

Article

Expert Knowledge as Basis for Assessing an Automatic Matching Procedure

Juan José Ruiz-Lendínez ^{*} , Francisco Javier Ariza-López  and Manuel Antonio Ureña-Cámara 

Department of Cartographic, Geodetic Engineering and Photogrammetry, University of Jaén, 23071 Jaén, Spain; fjariza@ujaen.es (F.J.A.-L.); maurena@ujaen.es (M.A.U.-C.)

* Correspondence: lendinez@ujaen.es; Tel.: +34-953-21-24-70; Fax: +34-953-21-28-54

Abstract: The continuous development of machine learning procedures and the development of new ways of mapping based on the integration of spatial data from heterogeneous sources have resulted in the automation of many processes associated with cartographic production such as positional accuracy assessment (PAA). The automation of the PAA of spatial data is based on automated matching procedures between corresponding spatial objects (usually building polygons) from two geospatial databases (GDB), which in turn are related to the quantification of the similarity between these objects. Therefore, assessing the capabilities of these automated matching procedures is key to making automation a fully operational solution in PAA processes. The present study has been developed in response to the need to explore the scope of these capabilities by means of a comparison with human capabilities. Thus, using a genetic algorithm (GA) and a group of human experts, two experiments have been carried out: (i) to compare the similarity values between building polygons assigned by both and (ii) to compare the matching procedure developed in both cases. The results obtained showed that the GA—experts agreement was very high, with a mean agreement percentage of 93.3% (for the experiment 1) and 98.8% (for the experiment 2). These results confirm the capability of the machine-based procedures, and specifically of GAs, to carry out matching tasks.

Keywords: machine learning; expert knowledge; automatic matching; spatial data accuracy; automatic assessment



Citation: Ruiz-Lendínez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A. Expert Knowledge as Basis for Assessing an Automatic Matching Procedure. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 289. <https://doi.org/10.3390/ijgi10050289>

Academic Editor: Wolfgang Kainz

Received: 24 February 2021

Accepted: 30 April 2021

Published: 2 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rise of machines to human-level performance in many complex recognition tasks—mainly due to the development of artificial intelligence (AI)—there have emerged a huge number of studies focused on comparing information processing in humans and machines [1–4], thereby reviving a longstanding debate and a competition between the merits of both [5]. The main purpose of these studies is to develop a deeper understanding of the mechanisms of human perception and to improve machine learning procedures. These procedures are capable of addressing certain types of tasks that, either by their complexity or by their volume, are difficult and costly to solve by humans, reaching accuracy levels similar to them. Among these tasks we highlight object recognition [6,7], depth estimation [8] and the most important one for our study: objects matching [1,9,10].

Matching techniques are basic tools for dealing with graphic information [11,12] and specifically geographic information (GI), whose main defining feature is the spatial position. Nowadays GI is very much in demand due to its use in many interest areas (such as business, tourism, fleet management, military development, land management, etc.) and to its economic importance [13]. This is why new cartographic products are continually emerging in reply to the demands of an increasingly expanding market. These cartographic products are usually obtained from new ways of mapping based on the integration of spatial data from heterogeneous sources, and with different degrees of detail [14–16], so their final quality levels—which are often unknown—depend on the quality levels of the initial sources. In short, it could be said that these new ways of mapping require new

and more efficient ways of assessing the positional accuracy of the resulting products for which traditional measures of data quality are not applicable, which in turn requires the automation of the associated processes.

According to the existing literature [17–19], the automation of the processes associated with the PAA of spatial data is based on the comparison between the locations of corresponding spatial objects from two GDB (called Reference GDB and Tested GDB) with different levels of accuracy and detail. In that regard, it must be noted that this methodological approach is based on the assumption that the accuracy of one of the GDB is high enough to ignore the unmeasured difference between it and the real one [20]. In order to automatically compare the locations of these corresponding objects links between them are required [17]. In this context, an automatic PAA procedure is essentially reduced to a pattern matching problem where there is a key issue: how to measure the similarity between corresponding spatial objects from two heterogeneous datasets (in our case, GDBs) in order to classify and match them in an automated way.

1.1. Automatic Matching as a Solution to PAA Procedures

Matching spatial objects from two heterogeneous datasets is a complex decision process [9]. According to these authors, there are two main aspects to address in this process: First, what is the most appropriate spatial object to determine the matching? Second, what are the similarity measures to employ? In response, and in most cases, matching algorithms use building polygons as spatial objects to match and geometric descriptors as similarity measures to perform the matching process. With regard to the criteria applied for the determination of the matching, these algorithms are based on the percentage of overlapped area [10,21,22], the context by means of both the Delaunay triangulation [23] and the Voronoi diagram [24], the distance between turning functions [25], the belief theory on position and orientation [26] and probability classifiers over a set of “evidence”, both geometric and attribute-based [27]. The results provided by most of the above-referred studies show a high matching success ratio between objects belonging to heterogeneous sources. However, despite this efficiency many aspects still need to be improved. These aspects are closely linked both to the generation of false matching pairs in the cases of 1:n or n:m correspondences—multiple matching cases often associated with generalization processes suffered by building polygons with relatively complex contours (Figure 1)—and to the manual selection of landmarks from within two datasets. In both cases, an increase of the level of automation would contribute to improve matching processes, avoiding the acceptance of erroneously-matched objects.

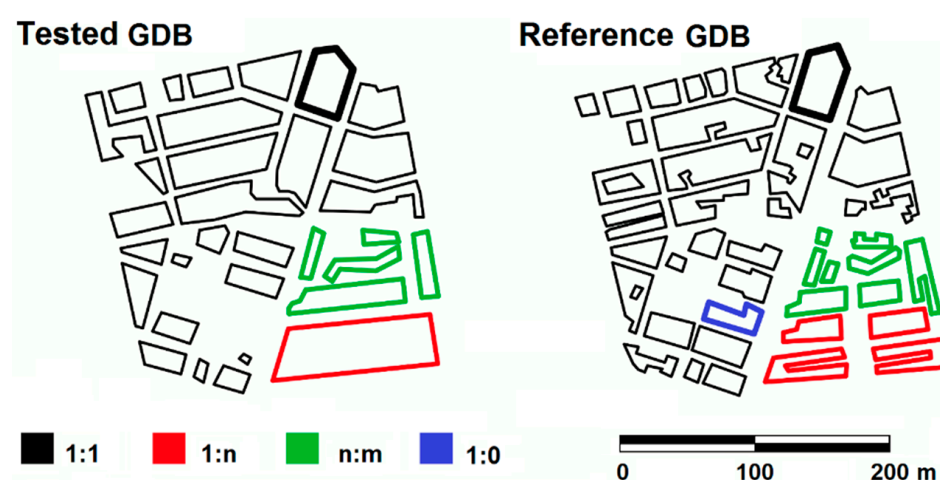


Figure 1. Some examples of correspondences between polygons. Example of 1:1 correspondence (black), example of 1:n correspondence (red), examples of n:m correspondences (green) and example of 1:0 correspondence (blue). Source: [1].

In order to solve, as far as possible, the aspects discussed above, in [1] we developed an innovative automated matching methodology for urban GDBs. These GDBs are presented as a set of georeferenced vector covers distributed by layers including a vector layer of buildings (city blocks). This methodology—initially proposed as a solution to PAA procedures—quantifies the similarity between two building polygonal shapes and match them by means of a weight-based classification procedure using the polygon's low-level feature descriptors and AI tools. Specifically, the weights assigned were calculated from a supervised training process using a GA [28,29]. The use of a GA allowed us to perform the categorization of the matching quality from a quantification of the similarity by means of a match accuracy value (MAV) ranging from zero to one (see [1,18,19]). Thus, two matched building polygons will be exactly equal if the MAV achieved is equal to one. This indicator was of great relevance for our work because it allowed us to: (i) select only 1:1 corresponding building polygon pairs among all the possible correspondences (using a thresholded value of MAV), thus avoiding the acceptance of both erroneously matched polygons and unpaired polygons; and (ii) set different similarity levels between 1:1 corresponding building polygon pairs. According to the results achieved, GA proved to be an appropriate and efficient tool on automated matching procedures using low-level feature descriptors for classifying data features as similar or nonsimilar to model features and assigning probabilities to those classes. This efficiency was achieved not only from a procedural perspective but also from both time and cost point of view. Both aspects were addressed in [18,19]. In the specific case of runtime, our matching tests revealed that for GDBs composed of between 2000 and 2500 polygons the average runtime reached was 150 s. This runtime must be multiplied by 110 in the case of a manual matching procedure.

1.2. Research Approach

Despite the positive and encouraging results above mentioned, there is a key issue that still needs to be analyzed in order to make automation a fully operational solution in matching techniques: to explore its capabilities in comparison with human capabilities.

There are many studies that analyze the way in which closed polygonal shapes are perceived by the human perceptual system and which are crucial to understanding differences in inferential processes when comparing humans and machines [2,4,30–36]. According to some of these authors, the overarching challenge in comparison studies between humans and machines seems to be the strong internal human interpretation bias [4]. In this sense, appropriate analysis tools such as AI tools and their training procedures help rationalize the interpretation of findings and put this internal bias into perspective. All in all, care has to be taken not to impose our human systematic bias when comparing human and machine perception [4].

The present study has been developed in response to the need to explore the real capabilities of our machine-based matching approach to quantify the similarity between two geographically referenced polygonal features (building polygons) and then match them. Thus, although our automated tool could emulate human operators, it is necessary to analyze its performance compared to them within a specific application and using concrete data. Specifically, using a GA, a group of human experts and two official cartographic databases, a comparison between a manual and an automated one matching procedure has been carried out. To this end, two different experiments have been developed: (i) to compare the similarity values between building polygons assigned by both the GA and the human experts; and (ii) to compare the matching procedure developed in both cases.

2. Materials and Methods

In order to improve the robustness of the results of the comparative approach applied in this study, we have used the same typology of data employed in the automated matching methodology developed in [1], that is to say, the locations of building polygons extracted from two official cartographic databases in Andalusia (southern Spain) (Figure 2a) with different levels of accuracy and which cover the same area: (i) the BCN25 (“Base Cartográfica

Numérica E25k”), the tested dataset, and (ii) the MTA10 (“Mapa Topográfico de Andalucía E10k”), the reference dataset (Figure 2b).

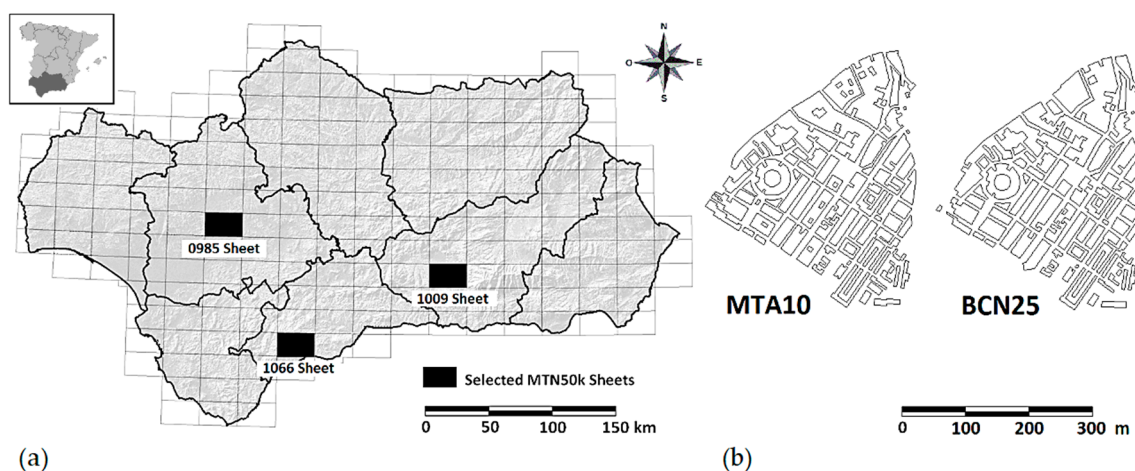


Figure 2. (a) Selected sheets of the National Topographic Map E50k of the region of Andalusia (southern Spain) and its administrative boundaries, and (b) examples of polygonal features belonging to Mapa Topográfico de Andalucía E10K (MTA10) and Base Cartográfica Numérica E25k (BCN25). Source: [1].

The MTA10 is produced by the Institute of Statistics and Cartography of Andalusia (Spain) and referenced to the ED50 datum. It is a topographic vector database with complete coverage of the Andalusian territory, which is obtained by manual photogrammetric restitution. As mentioned above, the MTA10 includes a vector layer of buildings (city blocks), which contains a sufficient quantity of geometrical information to be able to compute both the shape and geometric measures employed for assessing the geometric form of polygons. On the other hand, the BCN25 is produced by the National Geographic Institute of Spain and referenced to the European Terrestrial Reference System 1989 ETRS89 datum. This dataset covers the whole national territory of Spain—which logically includes the Andalusian territory. As in the previous case, the BCN25 is presented as a set of vector covers distributed by layers, including a vector layer of buildings (city-blocks) that contains the same type of geometrical information as the MTA10, thus allowing us to determine the degree of similarity between both data sets at building polygons level.

Finally, both databases must be interoperable, which means that they must be comparable both in terms of reference system and cartographic projection. In addition, they must be independently produced and neither of them, in turn, can be derived from another cartographic product of a larger scale through any process [37].

After describing the urban GDBs and their requirements, we must formally define the specific data used in the present study. These data were composed of pairs of building polygons (city-blocks) matched according to the accuracy criteria described in the next section and obtained from nine urban areas included in three sheets of the MTN50k (National Topographic Map of Spain at scale 1:50,000) (Figure 2a) in compliance as far as possible with all the variability requirements with regard to the typologies of buildings, that is to say, a wide geographical distribution and a large variety in their typology. In addition, we must note that these data were different from those employed to carry out the supervised training procedure of the GA.

2.1. Automatic Matching Process by Means of a GA

As mentioned in the introduction section, our automated matching methodology quantified the similarity between two building polygonal shapes and matched them by means of a weight-based classification methodology using the polygon’s low-level feature descriptors and a GA. Figure 3 shows the workflow of the process.

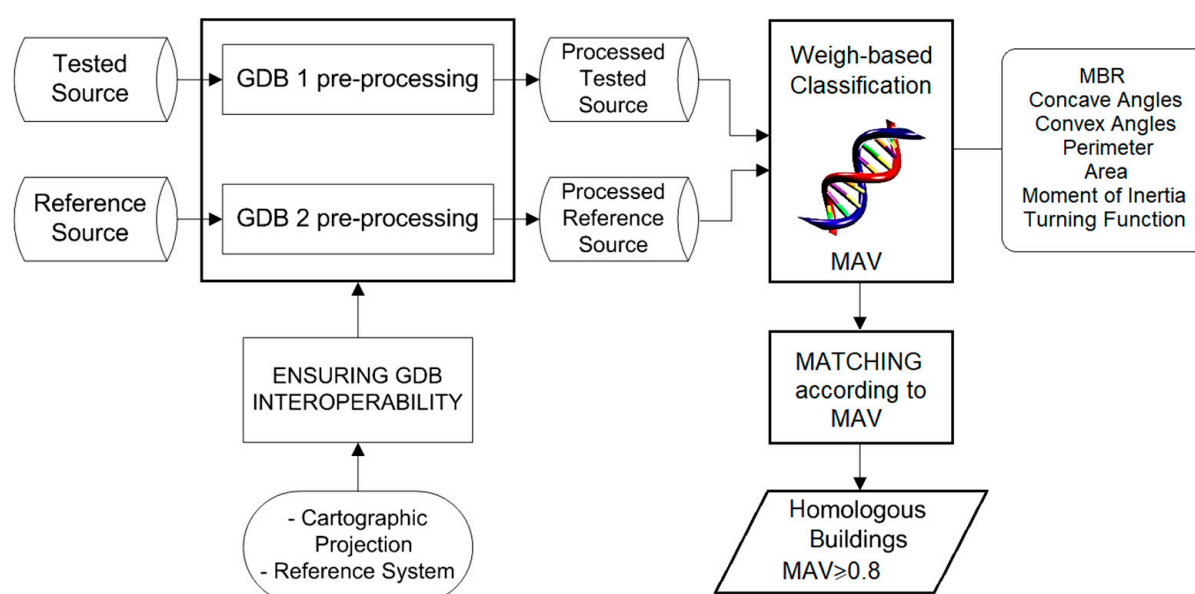


Figure 3. Workflow of the automated matching procedure.

Specifically, each building polygonal shape was characterized by means of its minimum bounding rectangle (MBR), defined by the coordinate values of its corners, number of angles (concave and convex), perimeter, area, moment of inertia and the area of the region below its turning function (see [1]). With regard to the GA, we used a GA based on real number representation (called real coded GA (RCGA)). Once again, we must note that all these aspects are addressed in detail in [1], so they will be not discussed here. There is however an important issue for the development of the present study: how to perform the discretization of the MAV values obtained. As stated previously, for the categorization of the matching quality the use of an RCGA allowed us to quantify the similarity by means of a MAV ranging from zero to one. However, it is necessary to discretize these results, defining ranges—adapted to the visual perception limit-based human discrimination capability—which allow a human operator to develop an external evaluation of such results. In this sense, the pairs of building polygons matched were classified into three different levels of similarity: low level ($MAV < 0.5$) (bad matching results), middle level ($0.5 \leq MAV < 0.8$) (intermediate matching results) and high level ($MAV \geq 0.8$) (good matching results). These thresholds between the levels were computed by assigning a confusion matrix to the GDB used to train the GA. All of them represent the shape similarity measure between two matched polygons—1:1 correspondences—regardless of the matching success. However, the confusion matrix showed that (i) a MAV greater than or equal to 0.8 allowed to avoid the acceptance of erroneously matched polygons (error of commission), (ii) the absence of this type of error was not guaranteed for values of MAV between 0.8 and 0.5 although its occurrence was unlikely, and (iii) a MAV lower than 0.5 did not guarantee the acceptance of erroneously matched polygons with a high degree of reliability [1]. Finally, and with regards to unmatched polygons (1:0 correspondences), these types of cardinality in the matching process were classified as unpaired or unmatched.

2.2. Manual Matching Carried Out by an Expert Group

Our comparative approach requires carrying out a manual matching procedure—expert conducted matching. To that end, first it is important to analyze the way both human and machines learn relationships between visual shapes. In this regard, a controlled experiment named synthetic visual reasoning test (SVRT) was developed in [2]. This SVRT was composed by 23 classification problems based on abstract reasoning and evaluated by 20 human experts and a machine-learning technique. By this method, these authors worked with planar random shapes and the classification was carried out at level of relationships,

such as “inside”, “in between”, “same”, etc. Specifically, one of the problems belonging to category 1 of their SVRT was about how to solve the question: Are there two identical shapes in the image? (Figure 4). This problem is closely related with matching procedures and, as will be seen below, it is the main goal of one of our experiments.

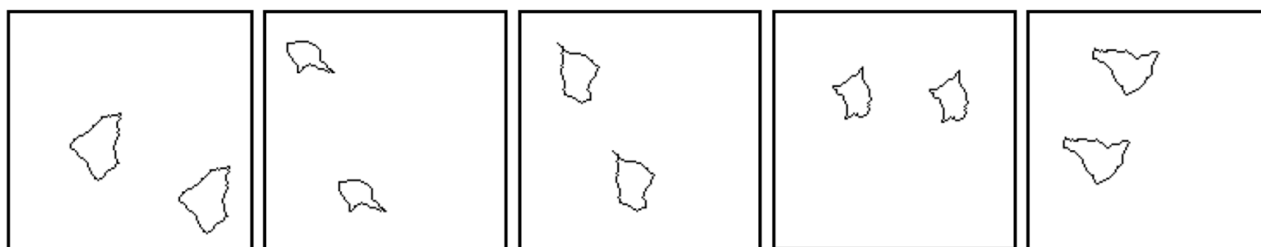


Figure 4. Shapes belonging to category 1 of the SVRT. Are there two identical shapes in the image? Source: [2].

With regard to the results, these authors obtained two important conclusions for our work: (i) humans needed only few examples to learn relationships between shapes and to solve a very high percentage of the problems, while the machine learning algorithm needed many more to reach a lower percentage; (ii) similarity criteria are easier to learn in complex shapes than in simple shapes. The first conclusion suggests that humans and their expert knowledge may be, a priori, a valuable tool for assessing an automated matching procedure which is—as mentioned in the next subsection—one of the objectives of our comparison process and the main hypothesis of the present work. With regard to the second conclusion, it suggests us a parallel hypothesis which must be also addressed in our study: similarity is easier to assess in complex shapes than in simple shapes.

After analyzing learning processes in humans and machines, there are three key aspects for defining manual matching procedures: (i) the selection of a representative sample of the elements for their evaluation (in our case, building polygons, city-blocks, that have been previously matched automatically by the GA), (ii) the selection of the group of experts that will carry out the procedure, and (iii) the definition of what has been termed agreement measures. These measures represent numerical values that can be used to analyze the efficiency of the results obtained from the proposed procedure. With regard to this last aspect, the measures can be classified into the following two categories:

- Level of agreement. Defined as the degree of similarity between the judgments issued independently by the different experts;
- Consistency. Defined as the degree of similarity between the judgments issued by the experts and the results provided by the automatic matching process (GA).

2.2.1. Objectives of the Comparison Process

As a first step, it was necessary to formulate the main objectives pursued by carrying out the experiments with the group of experts, which are:

- Provide an alternative MAV (expert value) to the MAV provided by the GA assessing the similarity between building polygons;
- Provide an alternative element matching to those provided by the GA assessing the level of effectiveness of the automatic matching process.

2.2.2. Selection of the Group of Experts

Expert knowledge plays an integral role in many fields of science. Thus, where empirical data are scarce or unavailable, expert knowledge is often regarded as the best or only source of information [38]. In addition, with the rise of machine learning procedures the role of expert knowledge has become ever more important. A deeper understanding of the mechanisms of human perception in order to improve machine learning procedures will require appropriate consideration of expert opinions. Therefore, selecting experts and

eliciting their opinions must be performed and handled carefully, with full recognition of the uncertainties inherent in those opinions [39].

As in other sciences, expert knowledge is very important for GIScience. However, although expert knowledge has long been a value input into geographical inquiry, in the context of automatic PAA procedures it has not been explored in detail. Traditional approaches to the PAA of GI are dominated by a paradigm borrowed from transactional data architectures where discrepancy metrics can be formulated easily [40]. Unfortunately, this approach has limited utility for automatic PAA procedures, where these metrics are highly dependent on the matching techniques previously employed. That is why these techniques must be subjected to external assessment processes.

There are many methods to select a panel of experts, highlighting quantitative selection methods. They are based on the calculation of coefficients of competence. In our case, the experts were selected employing this method and on the basis of their professional qualification, experience and prestige. With regard to the size, using a group as numerous as possible allowed us to improve the consistency of the results as well as to assess human consistency. Thus, we counted on the active participation of 24 experts from the following countries and institutions: Brazil (Universidade de Sao Paulo), Finland (National Land Survey of Finland), France (Institut Géographique National), Germany (Institut für Kartographie), Scotland (University of Edinburgh), Spain (from several institutions), Swiss (Iniversität Zurich), United States (University of Ohio) and Uruguay (Universidad ORT).

2.2.3. Working Framework and Documentation Provided

The working framework employed was strongly conditioned by the geographic spread of the participants. This circumstance did not allow us to carry out assessment procedures of a face-to-face nature, it being necessary to implement remote assessment methods. There are two basic ways to apply this type of method [38]: by analogue support (the traditional way) or by digital support. In our case, this last option was applied.

In comparison with traditional methods, assessment methods based on digital support have experienced an exceptional growth during recent years through the development of information and communication technologies (ICTs) and web services-based technologies (WSTs). Thus, the capacity to interact across the network provided by these technologies allows us to share resources such as data, processing modules and applications quickly and efficiently. WSTs are used widely in the geospatial domain, expanding the traditional focus on discovery and access to geospatial data. Geospatial WSTs are designed to integrate, edit and store a large amount of geospatial information and their corresponding metadata, promoting collaborative working and fulfilling users' requests [41]. However, our working framework does not include a volume of information large enough to implement a web service whose architecture is defined to support complex data infrastructures. With this in mind, we opted to develop an application which is easy to use in order to make the experts' work easier and to obtain the highest possible number of assessments. Our application was implemented in Visual-Basic.NET, and hosted in a virtual machine. Finally, access to the virtual machine—and therefore to our application—was carried out by means of a remote desktop connection. This way of working allowed us to enjoy the advantages of WST, avoiding the drawbacks associated with its architecture. Some of these advantages are:

- Simultaneous multiple access so that all experts can enter into the application at any time;
- Access restriction. This restriction was established at two different levels: (i) access restriction to any other type of information or software stored in the host computer—with any another way of working, this information would be totally exposed; and (ii) access restriction to the user accounts of other experts.

Finally, and together with the aforementioned application, the group of experts was provided with a matching guide consisting of a single document in which the guidelines to be followed to complete the experiments proposed were briefly presented. An important

section of this guide was composed by graphical examples of pairs of polygons with different degrees of similarity (Figure 5).

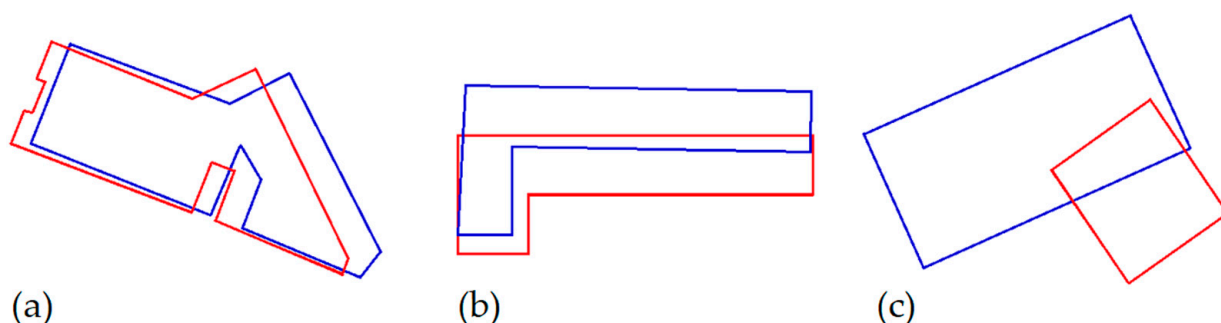


Figure 5. Degree of similarity between each pair of polygons previously matched. Examples belonging to our matching guide. Polygons from the BCN25 are represented by blue contours, while polygons from the MTA10 are represented by red contours. (a) High level of similarity; (b) middle level of similarity; and (c) Low level of similarity.

With this document we aimed to guide the experts in carrying out the different tasks. Since the final success of the evaluation will depend on its perfect understanding its writing must be clear and concise. This document was hosted together with the application in the virtual machine which can be retrieved through remote desktop connection.

2.2.4. Design of the Experiments

In order to accomplish the objectives proposed in Section 2.2.1, the assessment procedure was composed of two different experiments. An important aspect to take into account for their design was the time required to complete them. In this respect, there are two requirements that must be fulfilled: (i) the time required should be as short as possible, without affecting the achievement of the objectives, and (ii) the total number of cases to be evaluated by each expert should be as small as possible without affecting the significance of the results. According to these requirements, it was estimated that the time required should not exceed in any case 20 min. This is the maximum time interval recommended because the user's attention and therefore the quality of the response decreases after 20 min [42]. In addition and according to these last authors, it is very difficult to get a user to stay online for more than 20 min by filling out a questionnaire. Logically, this response time conditioned the final number of polygons used, which in any case were enough to reach statistically significant results.

MAV Experiment

This first experiment was applied over a sample of 18 pairs of homologous building polygons. In this case, each expert should provide an alternative MAV to the MAV provided by the GA when assessing the similarity between building polygons during the automatic matching process. To that end, two graphic windows were used so that each of the two building polygons was shown in a different graphic window (Figure 6a). After a visual check, the MAV provided by each expert could be included into the three discretized levels: low, medium and high level—depending on the degree of similarity between the polygons matched. The pairs of building polygons used were selected so that all the possible values of MAV were represented and in compliance, as far as possible, with all the variability requirements with regard to the typologies of polygons considering the variability with regard to the number of vertices. The use of this parameter allowed us to characterize intuitively the results, determining if there is some type of geometry to be especially problematic when assessing the degree of similarity between polygons.

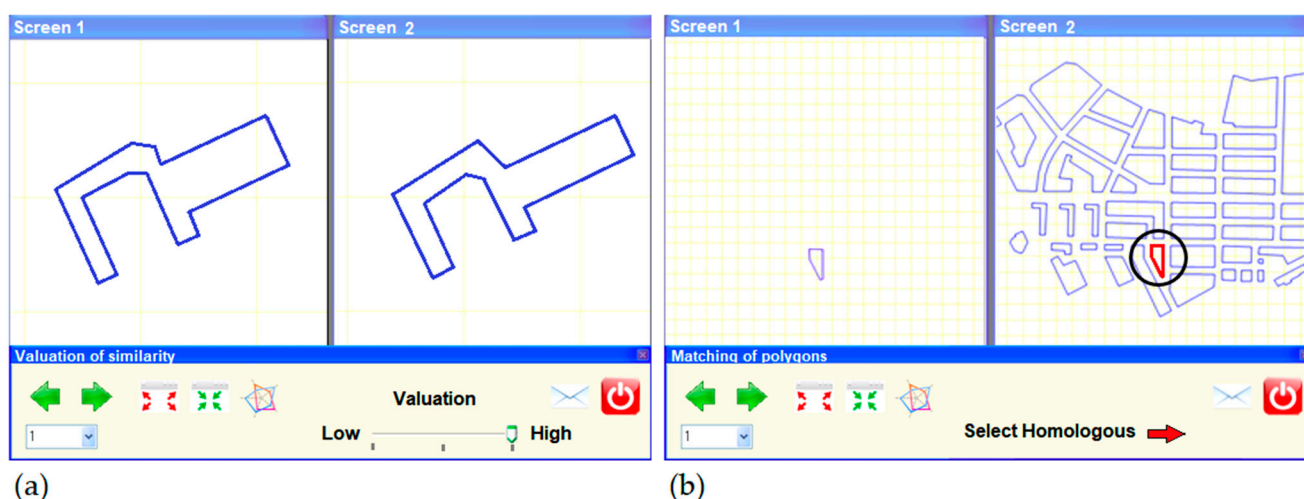


Figure 6. Graphical user interface. (a) Interface belonging to the MAV experiment; (b) interface belonging to the matching experiment.

In order to ensure the robustness of the results corresponding to this first experiment, it was necessary to transfer to the experts some of the results provided by the GA, explaining to them that not all descriptors used have the same importance in the weight-based classification methodology developed by the GA. The information related to this set of weighting criteria was provided to experts through various examples, ensuring that all of them give the same importance to each of the descriptors, and in this way it will be possible to homogenize the answers.

Matching Experiment

This second experiment was also applied over a sample of 18 pairs of homologous polygons. These polygons must be different to those belonging to the MAV experiment. The goal of this second experiment was to assess the efficiency achieved by the GA related to the tasks of identification and matching of homologous building polygons. As in the previous experiment, two graphic windows were used (Figure 6b). In this case, the left window shows the polygon whose homologue must be found by the expert in the right window, thus providing an alternative element matching to those provided by the GA assessing the level of effectiveness of the automatic matching process. The work of the experts in this case was of great simplicity; therefore, it was not necessary to provide them with any other type of additional information. However, in this case, visual assessment plays a very important role. That is why several editing options, such as zoom tools, grid system, buttons for overlapping the polygons according to their coordinates, etc., were implemented.

Experimental Design

With regard to the experimental design, the sample of building polygons included a total of 54 pairs of polygons for each of the two experiment types (MAV and matching), with which four different tests were performed. For the first three tests (numbers 1, 2 and 3), 18 pairs of different polygons were employed ($18+18+18=54$). These tests were assessed by a random selection of 15 experts from those included in the initial group. Test number 4 was performed with 18 pairs of polygons selected from those employed in the first three tests with each of them contributing six pairs of polygons. This last test was evaluated by the total number of experts included in the aforementioned group. Finally, we must note that the goal of this type of distribution was to increase the assessing pressure on certain cases, analyzing their statistical behavior in situations with different number of experts.

3. Results

3.1. Results Derived from the MAV Experiment

Table A1 (Appendix A) presents all the results obtained through the GA for each of the tests that were carried out (1, 2, 3, and 4). This table shows both the MAV computed—ranging from zero to one—and the discretized values corresponding to these MAV (high→ 2, middle→ 1, low→ 0). These results have been summarized in Figure 7. Specifically, Figure 7a shows the percentage distribution of matched polygons for each level of MAV—as mentioned above, this distribution was homogeneous—while Figure 7b sets out the results of the percentage distribution of matched polygons for each interval of MAV grouped by number of vertexes.

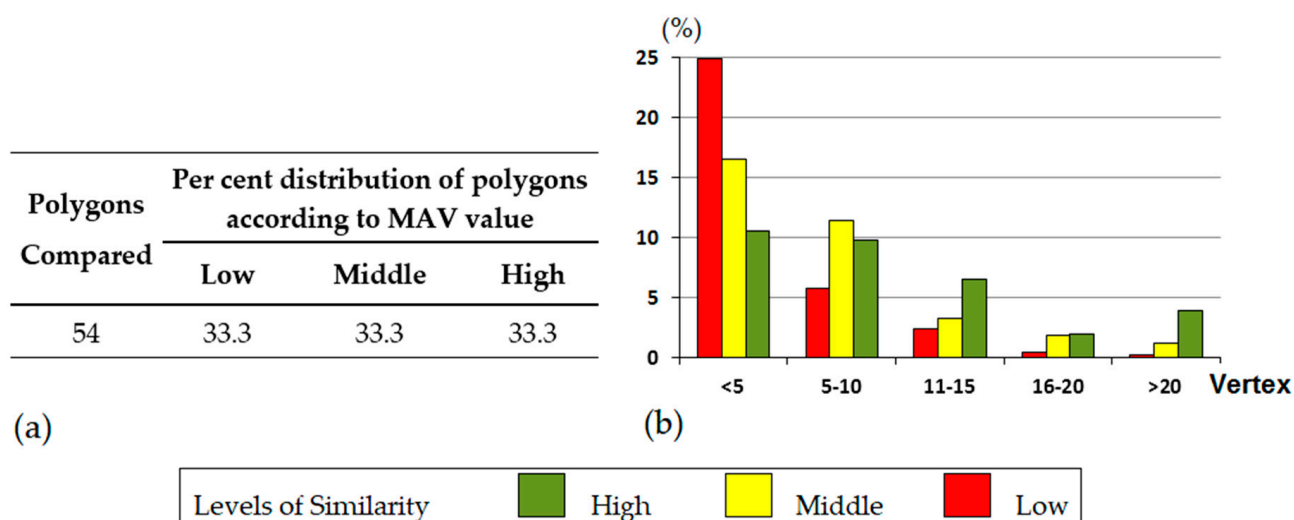


Figure 7. (a) Distribution of polygons according to the MAV computed by the GA, (b) percentage distribution of matched polygons for each level of MAV grouped by number of vertices.

With regard to the agreement results derived from the group of experts, and according to the experimental design, 612 assessments were conducted (270 assessments from the experiments 1, 2 and 3; and 342 new assessments from the experiment 4). Tables A2 and A3 (Appendix A) present the discretized MAV provided by the experts, which have also been summarized in Figure 8. For its processing, a double entry table model was used. This type of table shows all the options (C_{ij}) chosen by the n experts considered, and for the m pairs of polygons finally assessed. In addition, the MAV provided by the GA is also shown, therefore from the several C_{ij} values the levels of agreement and the consistency reached in the assessment were calculated and expressed as percentages of agreement. In order to facilitate the interpretation of the results, Figure 9 shows an example of the MAV computed by the GA (Figure 9c) and the discretized values assigned by an expert (Figure 9d) for five polygons (Figure 9b) belonging to the sample used in the present study.

3.2. Results Derived from the Matching Experiment

Table A4 (Appendix A) presents the polygon assigned as homologous in the matching process developed by the GA. With regard to the results derived from the group of experts, the tables of results have a similar structure to those of the previous section, responding to the double entry model in the data processing described above. Thus, in Tables A5 and A6 (Appendix A) each entry represents the building polygon (through its identifier) selected by the expert as homologous to another belonging to the selected sample. As in the MAV Experiment, in this experiment, 612 assignments were conducted. With regard to Table A4, it must be noted that for polygon numbers 7, 9, 13, 16, 24, 29 and 32, the GA was not able to find homologues to match these polygons, so they have been labelled as unmatched polygons. In the same way, none of the experts was able to match them (Tables A5 and A6).

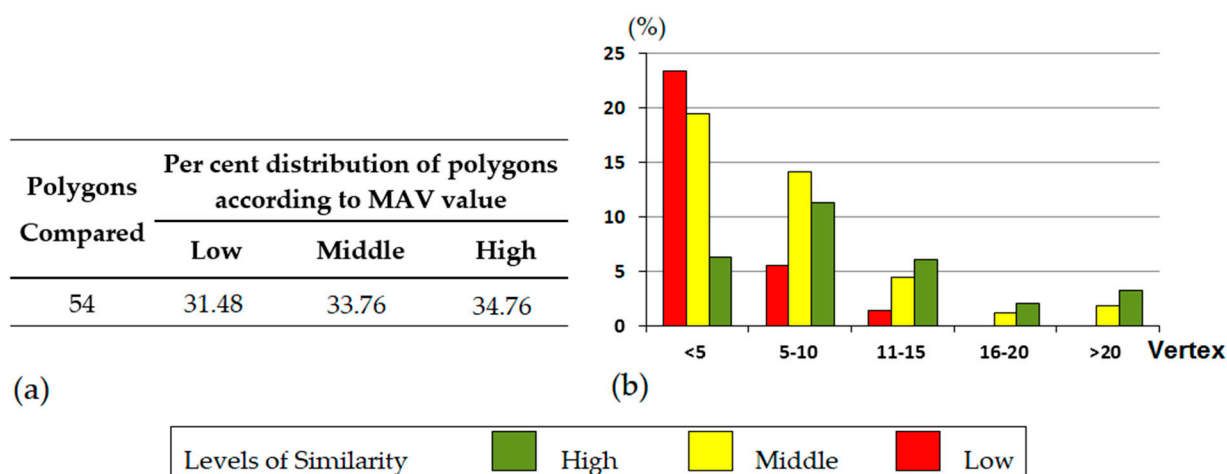


Figure 8. (a) Distribution of polygons according to the MAV computed by the group of experts, (b) percentage distribution of matched polygons for each level of MAV grouped by number of vertices.

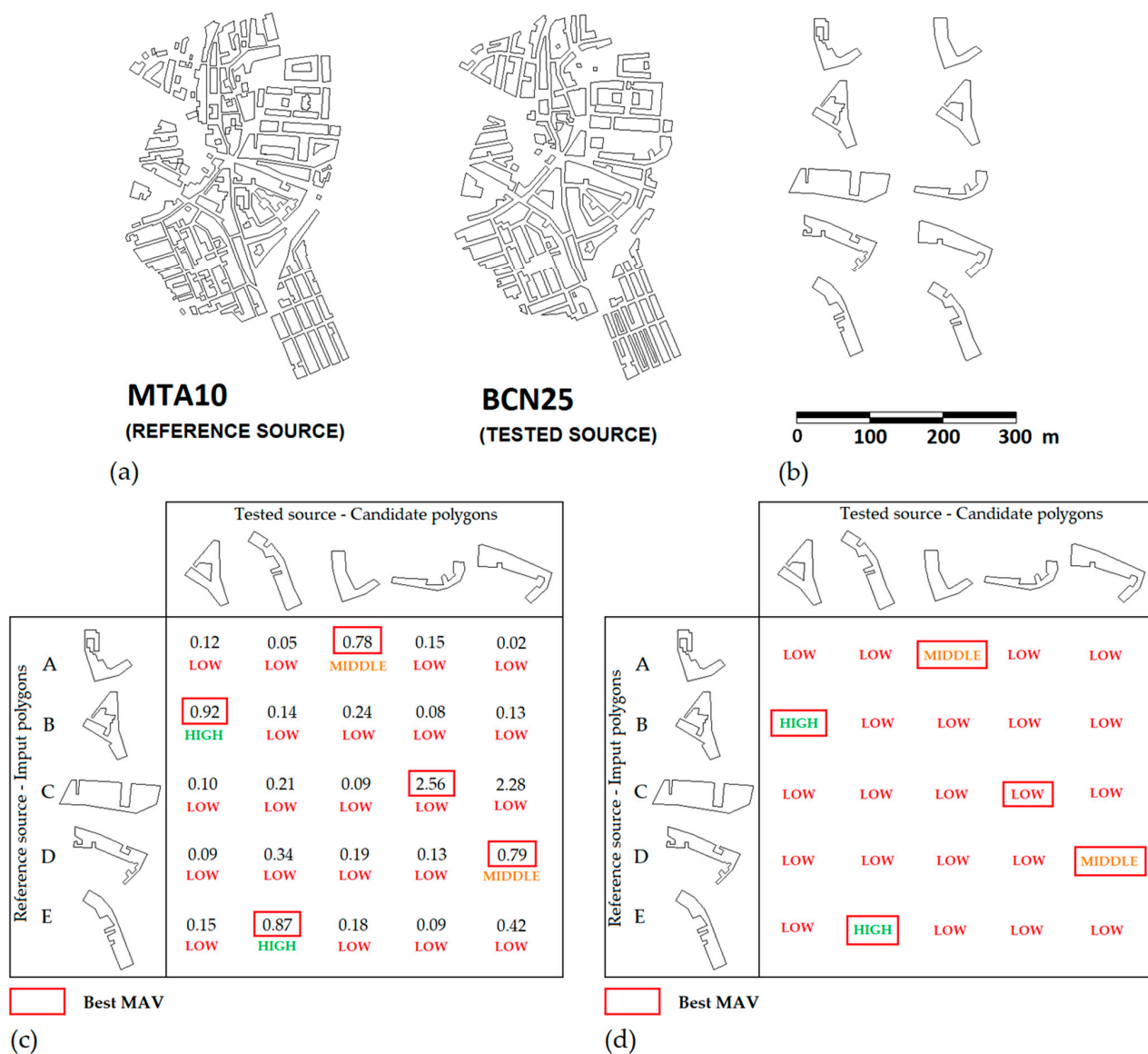


Figure 9. (a) One of nine urban areas employed in this study; (b) polygons belonging to this urban area and selected for the MAV experiment; (c) MAV computed by the GA; (d) discretized MAV assigned by a specific expert.

Unmatched polygons are associated mainly with one type of cardinality, 1:0. That is to say, polygons represented only in one of the data sources (Figure 10a). Specifically, this case corresponding to the polygon number 7 belonging to Tables A4 and A5. On the other hand, unmatched polygons can be also associated with m:n correspondences between polygons with a relatively complex contours which have been subjected to cartographic generalization processes (Figure 10b).

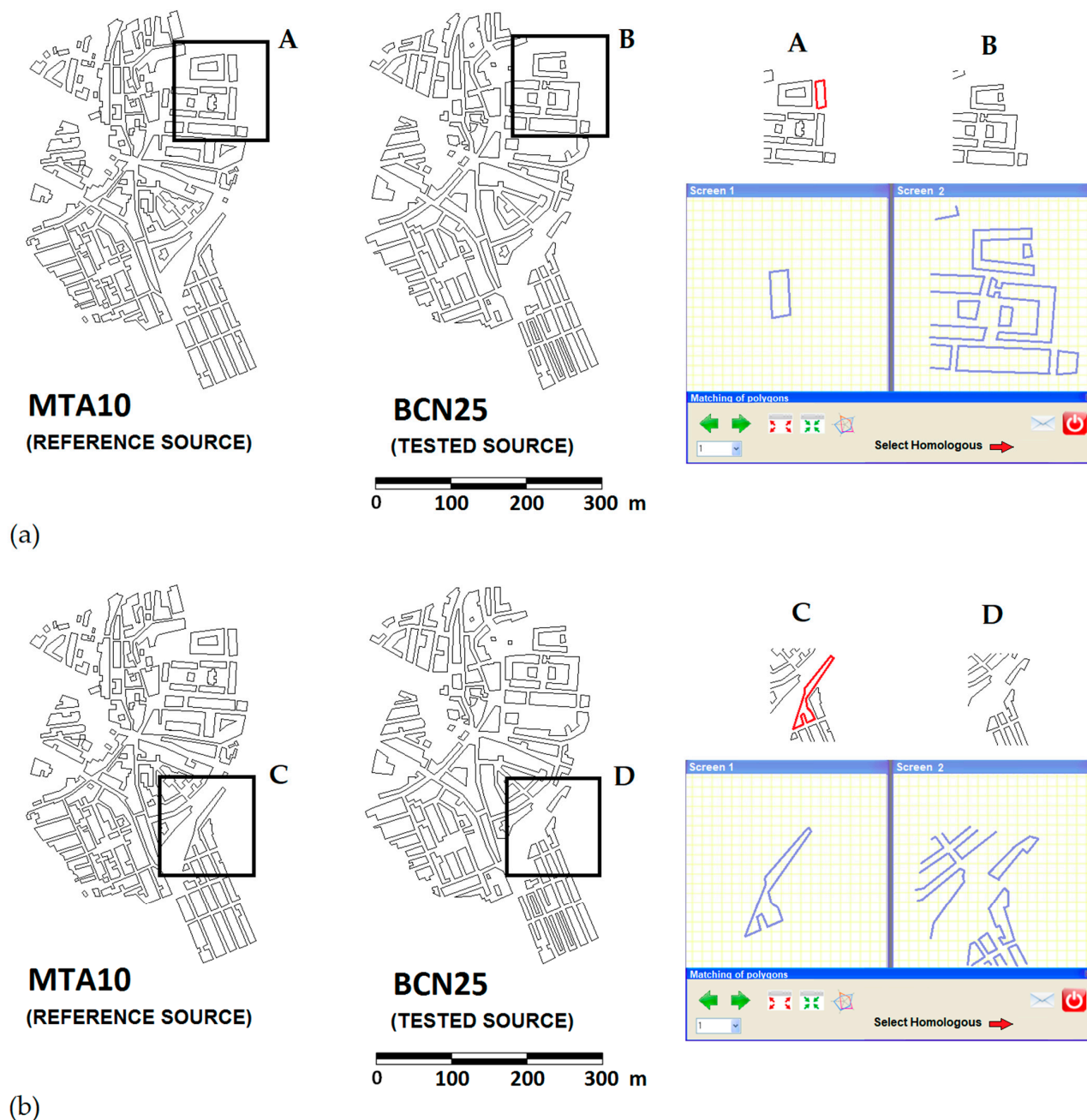


Figure 10. Unmatched polygons cases. (a) Unmatched polygons associated with 1:0 correspondences; (b) unmatched polygons associated with m:n correspondences.

4. Discussion

4.1. MAV Experiment

The bar charts from Figures 7b and 8b can help give a better interpretation of the results derived from the MAV experiment when polygons are characterized by their number

of vertices. Both bar charts show a similar behavior both in the case of the similarity assessment developed by the GA and in the case of the similarity assessment developed by the group of experts. In both cases, the proportion of low, middle and high MAV tends to be reversed as the number of vertexes increases. Specifically, the percentage of pairs of polygons assessed by the GA as having low level of similarity is 25% in the case of polygons with a number of vertexes lower than five. However, this percentage is practically nil for polygons with a number of vertices higher than twenty. This trend is even much sharper in the assessment carried out by the group of experts. In this last case, there are not any pairs of polygons with a number of vertices higher than twenty that have been assessed as low level of similarity. In the view of the above, one may conclude that both for GA and for human experts it is easier to assess the similarity between two matched polygons when they have a high number of vertex. This confirms one of the conclusions reached in [2] and verifies one of the hypothesis considered in our study, the similarity is easier to assess in complex shapes than in simple shapes. In the specific case of humans, it is due to the intuitive trend of decomposing pairs of complex objects into their constituent parts—that are individually recognizable—in order to facilitate identifying common shapes between these objects and then match them [43].

Regarding the agreement measures, it can be said that the consistency achieved for this first experiment (in all the tests performed) was very high, with a GA–experts mean agreement percentage—calculated over the total number of assessments of 612—of 93.3%, while for 75.9% of the building polygon pairs a 100% agreement percentage between the similarity results provided by the experts and those obtained through the AG was reached. Together with these high consistency percentages, the level of agreement between experts reveals an important aspect of the study; the confusion (disagreement GA–expert and expert–expert) when assessing the similarity between pairs of polygons whose MAV—computed by the GA—is close to the threshold values of 0.5 and 0.8. This type of behavior—the increase of uncertainty around threshold values—was already discussed in [44]. These authors analyzed the impact of estimation errors on reference data collected to validate land-cover maps, and concluded that these errors are more prevalent when the class proportions are close to the class definition threshold. Taking into account the particularities of our experiments—mainly related to the uncertainties inherent in the experts’ opinions—and with all due caution, one might argue that our case is similar to that addressed in [44]. Thus, Figure 11a shows that the lowest values of agreement correspond to pairs of polygons whose MAV is close to the threshold values of 0.5 and 0.8. Specifically, the grey line represents the mean value of agreement for each interval of 0.1 MAV units. On the other hand, Figure 11b represents the agreement reached in relation to the distance to the MAV threshold (DMT). In this last case, each blue point represents a pair of matched polygons, of which similarity was assessed with a certain level of agreement between the GA and the experts, and its distance to the nearest threshold value—0.5 or 0.8. Here the grey line represents the trend line followed by the data.

As a solution to minimize this type of error, the authors propose to carefully select the classification system with threshold values as far as possible from the modes of the distributions of land-cover proportions [44]. However, in our case this way to proceed makes no sense. We must recall that the location of the threshold values is already optimized by a confusion matrix whose aim is to avoid the acceptance of erroneously matched polygons (error of commission).

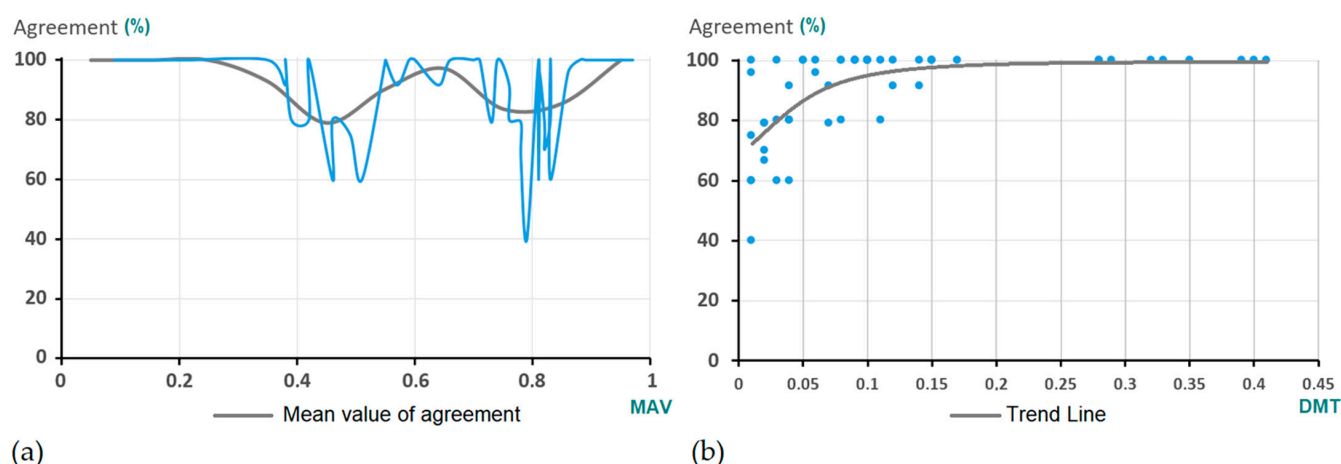


Figure 11. Uncertainty around threshold values; (a) level of agreement according to the MAV; (b) agreement according to the distance to the MAV threshold.

With regard to the specific cases, the polygons with highest level of disagreement in Figure 11 correspond to the numbers 1, 2, 9, 20, 21, 28, 31, 48 and 49 (Tables A2 and A3). Specifically, the cases of the polygons number 9 and 49 are particularly remarkable. In the case of polygon number 9, the MAV assigned by the GA is 0.79—just below the threshold of 0.8. This case corresponding to the polygon labeled as D in the Figure 9c. The case of polygon number 49 is very similar. In this case, which correspond to the polygon labeled as A in the Figure 9c, the MAV assigned by the GA is 0.78—also close to the threshold 0.8.

On the other hand, and with regard to the variability of the assessments provided—which represents the level of agreement between the judgments issued independently by the experts—it should be noted that in none of the pairs of polygons were the three types of values (0, 1 or 2)—belonging to MAV assessment—simultaneously assigned by the experts.

Finally, the results obtained increasing the evaluation pressure on certain cases (test 4) showed certain variability in the percentages of agreement reached in the following polygons:

- Polygon number 1. This polygon reached 100% agreement in test number 1, while in the number 4 it did not reach 80% (79.1%).
- Polygon number 4. In this case, the level of agreement reached decreased from 100% (test 1) to 91.6% (test 4).
- Something similar happened with polygons number 48 and 49. Both reached 100% agreement in test number 3, while in test number 4 they dropped to less than 80%.

As mentioned above, these last cases corresponded to building polygons whose degree of similarity is close to the two threshold values, which define the three different levels, established after the MAV discretization process.

4.2. Matching Experiment

The consistency achieved for this second experiment was even higher than for the case of the first one, with a GA–expert mean agreement percentage of 98.8%—calculated over the total number of assignments, 612. In this case the results were very strong, with a percentage of agreement of 100% in most cases. Even polygons that the AG was unable to match were recognized by experts as unmatched. However, there are striking exceptions. Such is the case of expert number 21—in the specific case of polygon number 9—and particularly the case of expert number 11. According to Table A6, this last expert was unable to assign a homologous element to the following polygons: 10, 26, 28, 31, 50 and 52. All these polygons are polygons with a square or rectangular shape for which it is helpful to employ the edition tools—such as the grid system or the button for overlapping the polygons according to their MBRs—in order to facilitate the assignment procedure. Therefore, the fact of not being able to carry out any assignment might be due to either some misinterpretation of the matching guide or not to use it.

Figure 12 presents an example that illustrates this situation: an urban area where predominates a constructive typology characterized by rectangular city-blocks. Specifically, it corresponds to the case of the polygon number 10. First, Figure 12a shows the graphical user interface used for assigning the homologous polygon (from the screen 2) corresponding to the polygon shown in the Screen 1. In this case, the assignment procedure can be very difficult if edition tools are not employed. Figure 12b shows the usefulness of these tools. Specifically, we have used the overlapping button. By this way, the uncertainty in the assignment is reduced significantly. Accordingly, and as discussed in Section 4.1, it can be concluded that it is not only easier to assess the similarity between two matched polygons when they have a high number of vertex but also is easier to match them. This reconfirms the conclusions reached in [2], and verifies the hypothesis considered in our study.

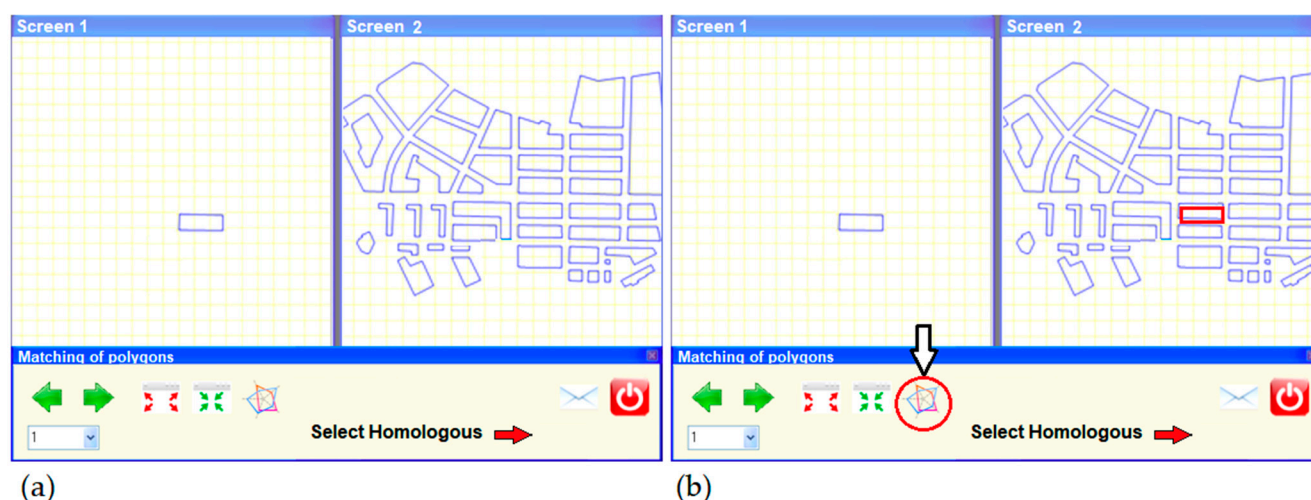


Figure 12. Graphical user interface belonging to the matching experiment. (a) Example of assignment without the help of the edition tools; (b) example of assignment using the edition tools.

Finally, the results obtained increasing the evaluation pressure on certain cases (test 4) showed certain variability in the percentages of agreement reached in the following polygons:

- Polygon number 31. This polygon reached 100% agreement in test number 2, while in test number 4 it did not reach 87.5%.
- Polygon number 52, whose level of agreement remained constant in tests 3 and 4. In addition, this case reached the lowest agreement level (79.1%) in test number 4. This polygon belongs to the polygons' typology described above, that is to say, polygons with a square or rectangular shape for which is advisable to employ the edition tools. As stated, the fact that the experts 2, 5, 7, 9 and 11 were not able to match it might be due to a failure to follow the guidelines established by the matching guide.

5. Conclusions

The present study explores the real capabilities of a machine-based matching approach to assess the similarity between two geographically referenced polygonal features (building polygons) and then match them. To that end, we selected a group of experts—composed of professionals of recognized prestige and with lengthy professional careers in the field of cartography—whose main objective was to provide alternative assessments (of similarity and matching) to those provided by our automated matching mechanism (GA). In both cases, our results show high levels of consistency, that is to say, a high degree of agreement between the judgments issued by experts and the results provided by the GA. Specifically, 93.3% in the case of the MAV experiment (similarity assessment) and 98.8% in the case of the matching experiment (homologous polygon assignment). In addition, the characterization of the pairs of polygons according to the number of vertices allowed us to corroborate the hypothesis that, in general, the similarity is easier to assess in complex shapes than in

simple shapes. It might be related with the fact that in complex shapes is easier to identify what is known as principles of perceptual organization [2], such as proximity, similarity, symmetry, inclusion, collinearity, etc.

However, and despite the high levels of consistency aforementioned, there are two aspects that must be highlighted: (i) the cases of disagreement between GA and expert and expert and expert are strongly linked to pairs of polygons whose MAV is close to the stipulated threshold values; and (ii) the matching experiment has revealed that human operators encounter difficulties in matching polygons with square or rectangular shapes when the geolocation of these polygons is not taken into consideration. Both aspects underline the dependence of the results on both the prevailing constructive typology in a certain urban GDB, and the experimental design itself. In this sense, efforts must lead towards the improvement of the experiments by a significant increase in both the number of polygons used in the study, and the number of experts who assess their similarity and match them.

In any case, and being careful with the interpretation of the results, we can conclude that the main hypothesis of the present work has been corroborated. That is to say, GAs have proved to be an efficient tool when assessing the similarity between polygonal features, and therefore in automated matching procedures using low-level feature descriptors.

Finally, and despite the guarantees provided by expert knowledge in general and by our group of experts in particular, our study does not cover an entire experimentation with automation tools, but confines itself to the use of one tool in particular (GAs). In addition, it has been applied on a specific spatial data set. In this sense, more studies using other tools of AI and other case studies are required to further understand the mechanisms of human perception and to improve machine learning procedures.

Author Contributions: Conceptualization, Francisco Javier Ariza-López; Methodology, Juan José Ruiz-Lendínez and Francisco Javier Ariza-López; Software, Juan José Ruiz-Lendínez and Manuel Antonio Ureña-Cámara; Investigation, Juan José Ruiz-Lendínez; Writing-original draft preparation, Juan José Ruiz-Lendínez; Writing-review and editing, Juan José Ruiz-Lendínez; Supervision, Juan José Ruiz-Lendínez, Francisco Javier Ariza-López and Manuel Antonio Ureña-Cámara; Project administration, Juan José Ruiz-Lendínez. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data also forms part of an ongoing study.

Acknowledgments: Authors acknowledge experts for accepting to take part in this study.

Conflicts of Interest: The authors declare no conflict of interest.

Table A1. Results provided by the GA. MAV Experiment.

| Test 1 | | | Test 2 | | | Test 3 | | | Test 4 | | |
|--------|------------------------------|--------|--------|------------------------------|--------|--------|-------------------------------|--------|--------|------------------------------|--------|
| Pol. | MAV Computed/ Discreticed | | Pol. | MAV Computed/ Discreticed | | Pol. | MAV Com- puted/Discreticed | | Pol. | MAV Computed/ Discreticed | |
| 1 | 0.82/2 | High | 19 | 0.17/0 | Low | 37 | 0.83/2 | High | 1 | 0.82/2 | High |
| 2 | 0.51/1 | Medium | 20 | 0.46/0 | Low | 38 | 0.09/0 | Low | 4 | 0.38/0 | Low |
| 3 | 0.71/1 | Medium | 21 | 0.39/0 | Low | 39 | 0.10/0 | Low | 5 | 0.55/1 | Medium |
| 4 | 0.38/0 | Low | 22 | 0.88/2 | High | 40 | 0.81/2 | High | 6 | 0.74/1 | Medium |
| 5 | 0.55/1 | Medium | 23 | 0.81/2 | High | 41 | 0.15/0 | Low | 7 | 0.86/2 | High |
| 6 | 0.76/1 | Medium | 24 | 0.70/1 | Medium | 42 | 0.57/1 | Medium | 8 | 0.11/0 | Low |
| 7 | 0.86/2 | High | 25 | 0.22/0 | Low | 43 | 0.90/2 | High | 22 | 0.88/2 | High |
| 8 | 0.11/0 | Low | 26 | 0.64/1 | Medium | 44 | 0.66/1 | Medium | 23 | 0.81/2 | High |
| 9 | 0.79/1 | Medium | 27 | 0.83/2 | High | 45 | 0.76/1 | Medium | 25 | 0.22/0 | Low |
| 10 | 0.83/2 | High | 28 | 0.49/0 | Low | 46 | 0.95/2 | High | 26 | 0.64/1 | Medium |
| 11 | 0.91/2 | High | 29 | 0.90/2 | High | 47 | 0.95/2 | High | 28 | 0.49/0 | Low |
| 12 | 0.89/2 | High | 30 | 0.60/1 | Medium | 48 | 0.38/0 | Low | 31 | 0.78/1 | Medium |
| 13 | 0.42/0 | Low | 31 | 0.78/1 | Medium | 49 | 0.78/1 | Medium | 43 | 0.90/2 | High |
| 14 | 0.59/1 | Medium | 32 | 0.46/0 | Low | 50 | 0.81/2 | High | 44 | 0.66/1 | Medium |
| 15 | 0.18/0 | Low | 33 | 0.97/2 | High | 51 | 0.42/0 | Low | 48 | 0.38/0 | Low |
| 16 | 0.90/2 | High | 34 | 0.82/2 | High | 52 | 0.35/0 | Low | 49 | 0.78/1 | Medium |
| 17 | 0.21/0 | Low | 35 | 0.73/1 | Medium | 53 | 0.70/1 | Medium | 50 | 0.81/2 | High |
| 18 | 0.22/0 | Low | 36 | 0.74/1 | Medium | 54 | 0.55/1 | Medium | 52 | 0.35/0 | Low |

Table A2. Results provided by the experts vs GA. MAV Experiment (Tests 1, 2 and 3).

| Test 1 | | | | | | | | Test 2 | | | | | | | | Test 3 | | | | | | | |
|---------|----|------------|---|---|----|---|---------------|---------|----|------------|---|---|----|----|---------------|---------|----|------------|---|---|---|----|---------------|
| Pol ID. | GA | Expert ID. | | | | | Agreement [%] | Pol ID. | GA | Expert ID. | | | | | Agreement [%] | Pol ID. | GA | Expert ID. | | | | | Agreement [%] |
| | | 16 | 5 | 2 | 21 | 8 | | | | 12 | 9 | 1 | 14 | 15 | | | | 3 | 4 | 7 | 6 | 24 | |
| 1 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 37 | 2 | 2 | 2 | 2 | 2 | 2 | 100 |
| 2 | 1 | 0 | 1 | 1 | 1 | 0 | 60 | 20 | 0 | 1 | 0 | 0 | 1 | 0 | 60 | 38 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 100 | 21 | 0 | 0 | 1 | 0 | 0 | 0 | 80 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 22 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 40 | 2 | 1 | 2 | 2 | 1 | 2 | 60 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 100 | 23 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 41 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| 6 | 1 | 2 | 1 | 1 | 1 | 1 | 80 | 24 | 1 | 1 | 1 | 1 | 1 | 1 | 100 | 42 | 1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 7 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 43 | 2 | 2 | 2 | 2 | 2 | 2 | 100 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 26 | 1 | 1 | 1 | 1 | 1 | 1 | 100 | 44 | 1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 9 | 1 | 1 | 2 | 2 | 1 | 2 | 40 | 27 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 45 | 1 | 1 | 1 | 2 | 1 | 1 | 80 |
| 10 | 2 | 2 | 2 | 1 | 2 | 2 | 80 | 28 | 0 | 0 | 0 | 1 | 0 | 0 | 80 | 46 | 2 | 2 | 2 | 2 | 2 | 2 | 100 |
| 11 | 2 | 1 | 2 | 2 | 2 | 2 | 100 | 29 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 47 | 2 | 2 | 2 | 2 | 2 | 2 | 100 |
| 12 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 30 | 1 | 1 | 1 | 1 | 1 | 1 | 100 | 48 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| 13 | 0 | 0 | 1 | 0 | 0 | 0 | 80 | 31 | 1 | 1 | 1 | 2 | 1 | 1 | 80 | 49 | 1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 14 | 1 | 2 | 1 | 1 | 1 | 1 | 100 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 50 | 2 | 2 | 2 | 2 | 2 | 2 | 100 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 33 | 2 | 2 | 2 | 2 | 2 | 2 | 100 | 51 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| 16 | 2 | 3 | 2 | 2 | 2 | 2 | 100 | 34 | 2 | 2 | 2 | 2 | 2 | 1 | 80 | 52 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 35 | 1 | 2 | 1 | 1 | 1 | 1 | 80 | 53 | 1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 36 | 1 | 1 | 1 | 1 | 1 | 1 | 100 | 54 | 1 | 1 | 1 | 1 | 1 | 1 | 100 |

Table A3. Results provided by the Experts vs GA. MAV Experiment (Test 4).

[illegible]

Table A4. Results provided by the GA. Matching experiment

| Test 1 | | Test 2 | | Test 3 | | Test 4 | |
|----------|---------------------|----------|---------------------|----------|---------------------|----------|---------------------|
| Pol. ID. | Matched Polygon ID. | Pol. ID. | Matched Polygon ID. | Pol. ID. | Matched polygon ID. | Pol. ID. | Matched Polygon ID. |
| 1 | 247 | 19 | 286 | 37 | 312 | 1 | 247 |
| 2 | 113 | 20 | 236 | 38 | 251 | 4 | 252 |
| 3 | 27 | 21 | 176 | 39 | 87 | 8 | 283 |
| 4 | 252 | 22 | 88 | 40 | 73 | 9 | Unmatched |
| 5 | 231 | 23 | 130 | 41 | 204 | 10 | 133 |
| 6 | 33 | 24 | Unmatched | 42 | 250 | 13 | Unmatched |
| 7 | Unmatched | 25 | 153 | 43 | 140 | 22 | 88 |
| 8 | 283 | 26 | 258 | 44 | 22 | 26 | 258 |
| 9 | Unmatched | 27 | 14 | 45 | 276 | 27 | 14 |
| 10 | 133 | 28 | 114 | 46 | 47 | 28 | 114 |
| 11 | 128 | 29 | Unmatched | 47 | 43 | 31 | 304 |
| 12 | 89 | 30 | 208 | 48 | 61 | 32 | Unmatched |
| 13 | Unmatched | 31 | 304 | 49 | 108 | 43 | 140 |
| 14 | 172 | 32 | Unmatched | 50 | 313 | 44 | 22 |
| 15 | 188 | 33 | 319 | 51 | 214 | 49 | 108 |
| 16 | Unmatched | 34 | 225 | 52 | 191 | 50 | 313 |
| 17 | 285 | 35 | 316 | 53 | 230 | 52 | 191 |
| 18 | 36 | 36 | 154 | 54 | 106 | 53 | 230 |

Table A5. Results provided by the experts vs GA. Matching experiment (Tests 1, 2 and 3).

| Test 1 | | | | | | | | Test 2 | | | | | | | | Test 3 | | | | | | | |
|---------|-----|------------|-----|-----|-----|-----|---------------|---------|-----|------------|-----|-----|-----|-----|---------------|---------|-----|------------|-----|-----|-----|-----|---------------|
| Pol ID. | GA | Expert ID. | | | | | Agreement [%] | Pol ID. | GA | Expert ID. | | | | | Agreement [%] | Pol ID. | GA | Expert ID. | | | | | Agreement [%] |
| | | 16 | 5 | 2 | 21 | 8 | | | | 12 | 9 | 1 | 14 | 15 | | | | 3 | 4 | 7 | 6 | 24 | |
| 1 | 247 | 247 | 247 | 247 | 247 | 247 | 100 | 19 | 286 | 286 | 286 | 286 | 286 | 286 | 100 | 37 | 312 | 312 | 312 | 312 | 312 | 312 | 100 |
| 2 | 113 | 113 | 113 | 113 | 113 | 113 | 100 | 20 | 236 | 236 | 236 | 236 | 236 | 236 | 100 | 38 | 251 | 251 | 251 | 251 | 251 | 251 | 100 |
| 3 | 27 | 27 | 27 | 27 | 27 | 27 | 100 | 21 | 176 | 176 | 176 | 176 | 176 | 176 | 100 | 39 | 87 | 87 | 87 | 87 | 87 | 87 | 100 |
| 4 | 252 | 252 | 252 | 252 | 252 | 252 | 100 | 22 | 88 | 88 | 88 | 88 | 88 | 88 | 100 | 40 | 73 | 73 | 73 | 73 | 73 | 73 | 100 |
| 5 | 231 | 231 | 231 | 231 | 231 | 231 | 100 | 23 | 130 | 130 | 130 | 130 | 130 | 130 | 100 | 41 | 204 | 204 | 204 | 204 | 204 | 204 | 100 |
| 6 | 133 | 133 | 133 | 133 | 133 | 133 | 100 | 24 | - | - | - | - | - | - | 100 | 42 | 250 | 250 | 250 | 250 | 250 | 250 | 100 |
| 7 | - | - | - | - | - | - | 100 | 25 | 153 | 153 | 153 | 153 | 153 | 153 | 100 | 43 | 140 | 140 | 140 | 140 | 140 | 140 | 100 |
| 8 | 283 | 283 | 283 | 283 | 283 | 283 | 100 | 26 | 258 | 258 | 258 | 258 | 258 | 258 | 100 | 44 | 22 | 22 | 22 | 22 | 22 | 22 | 100 |
| 9 | - | - | - | - | 74 | - | 80 | 27 | 14 | 14 | 14 | 14 | 14 | 14 | 100 | 45 | 276 | 276 | 276 | 276 | 276 | 276 | 100 |
| 10 | 133 | 133 | 133 | 133 | 101 | 133 | 80 | 28 | 114 | 114 | 114 | 114 | 114 | 114 | 100 | 46 | 47 | 47 | 47 | 47 | 47 | 47 | 100 |
| 11 | 128 | 128 | 128 | 128 | 128 | 128 | 100 | 29 | - | - | - | - | - | - | 100 | 47 | 43 | 43 | 43 | 43 | 43 | 43 | 100 |
| 12 | 89 | 89 | 89 | 89 | 89 | 89 | 100 | 30 | 208 | 208 | 208 | 208 | 208 | 208 | 100 | 48 | 61 | 61 | 61 | 61 | 61 | 61 | 100 |
| 13 | - | - | - | - | - | - | 100 | 31 | 304 | 304 | 304 | 304 | 304 | 304 | 100 | 49 | 108 | 108 | 108 | 108 | 108 | 108 | 100 |
| 14 | 172 | 172 | 172 | 172 | 172 | 172 | 100 | 32 | - | - | - | - | - | - | 100 | 50 | 313 | 313 | 313 | 313 | 313 | 313 | 100 |
| 15 | 188 | 188 | 188 | 188 | 188 | 188 | 100 | 33 | 319 | 319 | 319 | 319 | 319 | 319 | 100 | 51 | 214 | 214 | 214 | 214 | 214 | 214 | 100 |
| 16 | - | - | - | - | - | - | 100 | 34 | 225 | 225 | 225 | 225 | 225 | 225 | 100 | 52 | 191 | 191 | 191 | - | 191 | 191 | 80 |
| 17 | 285 | 285 | 285 | 285 | 285 | 285 | 100 | 35 | 316 | 316 | 316 | 316 | 316 | 316 | 100 | 53 | 230 | 230 | 230 | 230 | 230 | 230 | 100 |
| 18 | 36 | 36 | 36 | 36 | 36 | 36 | 100 | 36 | 154 | 154 | 154 | 154 | 154 | 154 | 100 | 54 | 106 | 106 | 106 | 106 | 106 | 106 | 100 |

Table A6. Results provided by the experts vs GA. Matching experiment (Test 4).[illegible]

References

1. Ruiz-Lendinez, J.J.; Ureña-Cámara, M.A.; Ariza-López, F.J. A Polygon and Point-Based Approach to Matching Geospatial Features. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 399. [\[CrossRef\]](#)
2. Fleuret, F.; Li, T.; Dubout, C.; Wampler, E.K.; Yantis, S.; Geman, D. Comparing machines and humans on a visual categorization test. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17621–17625. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Hassabis, D.; Kumaran, D.; Summerfield, C.; Botvinick, M. Neuroscience-inspired artificial intelligence. *Neuron* **2017**, *95*, 245–258. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Borowski, J.; Funke, C.; Stosio, K.; Brendel, W.; Wallis, T.; Bethge, M. The Notorious Difficulty of Comparing Human and Machine Perception. In Proceedings of the Conference on Cognitive Computational Neuroscience, Berlin, Germany, 13–16 September 2019.
5. Nyandwi, E.; Koeva, M.; Kohli, D.; Bennett, R. Comparing Human versus Machine-Driven Cadastral Boundary Feature Extraction. *Remote Sens.* **2019**, *11*, 1662. [\[CrossRef\]](#)
6. Quackenbush, L.J. A Review of Techniques for Extracting Linear Features from Imagery. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 1383–1392. [\[CrossRef\]](#)
7. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Proc. Syst.* **2012**, *25*, 1097–1105. [\[CrossRef\]](#)
8. Eigen, D.; Fergus, R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
9. Zhang, X.; Zhao, X.; Molenaar, M.; Stoter, J.; Kraak, M.; Tingshua, A. Pattern classification approaches to matching building polygons at multiple scales. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-2, XXII ISPRS Congress, Melbourne, Australia, 25 August–1 September 2012.
10. Fan, H.; Zipf, A.; Fu, Q.; Neis, P. Quality assessment for building footprints data on OpenStreetMap. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 700–719. [\[CrossRef\]](#)
11. Tang, J.; Jiang, B.; Zheng, A.; Luo, B. Graph matching based on spectral embedding with missing value. *Pattern Recognit.* **2012**, *45*, 3768–3779. [\[CrossRef\]](#)
12. Feng, W.; Liu, Z.; Wan, L.; Pun, C.; Jiang, J. A spectral-multiplicity-tolerant approach to robust graph matching. *Pattern Recognit.* **2013**, *46*, 2819–2829. [\[CrossRef\]](#)
13. Dold, J.; Groopman, J. Geo-spatial Information Science The future of geospatial intelligence. *Future Geospat. Intell.* **2017**, *20*, 5020–5151.
14. Devogele, T.; Trevisan, J.; Raynal, L. Building a multi-scale database with scale-transition relationships. In Proceedings of the 7th International Symposium on Spatial Data Handling, Delft, The Netherlands, 12–16 August 1996; Taylor & Francis: Oxford, UK, 1996; pp. 337–351.
15. Yuan, S.; Tao, C. Development of conflation components. In Proceedings of the Geoinformatics'99 Conference, Ann Arbor, MI, USA, 19–21 June 1999; pp. 1–13.
16. Song, W.; Keller, J.; Haithcoat, T.; Davis, C. Relaxation-Based Point Feature Matching for Vector Map Conflation. *Trans. GIS* **2011**, *15*, 43–60. [\[CrossRef\]](#)
17. Stoter, J.; Burghardt, D.; Duchêne, C.; Baella, B.; Bakker, N.; Blok, C.; Pla, M.; Regnauld, N.; Touya, G.; Schmid, S. Methodology for evaluating automated map generalization in commercial software. *Comput. Environ. Urban Syst.* **2009**, *33*, 311–324. [\[CrossRef\]](#)
18. Ruiz-Lendinez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A. Automatic positional accuracy assessment of geospatial databases using line-based methods. *Surv. Rev.* **2013**, *45*, 332–342. [\[CrossRef\]](#)
19. Ruiz-Lendinez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A. A point-based methodology for the automatic positional accuracy assessment of geospatial databases. *Surv. Rev.* **2016**, *48*, 269–277. [\[CrossRef\]](#)
20. Goodchild, M.; Hunter, G. A Simple Positional Accuracy Measure for Linear Features. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 299–306. [\[CrossRef\]](#)
21. Hastings, J.T. Automated conflation of digital gazetteer data. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 1109–1127. [\[CrossRef\]](#)
22. Huh, Y.; Yu, K.; Heo, J. Detecting conjugate-point pairs for map alignment between two polygon datasets. *Comput. Environ. Urban Syst.* **2011**, *35*, 250–262. [\[CrossRef\]](#)
23. Samal, A.; Seth, S.; Cueto, K. A feature-based approach to conflation of geospatial source. *Int. J. Geogr. Inf. Sci.* **2004**, *18*, 459–489. [\[CrossRef\]](#)
24. Kim, J.; Yu, K.; Heo, J.; Lee, W. A new method for matching objects in two different geospatial datasets based on the geographic context. *Comput. Geosci.* **2010**, *36*, 1115–1122. [\[CrossRef\]](#)
25. Arkin, E.; Chew, L.; Huttenlocher, D.; Kedem, K.; Mitchell, J. An efficiently computable metric for comparing polygonal shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 209–216. [\[CrossRef\]](#)
26. Olteanu-Raimond, A.; Mustière, S.; Ruas, A. Knowledge formalization for vector data matching using belief theory. *J. Spat. Inf. Sci.* **2015**, *10*, 21–46. [\[CrossRef\]](#)
27. Jones, C.B.; Ware, J.M.; Miller, D.R. A probabilistic approach to environmental change detection with area-class map data. In *Integrated Spatial Databases*; Agouris, P., Stefanidis, A., Eds.; Springer: Berlin/Heidelberg, Germany, 1999; pp. 122–136.
28. Herrera, F.; Lozano, M.; Verdegay, J. Tackling Real-Coded Genetic Algorithms: Operators and Tools for Behavioural Analysis. *Artif. Intell. Rev.* **1998**, *12*, 265–319. [\[CrossRef\]](#)

29. Herrera, F.; Lozano, M.; Sánchez, A. A Taxonomy for the Crossover Operator for Real-coded Genetic Algorithms: An Experimental Study. *Int. J. Intell. Syst.* **2003**, *18*, 309–338. [\[CrossRef\]](#)
30. Elder, J.; Zucker, S. The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vis. Res.* **1993**, *33*, 981–991. [\[CrossRef\]](#)
31. Kovacs, I.; Julesz, B. A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7495–7497. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Ringach, D.L.; Shapley, R. Spatial and temporal properties of illusory contours and amodal boundary completion. *Vis. Res.* **1996**, *36*, 3037–3050. [\[CrossRef\]](#)
33. Tversky, T.; Geisler, W.S.; Perry, J.S. Contour grouping: Closure effects are explained by good continuation and proximity. *Vis. Res.* **2004**, *44*, 2769–2777. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Ullman, S.; Assif, L.; Fetaya, E.; Harari, D. Atoms of recognition in human and computer vision. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2744–2749. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Majaj, N.J.; Pelli, D.G. Deep learning-using machine learning to study biological vision. *J. Vis.* **2018**, *18*, 22. [\[CrossRef\]](#)
36. Kar, K.; Kubilius, J.; Schmidt, K.; Issa, E.B.; DiCarlo, J.J. Evidence that recurrent circuits are critical to the ventral stream’s execution of core object recognition behaviour. *Nat. Neurosc.* **2019**, *22*, 974. [\[CrossRef\]](#)
37. Ruiz-Lendinez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A. Study of NSSDA Variability by Means of Automatic Positional Accuracy Assessment Methods. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 552. [\[CrossRef\]](#)
38. Kuhnert, P.M.; Martin, T.G.; Griffiths, S.P. A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecol. Lett.* **2010**, *7*, 900–914. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Ayyub, B.M. *Elicitation of Expert Opinions for Uncertainty and Risks*; CRC Press: Boca Ratón, FL, USA, 2001.
40. Roberson, C.; Feick, R. Defining Local Experts: Geographical Expertise as a Basis for Geographic Information Quality. In *Proceedings of Workshops and Posters at the 13th International Conference on Spatial Information Theory (COSIT 2017)*; Clementini, E., Fogliaroni, P., Ballatore, A., Eds.; Springer: L’Aquila, Italy, 2017; Article No. 22; pp. 22:1–22:14.
41. Brodaric, B.; Fox, P.; McGuinness, D.L. Geoscience Knowledge representation in cyber infrastructure. *Comput. Geosci.* **2009**, *35*, 697–699. [\[CrossRef\]](#)
42. Casas, J.; Repullo, J.; Donado, J. La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I). *Aten. Primaria* **2003**, *31*, 527–538. [\[CrossRef\]](#)
43. Fischler, M.; Elschlager, R. The representation and matching of pictorial structures. *IEEE Trans. Comput.* **1973**, *22*, 67–92. [\[CrossRef\]](#)
44. Radoux, J.; Waldner, F.; Bogaert, P. How Response Designs and Class Proportions Affect the Accuracy of Validation Data. *Remote Sens.* **2020**, *12*, 257. [\[CrossRef\]](#)