Int J Appl Earth Obs Geoinformation 81 (2019) 146–153

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Sensitivity of multiresolution segmentation to spatial extent

Lucian Drăguţ[a,*], Mariana Belgiu[b], George Popescu[a], Peter Bandura[c]

[a] Department of Geography, West University of Timişoara, Romania
[b] Faculty of Geo-Information Science and Earth Observation, University of Twente, Romania
[c] Department of Physical Geography and Geoecology, Comenius University in Bratislava, Romania

ARTICLE INFO

Keywords:
Scale
Buildings
Crop fields
Image segmentation
Estimation of Scale Parameters

ABSTRACT

Spatial extent (i.e. the size of the study area) is acknowledged as an important component of scale, together with grain (i.e. cell size). While the influence of grain on multiresolution segmentation has been evaluated, the impact of spatial extent is still poorly understood. The main goal of our study was to evaluate how changing the extent affects multiresolution segmentation, in respect to the geometric accuracy of the resulting image objects.

The experiments were carried out on very-high resolution optical images in four study areas: the City of Manchester (UK), the region of Normandy (France), the City of Tampa (Florida), USA, and the province of Flevoland (the Netherlands). Data sets of various extents were created by partitioning each image into regular tiles with eCognition® Server. The smallest tile size was 100 × 100 pixels, which doubled iteratively, until no further partition was possible, so that the image was processed at its full extent. Each tile was segmented with the Estimation of Scale Parameters (ESP2) algorithm and for each of the three generated levels the degree of overlap between the image objects and the reference polygons representing buildings and crop fields was checked. Segmentation accuracy was performed with the following metrics: Area Fit Index, Under-Segmentation, Over-Segmentation, D- index, and Quality Rate.

The results show that the geometric accuracy improved by 8–19% in Quality Rate when multiresolution image segmentation was performed in the smallest extent (100 × 100 pixels), as compared to the segmentation of whole images. These findings challenge previous assumptions and findings that partitioning an image into regularly-sized tiles can bias segmentation, and are relevant to guiding the setup of tile size in a distributed computing framework.

1. Introduction

Within the context of increasing availability of high-resolution imagery, image segmentation is regarded as a solution to automate conversion of the raw data into tangible information, which is required in many application domains (Blaschke, 2010). Multiresolution segmentation (MRS) is now one of the most important algorithms in the object-oriented analysis of remote sensing imagery (Cheng and Han, 2016). MRS is a region-growing algorithm that relies on the homogeneity criteria for partitioning an image into object primitives, which are the basic entities for further processing procedures (Baatz and Schäpe, 2000). As the name suggests, MRS is intimately related to scale in terms of outputs: a *scale parameter (SP)* controls the internal homogeneity of object primitives, which is inversely correlated with their size. A scene can therefore be segmented at a variety of scale levels, ideally emulating the scale levels of geographical features distinguishable in that scene (Costa et al., 2018). However, the scale of segmentation outputs depends on the scale characteristics of the input imagery, namely spatial resolution and spatial extent as the most important components of the geospatial data (Goodchild, 2001).

While the relationship between segmentation scale and spatial resolution has been relatively well addressed, the impact of spatial extent on segmentation results has been basically ignored. For instance, an ample review of supervised object-based land-cover classification studies (Ma et al., 2017) has identified 92 case studies which provided the necessary information to determine the former relationship, and none relevant to the later one. The cited study reveals a significant ($p < 0.05$), but weak ($R^2 = 0.058$) inverse relationship between grid size and optimal segmentation scales reported in the 92 studies, and speculates that the size of the study area might have influenced these statistics. Indeed, the size of the study area would likely impact on the results of region-growing algorithms, such as MRS, which rely on

* Corresponding author at: Department of Geography, West University of Timisoara, Blvd. V. Parvan 4, Timisoara 300223, Romania.
E-mail addresses: lucian.dragut@fulbrightmail.org (L. Drăguţ), m.belgiu@utwente.nl (M. Belgiu), georgepopescu594@gmail.com (G. Popescu), peter.bandura@uniba.sk (P. Bandura).

https://doi.org/10.1016/j.jag.2019.05.002
Received 1 February 2019; Received in revised form 1 May 2019; Accepted 8 May 2019
Available online 22 May 2019
0303-2434/

homogeneity, as it is known that increasing the extent of a study area will increase heterogeneity (Atkinson and Tate, 2000). Understanding the impact of extent upon segmentation is urged by recent developments in parallel processing of geospatial data, as for instance Google Earth Engine (Gorelick et al., 2017). While offering massive computational capabilities, such facilities require tiling a dataset, thus a decision on the processing extent.

We have not identified studies reporting directly on the impact of extent upon segmentation. However, a couple of publications aiming at improving segmentation via more localized approaches are relevant to the topic. For instance, d'Oleire-Oltmanns and Tiede (2014) proposed a procedure to subset a QuickBird image so that MRS applied only to those regions of the image where gullies were dominant. The study found that segmentation results improved as compared to segmentation of the whole image, likely due to reduction in heterogeneity. Kavzoglu et al. (2017) introduced a two-stage segmentation to classify semi-urban landscapes: an initial partitioning of images into broad regions sharing similar characteristics, followed by MRS parameterization for each region. The study found that adjusting the *SP* to the individual regions led to improvements in classification accuracy up to 5% as compared to segmentations with a global *SP*. Similarly, Grippa et al. (2017) and Georganos et al. (2018) found that optimizing segmentation parameters locally rather than globally helps in mapping heterogeneous urban environments. Local optimization methods led to up to 1.5% increase in overall accuracy as well as improvements in the geometric accuracy of the image objects. Comparable findings were also reported by Bandura et al. (2018), where partitioning of the complex full scene into more homogeneous domains led to delineation of geomorphological features, which vary in shape and homogeneity properties. The authors locally adjusted SP for each domain, which, complemented with subsequent removal of distinct geomorphological features, positively affected the overall level of homogeneity in the MRS. This approach led to the decrease of inner variability of input layers in the segments by 2.5% when compared to the globally-set MRS.

The results reported so far suggest that image segmentation works better on smaller extents. However, those extents have been primarily defined by their relative homogeneity, rather than size per se. Therefore, we only know that segmentation improves due to reducing heterogeneity via regionalization, and are still uncertain what to expect when considering 'pure' extent, i.e. defined by regular tiles rather than homogeneous regions. Georganos et al. (2018) found some merits of regularly-shaped rectangular tiles, but investigations were not systematic as the aim of the study was different. Within the context of parallel processing, constructing homogeneous regions would require additional work if performed in a more automatic way (as in Kavzoglu et al., 2017 or Georganos et al., 2018), or would be practically impossible if regions are delineated manually as in d'Oleire-Oltmanns and Tiede (2014) or Grippa et al. (2017). Thus, more knowledge is required to support decision regarding the size of the tiles to be submitted to processing.

This study aims at evaluating the sensitivity of MRS to spatial extent, in order to answer two questions: 1) does spatial extent impact on the geometric accuracy of the image objects?, and 2) if so, is there any relationship between extent and segmentation results? Based on the results of the reviewed studies, we hypothesize that MRS works better on the smallest extents. However, the geometric accuracy of image objects might suffer because of the edge effect caused by an excessive tiling, as for instance splitting of building roofs along tile edges (Georganos et al., 2018). We restrict the evaluation to the geometric accuracy and do not consider the thematic one as well because the relationship between segmentation and classification is not straightforward, as shown in literature (e.g. Belgiu and Drăguţ, 2014). The experiments are conducted in a scenario of extracting buildings and crop fields from high-resolution images.

Fig. 1. Study areas located in Manchester, UK (up left), Normandy, France (bottom left), and Tampa, FL, USA (right), illustrated in RGB natural-colour composite. The data sets are available at http://web.eee.sztaki.hu/remotesensing/building_benchmark.html and https://spacenetchallenge.github.io/datasets/datasetHomePage.html.

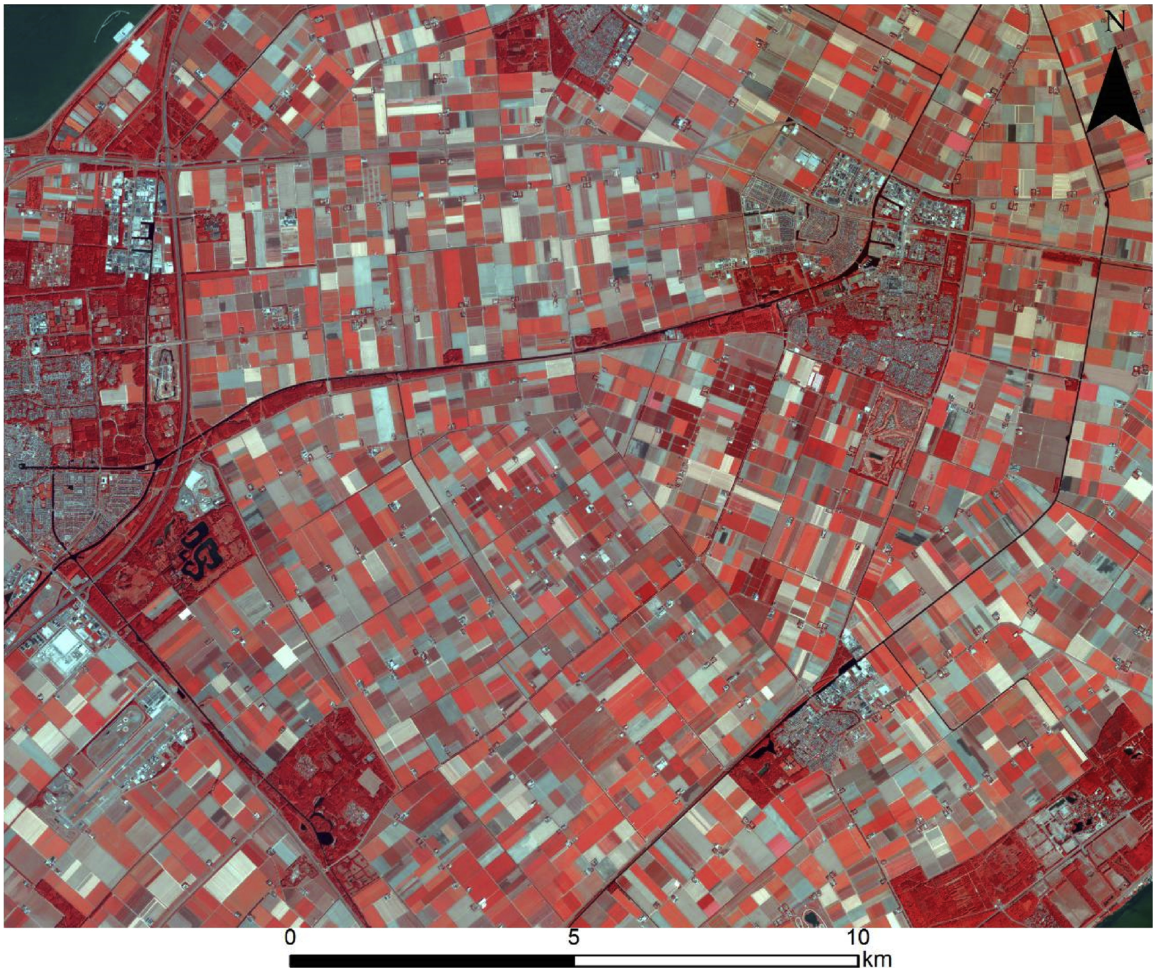


Fig. 2. The study area of Flevoland, the Netherlands illustrated in false-colour composite (Source PlanetScope, https://www.planet.com/explorer).

2. Study areas and data

Three study areas were chosen for the analysis of buildings: two located in Europe – the City of Manchester (UK) and region of Normandy (France) – and one located in the USA – the City of Tampa (Florida) (Fig. 1). Each study area represents a different urban pattern and distribution of buildings, bringing larger variety into the experiments. While the areas of Manchester and Normandy represent a residential type, the area of Tampa is more of a mixed type (urban, residential as well as commercial).

Analysis of crop fields was conducted in Flevoland (Fig. 2), which is one of the twelve provinces of the Netherlands. This province was reclaimed from the sea for the agricultural expansion purpose. The crop fields are variable in size and regular in shape.

The input data were obtained from the existing building detection benchmark datasets available online. The data are in the form of very-high resolution optical images generated by commercial satellites. Both images from Europe (smaller scenes) were downloaded from SZTAKI-INRIA Building Detection Benchmark dataset (Benedek et al., 2012) as raw images. The image of Tampa (larger scene) was downloaded from the URBAN3D Challenge dataset (Goldberg et al., 2017) provided by VRICON and hosted by SpaceNet as a true orthorectified RGB image. Along with the images, the reference polygons – ground truth footprints of buildings – are included. PlanetScope 4-band images were used for the segmentation of the crop fields. Ground truth data were downloaded from the PDOK platform, which is the Dutch open platform (https://www.pdok.nl/) used to share geodata with the public at large. Table 1 provides an overview of the study areas and input data.

3. Methods

To investigate the impact of changing extent on segmentation, data sets of various extents were created by partitioning each image into regular tiles with eCognition® Server (Fig. 3). The smallest tile size was 100 × 100 pixels (E1 in Fig. 3), which doubled iteratively (E2 to E4 in Fig. 3), until no further partition was possible, so that the image was processed at its full extent (E5 in Fig. 3). Reference polygons were split as well when located at the edges of tiles. Thus, five data sets (E1 to E5) resulted for Manchester and Normandy case studies and seven (E1 to E7) for Tampa and Flevoland, respectively. Tiling of the Tampa image stopped at 6400 × 6400 because further partitioning would produce imbalanced-sized tiles, while processing the full image (more than 250 million pixels) exceeded the available computational resources. According to the full extents of images (Table 1), the number of tiles is variable in each data set, as shown in Table 2.

Each tile was then segmented with the ESP2 tool (Drăguţ et al., 2014), which automatically produced image objects by calibrating the SPs as a function of the Local Variance specific to the tile. The tool was run with the default parameters (Drăguţ et al., 2014) and delivered image objects at three scale levels for each tile, from L1 (the finest) to L3 (the broadest). This processing step aimed at consistency, rather than accuracy. Tiles were further stitched with eCognition® Server, so that each data set (E1 to E7) tripled according to the number of generated scale levels (L1 to L3). The stitched images, including the tile boundaries and the objects split by them, were further evaluated for geometric accuracy.

The data sets produced as above were evaluated for geometric

Table 1
Study areas and characteristics of the input data.

No.	Location	Source	Image data			Reference objects (n)
			Spatial resolution (m)	Bands	Image size (pixels)	
1	Manchester	Google Earth	1	R, G, B	1412 × 797	196
2	Normandy	Google Earth	1	R, G, B	1437 × 814	172
3	Tampa	Vricon	0.5	R, G, B	16320 × 15642	36139
4	Flevoland	PlanetScope	3	R, G, B, NIR	6746 × 5421	3479

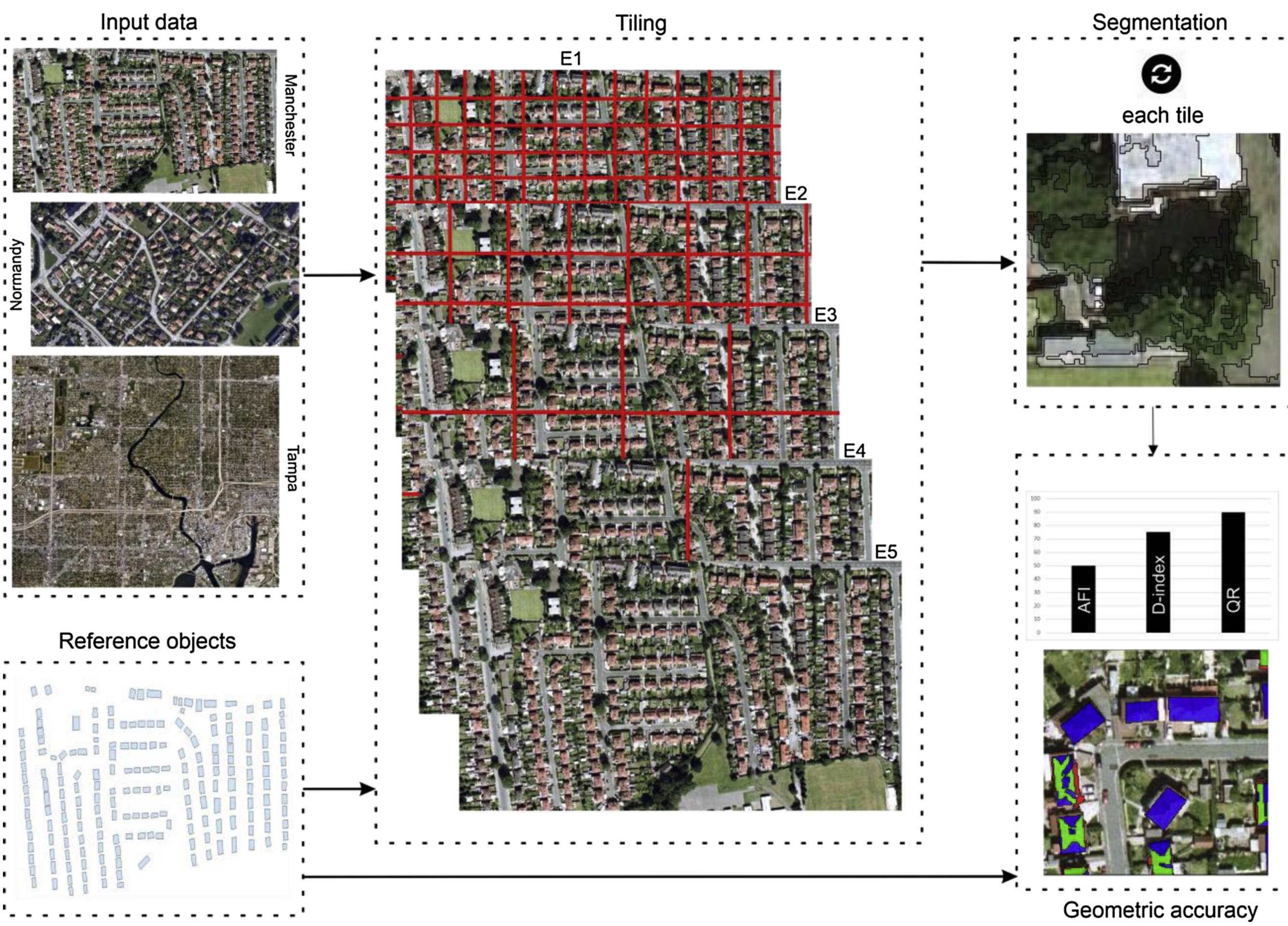


Fig. 3. Overview of the experiments to evaluate sensitivity of multiresolution segmentation to spatial extent.

Table 2
Characteristics of data sets created by tiling: number of data sets per case study; number of tiles and reference polygons per data set.

Data set	Number of tiles/reference polygons							
	Manchester		Normandy		Tampa		Flevoland	
	No tiles	No polygons	No tiles	No polygons	No tiles	No polygons	No tiles	No polygons
E1	120	337	135	315	25748	63019	3740	15059
E2	32	259	40	231	6478	48873	952	9043
E3	8	228	12	196	1640	42152	238	6544
E4	2	196	4	181	420	39072	63	5408
E5	1	196	1	172	110	37641	20	4901
E6	–	–	–	–	30	36952	6	4625
E7	–	–	–	–	9	36571	1	3479

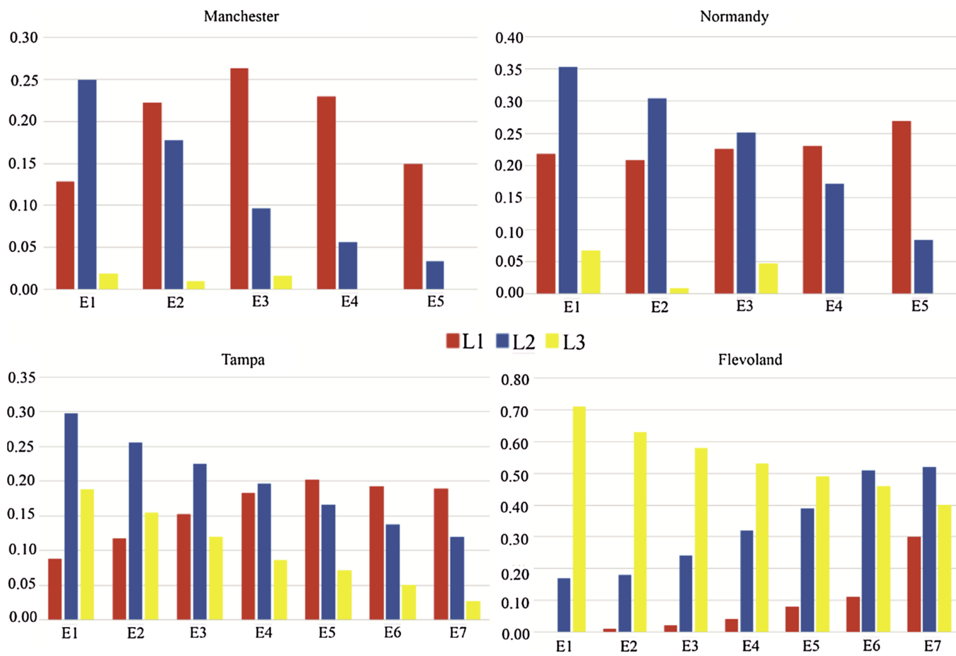


Fig. 4. Quality rate of the image objects obtained with ESP2 at three scale levels: L1, L2, and L3.

accuracy with the help of an algorithm introduced by Eisank et al. (2014). Image objects produced through segmentation were compared with the reference polygons. First, a mutual spatial overlap of minimum 50% was determined, and then segmentation accuracy metrics were calculated. Image objects were considered candidates in evaluation only if they overlapped at least half of the corresponding reference polygon AND the reference polygon overlapped at least half of their area (Clinton et al., 2010). The following segmentation accuracy metrics were computed: Area Fit Index (AFI) (Lucieer and Stein, 2002), Under-Segmentation (US), Over-Segmentation (OS), D- index, that combines US and OS (Clinton et al., 2010), and the Quality Rate (QR) (Winter, 2000). All of the metrics range from 0 to 1, where 0 indicates perfect spatial match between reference polygons and individual image objects, except for QR, where 1 is the ideal value. Images at maximum extents served as baseline to evaluate the impact of the extent on geometric accuracy of the segmented objects. At lower extents, the stitched images were evaluated at once, rather than averaging the results of the individual tiles.

4. Results

The results of geometric accuracy assessment of segmentation are shown in Fig. 4 and Tables 3–6. As expected, the segmentation accuracy metrics are consistent, i.e. the best score of one metric coincides with the best scores or combinations of the others. Other studies (e.g. Chen et al., 2018) also found strong correlations between metrics such as QR and AFI. Therefore QR has been chosen to comparatively illustrate the results across the study areas, extents, and ESP2 levels (Fig. 4).

For the buildings case studies, it is apparent that the finest ESP2 level (L1) tends to work better in larger extents, while the broader levels (L2 and L3) tend to perform better in smaller extents. This means that larger extents require lower SP values for segmentation than smaller extents, which might look counter-intuitive. Within the context of building extraction, these results look normal: lower SP values produce smaller and more homogeneous objects, which fit the size of buildings in large extents, while in smaller extents the same values would lead to over-segmentation of buildings. Conversely, higher SP values produce larger image objects, which fit the size of buildings in small extents, while exceeding their size in larger extents. In some cases (E4 and E5 in the Manchester and Normandy areas), L3 contained objects so large that no reference polygon was matched according to the "50%" rule, and therefore some segmentation accuracy metrics remained undefined (Tables 3 and 4).

An opposite trend is visible in the case of Flevoland: the broadest

Table 3
Segmentation accuracy results for the Manchester area.

Extent	ESP Level	AFI	D Index	Over Segmentation	Quality Rate	Under Segmentation
1	1	0.84	0.63	0.87	0.13	0.17
	2	0.61	0.55	0.72	0.25	0.29
	3	0.98	0.71	0.98	0.02	0.18
2	1	0.70	0.56	0.76	0.22	0.21
	2	0.70	0.62	0.80	0.18	0.34
	3	0.98	0.77	0.99	0.01	0.47
3	1	0.60	0.54	0.71	0.26	0.26
	2	0.86	0.67	0.90	0.10	0.30
	3	0.98	0.71	0.98	0.02	0.20
4	1	0.63	0.57	0.74	0.23	0.31
	2	0.91	0.71	0.94	0.06	0.34
	3	Undefined	Undefined	Undefined	Undefined	Undefined
5	1	0.77	0.63	0.84	0.15	0.31
	2	0.95	0.72	0.97	0.03	0.32
	3	Undefined	Undefined	Undefined	Undefined	Undefined

Table 4
Segmentation accuracy results for the Normandy area.

Extent	ESP Level	AFI	D Index	Over Segmentation	Quality Rate	Under Segmentation
1	1	0.74	0.56	0.77	0.22	0.14
	2	0.53	0.45	0.62	0.35	0.19
	3	0.91	0.68	0.93	0.07	0.26
2	1	0.74	0.57	0.78	0.21	0.16
	2	0.58	0.50	0.67	0.30	0.21
	3	0.99	0.73	0.99	0.01	0.29
3	1	0.71	0.55	0.76	0.23	0.19
	2	0.63	0.55	0.72	0.25	0.26
	3	0.93	0.70	0.95	0.05	0.29
4	1	0.69	0.55	0.75	0.23	0.22
	2	0.74	0.61	0.82	0.17	0.28
	3	Undefined	Undefined	Undefined	Undefined	Undefined
5	1	0.63	0.53	0.71	0.27	0.22
	2	0.88	0.68	0.91	0.08	0.30
	3	Undefined	Undefined	Undefined	Undefined	Undefined

Table 5
Segmentation accuracy results for the Tampa area.

Extent	ESP Level	AFI	D Index	Over segmentation	Quality rate	Under segmentation
1	1	0.90	0.65	0.91	0.09	0.13
	2	0.63	0.50	0.69	0.30	0.15
	3	0.76	0.58	0.80	0.19	0.18
2	1	0.86	0.63	0.88	0.12	0.15
	2	0.68	0.53	0.73	0.26	0.15
	3	0.81	0.61	0.84	0.15	0.16
3	1	0.81	0.61	0.84	0.15	0.16
	2	0.72	0.55	0.77	0.22	0.16
	3	0.85	0.63	0.88	0.12	0.15
4	1	0.78	0.58	0.81	0.18	0.16
	2	0.76	0.57	0.80	0.20	0.15
	3	0.90	0.65	0.91	0.09	0.14
5	1	0.75	0.57	0.79	0.20	0.15
	2	0.80	0.59	0.83	0.17	0.14
	3	0.92	0.66	0.93	0.07	0.12
6	1	0.77	0.58	0.80	0.19	0.15
	2	0.84	0.62	0.86	0.14	0.14
	3	0.94	0.68	0.95	0.05	0.13
7	1	0.77	0.58	0.80	0.19	0.15
	2	0.86	0.63	0.88	0.12	0.15
	3	0.97	0.70	0.97	0.03	0.17

ESP2 level (L3) works better in smaller extents, while the finer levels (L1 and L2) tend to perform better in larger extents. Smaller extents require higher SP values for segmentation than larger extents, to adjust to the larger size of the crop fields. Higher SP values produce larger and more heterogeneous objects, which fit the size crop field in small extents, while lower SP values would lead to their over-segmentation. Over-segmentation was so severe in E1, that the accuracy metrics remained undefined for L1 (Table 6). The broadest ESP2 level (L3) tends to produce over-segmented objects, while the finer levels (particularly L2) adjust better to the size of crop fields.

While L1 and L3 do not exhibit consistent trends in evolution of QR across extents, L2 clearly shows declines in QR with increase in extent, in all building case studies. Image objects in L2 best match the reference buildings in the smallest extent in all three cases. In larger extents, objects in L1 start matching better the reference polygons, at various tipping points, which are likely caused by differences in image heterogeneity. The tipping points are E2 in Manchester, E4 in Normandy, and E5 in Tampa (Fig. 4). In Flevoland, all ESP levels show clear trends: L3 indicates declining QR with increase in extent, while L1 and L2 exhibit opposite trends. The best match between image objects and crop field polygons is provided by L3 at smaller extents (E1 to E5), and by L2 at the broadest extents (E6 and E7)(Fig. 4).

The best results achieved at maximum extents are inferior or equal to the best results displayed by the smaller extents, except for E3 and E4

Table 6
Segmentation accuracy results for the Flevoland area.

Extent	ESP Level	AFI	D Index	Over segmentation	Quality rate	Under segmentation
1	1	Undefined	Undefined	Undefined	Undefined	Undefined
	2	0.83	0.59	0.83	0.17	0.02
	3	0.25	0.20	0.27	0.71	0.03
2	1	0.99	0.70	0.99	0.01	0.02
	2	0.82	0.58	0.82	0.18	0.01
	3	0.34	0.26	0.36	0.63	0.03
3	1	0.98	0.69	0.98	0.02	0.02
	2	0.76	0.54	0.76	0.24	0.01
	3	0.40	0.29	0.41	0.58	0.02
4	1	0.96	0.68	0.96	0.04	0.01
	2	0.68	0.48	0.68	0.32	0.01
	3	0.45	0.33	0.46	0.53	0.02
5	1	0.92	0.65	0.92	0.08	0.01
	2	0.60	0.43	0.61	0.39	0.01
	3	0.49	0.35	0.50	0.49	0.02
6	1	0.89	0.63	0.89	0.11	0.01
	2	0.48	0.34	0.49	0.51	0.02
	3	0.53	0.38	0.54	0.46	0.02
7	1	0.70	0.50	0.70	0.30	0.01
	2	0.47	0.34	0.48	0.52	0.02
	3	0.59	0.42	0.60	0.40	0.02

in Normandy, and E5 and E6 in Flevoland. The ability of segmentation to delineate buildings improved by 8–11% in QR when performed on smaller extents, as compared to the segmentation of maximum extents. Moreover, segmentation of the smallest extent (E1) produced the best results, except for the Manchester case, where objects in L1 show slightly better accuracy results (+1%) at E3 (Table 3). Segmentation of crop fields improved in geometric accuracy by 19% when performed in the smallest extent, as compared to the largest one (Table 6).

5. Discussion

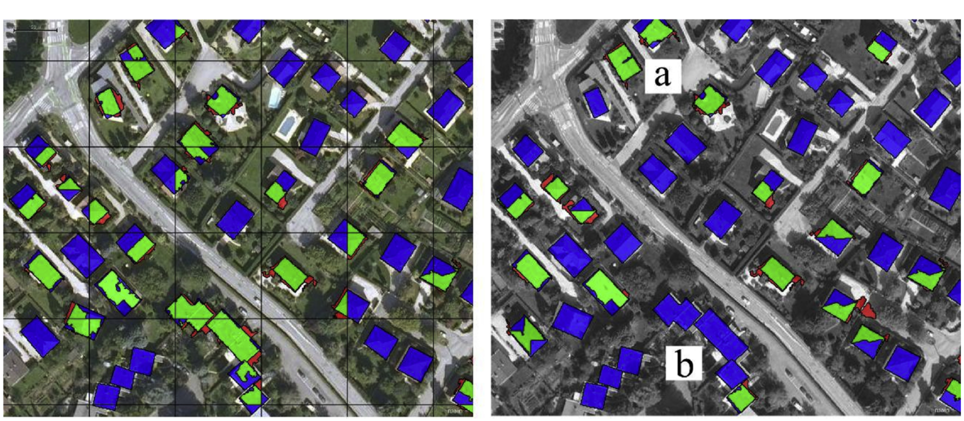
The overall goal of this study was to explore the relation between the spatial extent of an image and geometric accuracy of the objects delineated through multiresolution image segmentation. We found that segmentation performed at smaller spatial extents improves the geometric accuracy of the primitive objects. This finding is relevant in guiding the setup of tile size in a distributed computing framework such as the Google Earth Engine (GEE) platform (Gorelick et al., 2017). According to Lassalle et al. (2015) and Gorelick et al. (2017), classical image clustering procedures (including region growing segmentation) perform poorly on distributed computing platforms because of the image tiling, which introduces artefacts in the results. However, our results do not support this assumption. We found that segmentation is not only insensitive to the lack of information on the global image heterogeneity, but adjusting segmentation to local conditions actually improves results.

Previous studies have already reported that increasing the homogeneity of the study areas improves the segmentation results. The way that partitioning was performed varied among these studies. Grippa et al. (2017), for example, used manually delineated urban land use zones to optimize the segmentation procedure at the local level. d'Oleire-Oltmanns and Tiede (2014) also delineated manually the spatial units used to constrain the segmentation to high density gullies areas. While these local segmentation approaches are good at improving the segmentation results, they rely on manual delineation of the homogeneous spatial units, which makes them time-consuming and less appropriate for use in operational image analysis environments. Alternatively, Cánovas-García and Alonso-Sarría (2015) used existing agricultural plots as homogeneous spatial units to delineate different cropland areas. Despite the fact that the quality of the segmentation results improved considerably, this study relies on auxiliary data that are quite often missing in the investigated areas. Georganos et al. (2018) tested the efficiency of automatic splitting of the image into tiles of equal area and reported that this approach can cause errors at the tiling borders such as splitting building roofs in several parts. Consequently, they proposed the partition of the input image into several tiles by taking into account the linear features (such as roads) present in that image. However, our study revealed that even if the image tiles impact the geometric accuracy of the primitive objects located at the tiles edges, the overall segmentation errors occurring through tiling are lower than those obtained when the whole image is used as input, as illustrated in Fig. 5. On the one hand, the geometry of buildings may be damaged by tiling, as for instance the two buildings marked by *a*. On the other hand, many other buildings that were not delineated in the segmentation of the full image show up in the tiled one, even when buildings were split into multiple parts, as for instance the group of objects marked by *b*. Therefore, we recommend the automatic splitting of images into tiles of equal size for increasing the quality of the multiresolution segmentation results. To remove the artefacts occurring at the tile edges, one of the solutions available in the literature such as those proposed by Lassalle et al. (2015) can be applied.

The segmentation performs better when the image is partitioned into smaller tiles, likely due to the reduction in image heterogeneity, which is related to the number, size and patterns of the objects present in the image (Stein and De Beurs, 2005). Intrinsic image complexity, i.e. the amount of information present in an image, is less likely responsible for these results since different images can have the same complexity, as proved by Stein and De Beurs (2005). Thus, by reducing the size of the tiles used for segmentation, the inherent pattern complexity of the image is reduced and consequently, target objects are better delineated. Further, the reduction of image heterogeneity has a direct impact on the scale parameter which is lower when the homogeneity of the image increases (Cánovas-García and Alonso-Sarría, 2015).

In general, the geometric quality of the delineated primitive objects was not very high (Fig. 4), mainly because of the over-segmentation of the buildings into several objects. Over-segmentation is caused by the contrast between the lighted and shadow sides of the roofs in the three study areas, which challenges segmentation in general regardless of the adopted segmentation method. Despite this relatively low geometric accuracy, the accuracy of buildings classification can be improved considerably by merging the objects belonging to the same building in the further classification steps (Belgiu and Drăguţ, 2014). In contrast, segmentation of crop fields performed much better in terms of geometric accuracy, with a QR that achieved a value of 0.71, as compared to the maximum value of 0.35 achieved by the segmentation of buildings.

Our findings enable informed decision on the size of the tiles to be used for multiresolution image segmentation, following the rule "the smaller, the better". However, these results should be regarded with some caution in case of large objects, given that the relationship between the extent and the size of the target objects has not been fully considered in this study. The ratio extent/object size at the minimum extent (E1) ranges between 8.71 (Tampa, minimum extent = 2500 sq m; average object size = 287 sq m) and 0.14 (Flevoland, minimum extent = 9000 sq m; average object size = 63,517 sq m). More systematic studies on the relationship between the extent and the size of the target objects would be necessary to find whether the rule "the smaller, the better" holds true for other categories of objects.



E1L2 E5L1

Fig. 5. A detailed view of an approximately 500 x 500 m area in the Normandy case study to visualize the impact of tiling on segmentation. Left: the image tiled in 100 × 100 pixels. Right: the same area segmented at the full image extent. Both images display the best segmentation results achieved at ESP level 2 and 1, respectively. Polygon colours represent the degree of matching between image objects and reference buildings as follows: green = overlap; blue = under-estimation; red = over-estimation. See explanations in text for *a* and *b* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

6. Conclusions

This paper evaluated the impact of the spatial extent on the geometric accuracy of the objects delineated through multiresolution image segmentation. The experiments revealed that the geometric accuracy improved by 8–19% in Quality Rate when multiresolution segmentation was performed in smaller extents, as compared to the segmentation of whole images. Moreover, segmentation within the smallest extent (100 × 100 pixels) produced the best results in terms of geometric accuracy. These findings challenge previous assumptions and findings that partitioning an image into regularly-sized tiles can bias segmentation. Even though edge errors might occur, the quality of segmentation increases with the decrease of the size of tiles. The results suggest that multiresolution image segmentation within the framework of distributed computing should be performed in small tiles.

Declarations of interest

None.

Acknowledgments

This work was supported by a grant of Ministry of Research and Innovation, CNCS – UEFISCDI, project number PN-III-P4-ID-PCE-2016-0222, within PNCDI III. Andreea Scorobete and Emanuela Mihuţ were involved in data processing in the early stage of this research. Peter Bandura was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0054. The paper benefited from constructive comments received from the editor Freek van der Meer and three anonymous reviewers.

References

Atkinson, P.M., Tate, N.J., 2000. Spatial scale problems and geostatistical solutions: a review. Prof. Geogr. 52 (4), 607–623.
Baatz, M., Schäpe, A., 2000. Multiresolution segmentation-an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), Angewandte Geographische Informationsverarbeitung. Wichmann-Verlag, Heidelberg, pp. 12–23.
Bandura, P., Minár, J., Drăguţ, L., Bielik, M., 2018. Physically-based segmentation of the western carpathians (Central Europe). PeerJ Prepr. 6, e27083v1. https://doi.org/10.7287/peerj.preprints.27083v1.
Belgiu, M., Drăguţ, L., 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. Iprs J. Photogramm. Remote Sens. 96 (0), 67–75.
Benedek, C., Descombes, X., Zerubia, J., 2012. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 34 (1), 33–50.
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote. Sens. 65 (1), 2–16.
Cánovas-García, F., Alonso-Sarría, F., 2015. A local approach to optimize the scale parameter in multiresolution segmentation for multispectral imagery. Geocarto Int. 30 (8), 937–961.
Chen, Y., Ming, D., Zhao, L., Lv, B., Zhou, K., Qing, Y., 2018. Review on high spatial resolution remote sensing image segmentation evaluation. Photogramm. Eng. Remote Sensing 84 (10), 629–646.
Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogramm. Remote. Sens. 117, 11–28.
Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. Photogramm. Eng. Remote Sensing 76 (3), 289–299.
Costa, H., Foody, G.M., Boyd, D.S., 2018. Supervised methods of image segmentation accuracy assessment in land cover mapping. Remote Sens. Environ. 205, 338–351.
d'Oleire-Oltmanns, S., Tiede, D., 2014. Specific target objectsspecific scale levels? Application of the estimation of scale parameter 2 (ESP 2) tool for the identification of scale levels for distinct target objects. South Eastern Eur. J. Earth Observ. Geom. 3 (2S), 579–584.
Drăguţ, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. Isprs J. Photogramm. Remote. Sens. 88 (0), 119–127.
Eisank, C., Smith, M., Hillier, J.K., 2014. Multiresolution segmentation to automatically delimit landforms in DEMs: tests using synthetic drumlins. Geomorphology 214, 452–464.
Georganos, S., Grippa, T., Lennert, M., Vanhuysse, S., Johnson, B., Wolff, E., 2018. Scale matters: spatially partitioned unsupervised segmentation parameter optimization for large and heterogeneous satellite images. Remote Sens. 10 (9), 1440.
Goldberg, H., Brown, M., Wang, S., 2017. A benchmark for building footprint classification using orthorectified rgb imagery and digital surface models from commercial satellites. In: 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE. pp. 1–7.
Goodchild, M.F., 2001. Metrics of scale in remote sensing and GIS. Int. J. Appl. Earth Obs. Geoinf. 3 (2), 114–120.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.
Grippa, T., Georganos, S., Lennert, M., Vanhuysse, S., Wolff, E., 2017. A local segmentation parameter optimization approach for mapping heterogeneous urban environments using VHR imagery. Remote Sensing Technologies and Applications in Urban Environments II. International Society for Optics and Photonics p. 104310G.
Kavzoglu, T., Erdemir, M.Y., Tonbul, H., 2017. Classification of semiurban landscapes from very high-resolution satellite images using a regionalized multiscale segmentation approach. J. Appl. Remote Sens. 11 (3), 035016.
Lassalle, P., Inglada, J., Michel, J., Grizonnet, M., Malik, J., 2015. A scalable tile-based framework for region-merging segmentation. IEEE Trans. Geosci. Remote. Sens. 53 (10), 5473–5485.
Lucieer, A., Stein, A., 2002. Existential uncertainty of spatial objects segmented from satellite sensor imagery. IEEE Trans. Geosci. Remote. Sens. 40 (11), 2518–2521.
Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. ISPRS J. Photogramm. Remote. Sens. 130, 277–293.
Stein, A., De Beurs, K., 2005. Complexity metrics to quantify semantic accuracy in segmented Landsat images. Int. J. Remote Sens. 26 (14), 2937–2951.
Winter, S., 2000. Location similarity of regions. ISPRS J. Photogramm. Remote. Sens. 55 (3), 189–200.