rising morphologic complexity. With respect to the mapping procedure we observe a high degree of individuality, yet a personal consistency.

After having assessed and interpreted the results, we draw the conclusion that the MVII has to be challenged being the most accurate methodology for complex urban areas at the single building level. Our findings about uncertainty in the MVII, considering deviations but also consistency, must be taken into account in future, because it can influence machine learning approaches significantly. Furthermore, EO research has to consider this uncertainty also by finding new methods for a better approximation to a "ground truth". This concerns further studies about cognitive perception [49] as well as LoD-1 machine detection. A following task would reach beyond the sole spatial analysis and explore potential variety over time, for instance, by a multitemporal panel study. Will the mapping "behavior" of a human being remain constant when placed in a multitemporal experiment?

Finally, one interpreter stated that *"visual interpretation can often be more precise than automatic processing but this depends on the experience of the interpreter, his/her knowledge with the influence of the subjective perception, the image quality and other factors."* The MVII as a methodology still plays a key role in complex urban morphologies, may it be for mapping, validation purposes or to generate training data for machine-learning algorithms; and cognitive advantages of the brain are demonstrably not to be underestimated. Nevertheless, this article shows that the seemingly simple process of visual mapping in optical satellite data is obviously much more difficult than expected. This leaves us with the final remark that many validation approaches in remote sensing studies must be fundamentally questioned.

Appendix

1) Digitization guideline for interpreters.
2) Questionnaire.
3) Table with elements of image interpretation and digitization parameters.

Acknowledgment

The authors would like to thank S. Bauer, A. Droin, L. Gorzakowski, J. Kauffert, B. D. Kraff, S. Nolting, I. E. Pistopoulou, T. Spiess, and M. Vogler as interpreters for data digitization and responses to the questionnaire and also would like to thank B. Peil for the mathematical support regarding Fleiss-Kappa calculation.

References

[1] J. Zhang and M. F. Goodchild. *Uncertainty in Geographical Information*. Boca Raton, FL, USA: CRC Press, 2002.

[2] M. Kuffer, K. Pfeffer, and R. Sliuzas, "Slums from space—15 years of slum mapping using remote sensing," *Remote Sens.*, vol. 8, no. 6, pp. 455–484, 2016.

[3] H. Taubenböck, N. J. Kraff, and M. Wurm, "The morphology of the arrival city-A global categorization based on literature surveys and remotely sensed data," *Appl. Geography*, vol. 92, pp. 150–167, 2018.

[4] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 150, pp. 59–69, 2019.

[5] R. Sliuzas, G. Mboup, and A. de Sherbinin, "Report of the expert group meeting on slum identification and mapping report by CIESIN, UN-HABITAT," Int. Inst. for Geo-Inf. Sci. and Earth Observ., Enschede, The Netherlands, 2008.

[6] I. Baud, M. Kuffer, K. Pfeffer, R. V. Sliuzas, and S. Karuppannan, "Understanding heterogeneity in metropolitan India: The added value of remote sensing data for analyzing sub-standard residential areas," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 12, no. 5, pp. 359–374, 2010.

[7] R. Mahabir, A. Croitoru, A. Crooks, P. Agouris, and A. Stefanidis, "A critical review of high and very high-resolution remote sensing approaches for detecting and mapping Slums: Trends, challenges and emerging opportunities," *Urban Sci.*, vol. 2, no. 1, pp. 8–46, 2018.

[8] T. Veljanovski, U. Kanjir, P. Pehani, K. Oštir, and P. Kovačič, "Object-based image analysis of VHR satellite imagery for population estimation in informal settlement Kibera-Nairobi, Kenya," *Remote Sens. Appl.*, 2012, pp. 407–434. [Online]. Available: https://www.intechopen.com/books/remote-sensing-applications

[9] J. Yuan and A. M. Cheriyadat, "Learning to count buildings in diverse aerial scenes," in *Proc. 22nd ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2014, pp. 271–280.

[10] M. Wurm and H. Taubenböck, "Detecting social groups from space–Assessment of remote sensing-based mapped morphological slums using income data," *Remote Sens. Lett.*, vol. 9, no. 1, pp. 41–50, 2018.

[11] P. Hurskainen and P. Pellikka, "Change detection of informal settlements using multi-temporal aerial photographs–the case of Voi, SE-Kenya," in *Proc. 5th Afr. Assoc. Remote Sens. Environ. Conf.*, 2004, pp. 1–9.

[12] D. Kohli, A. Stein, and R. Sliuzas, "Uncertainty analysis for image interpretations of urban slums," *Comput., Environ. Urban Syst.*, vol. 60, pp. 37–49, 2016.

[13] T. Stark, M. Wurm, H. Taubenböck, and X. X. Zhu, "Slum mapping in imbalanced remote sensing datasets using transfer learned deep features," in *Proc. Joint Urban Remote Sens. Event*, 2019, pp. 1–4.

[14] F. M. Van Coillie, S. Gardin, F. Anseel, W. Duyck, L. P. Verbeke, and R. R. De Wulf, "Variability of operator performance in remote-sensing image interpretation: The importance of human and external factors," *Int. J. Remote Sens.*, vol. 35, no. 2, pp. 754–778, 2014.

[15] C. Kinkeldey, "Evaluating the use of uncertainty visualization for exploratory analysis of land cover change: A qualitative expert user study," *Comput. Geosci.*, vol. 84, pp. 46–53, 2015.

[16] H. Taubenböck and N. J. Kraff, "The physical face of slums: A structural comparison of slums in Mumbai, India, based on remotely sensed data," *J. Housing Built Environ.*, vol. 29, no. 1, pp. 15–38, Mar. 2014.

[17] M. Kuffer and J. Barros, "Urban morphology of unplanned settlements: The use of spatial metrics in VHR remotely sensed images," *Procedia Environ. Sci.*, vol. 7, no. 11, pp. 152–157, 2011.

[18] M. Wurm, H. Taubenböck, M. Weigand, and A. Schmitt, "Slum mapping in polarimetric SAR data using spatial features," *Remote Sens. Environ.*, vol. 194, pp. 190–204, 2017.

[19] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection-systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.

[20] P. Hofmann, J. Strobl, T. Blaschke, and H. Kux, "Detecting informal settlements from quickbird data in Rio de Janeiro using an object based approach. *Object-Based Image Anal.*, pp. 531–553, 2008. [Online]. Available: https://www.springer.com/de/book/9783540770572

[21] M. Kuffer *et al.*, "The role of earth observation in an integrated deprived area mapping "System" for low-to-middle income countries," *Remote Sens.*, vol. 12, no. 6, pp. 982–1007, 2020.

[22] T. Grippa *et al.*, "Mapping urban land use at street block level using openstreetmap, remote sensing data, and spatial metrics," *ISPRS Int. J. Geoinformation*, vol. 7, no. 7, pp. 246–266, 2018.

[23] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.
[24] C. Persello and A. Stein, "Deep fully convolutional networks for the detection of informal settlements in VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2325–2329, Dec. 2017.
[25] H. Taubenböck and N. J. Kraff, *Das globale Gesicht urbaner Armut? Siedlungsstrukturen in slums. – In: Globale Urbanisierung (S. 107-119)*. Berlin, Germany: Springer, 2015.
[26] D. Kohli, R. V. Sliuzas, N. Kerle, and A. Stein, "An ontology of slums for image-based classification," *Comput. Environ. Urban Syst.*, vol. 36, no. 2, pp. 154–163, 2012.
[27] N. J. Kraff, H. Taubenböck, and M. Wurm, "How dynamic are slums? EO-based assessment of Kibera's morphologic transformation," in *Proc. IEEE Joint Urban Remote Sens. Event*, 2019, pp. 1–4.
[28] H. Shu, S. Spaccapietra, C. Parent, and D. Q. Sedas, "Uncertainty of Geographic Information and its Support in MADS," in *Proc. 2nd Int. Symp. Spatial Data Qual.*, Hong Kong, Mar. 2003, pp. 1–13.
[29] M. F. Goodchild, "Geographical information science," *Int. J. Geographical Inf. Syst.*, vol. 6, no. 1, pp. 31–45, 1992.
[30] A. M. MacEachren *et al.*, "Visualizing geospatial information uncertainty: What we know and what we need to know," *Cartography Geographic Inf. Sci.*, vol. 32, no. 3, pp. 139–160, 2005.
[31] M. Molenaar, "Three conceptual uncertainty levels for spatial objects," *Int. Archives Photogrammetry Remote Sens.*, vol. 33, pp. 670–677, 2000.
[32] J. E. Estes, E. J. Hajic, and L. R. Tinney, "Fundamentals of image analysis: Analysis of visible and thermal infrared data," *Manual Remote Sens.*, vol. 1, pp. 987–1124, 1983.
[33] T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation*, 1979 7th ed. Hoboken, NJ, USA: Wiley,, 2015.
[34] G. M. Foody, "Status of land cover classification accuracy assessment. *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, 2002.
[35] J. Albertz, Einführung in die Fernerkundung. (Wissenschaftliche Buchgesellschaft). 3. Aufl., Darmstadt. (in German), 2007.
[36] W. Gerbino and D. Salmaso, "The effect of amodal completion on visual matching," *Acta Psychologica*, vol. 65, no. 1, pp. 25–46, 1987.
[37] M. Herold, X. H. Liu, and K. C. Clarke, "Spatial metrics and image texture for mapping urban land use,“ *Photogrammetric Eng. Remote Sens.*, vol. 69, pp. 991–1001, 2003.
[38] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, Mar. 1977.
[39] A. Diekmann, Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. (rowohlts enzyklopädie). 15. Aufl., Reinbek. (in German), 2006.
[40] T. Leichtle, C. Geiß, M. Wurm, T. Lakes, and H. Taubenböck, "Unsupervised change detection in VHR remote sensing imagery–an object-based clustering approach in a dynamic urban environment," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 54, pp. 15–27, 2017.
[41] K. Jacobsen, and G. Büyüksalih, "Topographic mapping from space," in *Proc. 4th EARSeL Workshop Remote Sens. Developing Countries/GISDECO*, Istanbul, Turkey, Jun. 2008, pp. 1–10.
[42] G. J. Hay, T. Blaschke, D. J. Marceau, and A. Bouchard, "A comparison of three image object methods for the multiscale analysis of landscape structure," *ISPRS J. Photogrammetry Remote Sens.*, vol. 57, pp. 327–345, 2003.
[43] J. Pratomo, M. Kuffer, J. Martinez, and D. Kohli, "Coupling uncertainties with accuracy assessment in object-based slum detections, case study: Jakarta, Indonesia," *Remote Sens.*, vol. 9, no. 11, pp. 1164–1180, 2017.
[44] E. Nyandwi, M. Koeva, D. Kohli, and R. Bennett, Comparing Human versus Machine-Driven Cadastral Boundary Feature Extraction, 2019.
[45] National Service for the Implementation of Regulations Netherlands. The Monitoring Agricultural Resources (MARS) Unit Mission. European Commission. Joint Research Centre (JRC). LPIS Workshop Malta, Jun. 2012.
[46] J. R. Jensen and D. C. Cowen, "Remote sensing of urban suburban infrastructure and socio-economic attributes," *Photogrammetric Eng. Remote Sens.*, vol. 65, pp. 611–622, 1999.
[47] C. Brelsford, T. Martin, J. Hand, and L. M. Bettencourt, "Toward cities without slums: Topology and the spatial evolution of neighborhoods," *Sci. Adv.*, vol. 4, no. 8, 2018, Art. no. eaar4644.
[48] Q. Li, Y. Shi, X. Huang, and X. X. Zhu, "Building footprint generation by integrating convolution neural network with feature pairwise conditional random field (FPCRF)," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.2973720.
[49] M. Wurm, J. Goebel, G. G. Wagner, M. Weigand, S. Dech, and H. Taubenböck, "Inferring floor area ratio thresholds for the delineation of city centers based on cognitive perception," Environment and Planning B: Urban Analytics and City Science, 2019. [Online]. Available: https://doi.org/10.1177/2399808319869341

Nicolas Johannes Kraff received the diploma degree (Dipl.-Geogr.) in human geography with a specialization in spatial development, remote sensing, and cartography from the University of Trier, Trier, Germany, in 2011. He is currently working toward the Ph.D. degree with the DLR.

He was with UNEP and the German Federal Office of Civil Protection and Disaster Assistance as an intern. Since 2013, he has been working in the area of European funding in Wiesbaden, Germany and next to it provides research about urban geography in cooperation with the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. His research focuses on mapping methodologies, global physical and morphological appearances of urban poverty areas and temporal change.

Mr. Kraff won together with M. Wurm and H. Taubenböck the award for the Best Poster at the Joint Urban Remote Sensing Event, which was co-sponsored by IEEE.

Michael Wurm received the Diploma degree (Mag. rer. nat.) in geography with a specialization in remote sensing, GIS, and spatial research from the University of Graz, Graz, Austria, in 2007, and the Ph.D. degree (Dr. rer. nat.) in surveying and geoinformation from the Graz University of Technology, Graz, Austria, in 2013.

He was with the Institute of Digital Image Processing, Joanneum Research, Graz, in 2007. In 2008, he joined the University of Wurzburg, Germany, where he was engaged in interdisciplinary research between earth observation data and social sciences. Since 2011, he is with the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Weßling, Germany, where he is involved in topics on urban geography, urban remote sensing, and urban morphology, and slum mapping research. Since 2013, he has been a Lecturer with the University of Graz.

Hannes Taubenböck received the diploma in geography from the Ludwig-Maximilians-Universität München, Munich, Germany, in 2004, and the Ph.D. degree (Dr. rer. nat.) in geography from the Julius-Maximilians-Universität Würzburg, Würzburg, Germany, in 2008.

In 2005, he joined the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. After a postdoctoral research phase with the University of Würzburg (2007–2010), he returned in 2010 to DLR–DFD as a scientific employee. In 2013, he became the Head of the "City and Society" team. In 2019, he habilitated with the University of Würzburg in geography. His current research focuses on urban remote sensing topics, from the development of algorithms for information extraction to value adding to classification products for findings in urban geography.