Article

A Polygon and Point-Based Approach to Matching Geospatial Features

Juan J. Ruiz-Lendínez *, Manuel A. Ureña-Cámara [iD] **and Francisco J. Ariza-López**

Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría, Escuela Politécnica Superior de Jaén, Universidad de Jaén, 23071 Jaén, Spain; maurena@ujaen.es (M.A.U.-C.); fjariza@ujaen.es (F.J.A.-L.)
* Correspondence: lendinez@ujaen.es; Tel.: +34-953-21-24-70; Fax: +34-953-21-28-54

Abstract: A methodology for matching bidimensional entities is presented in this paper. The matching is proposed for both area and point features extracted from geographical databases. The procedure used to obtain homologous entities is achieved in a two-step process: The first matching, polygon to polygon matching (inter-element matching), is obtained by means of a genetic algorithm that allows the classifying of area features from two geographical databases. After this, we apply a point to point matching (intra-element matching) based on the comparison of changes in their turning functions. This study shows that genetic algorithms are suitable for matching polygon features even if these features are quite different. Our results show up to 40% of matched polygons with differences in geometrical attributes. With regards to point matching, the vertex from homologous polygons, the function and threshold values proposed in this paper show a useful method for obtaining precise vertex matching.

Keywords: area entities matching; point entities matching; genetic algorithm; turning function descriptor; geographical databases

1. Introduction

Pattern recognition and matching of features are essential processes in many problems when dealing with graphic information [1,2]. Whether in vector format (object-oriented) or in raster format, there are many studies related to graphic information. One specific type of graphic information is geographic information or spatial data, which is very much in demand nowadays because of the use of location in many areas (business, military, international aid, tourism, fleet management, smart territories, etc.) and has great economic importance [3]. The main feature of spatial data is the geographical position, which means that spatial data refer to positions in the real world. Position is the basis of mapping and navigation, which are essential for engineering, natural sciences, land management, etc. Different matching techniques based on position have been developed in response to the need to create new cartographic products, which are the result of integrating Geospatial Databases (GDB) from heterogeneous sources with levels of quality equally heterogeneous [4–6]. This integration process has not only allowed the development of new ways of mapping based on the Volunteered Geographic Information (VGI) and the Spatial Data Infrastructures (SDI), but has also been the cause of an ever-increasing necessity to implement more efficient accuracy assessment procedures for evaluating the quality of these products.

Under this new geospatial paradigm, the knowledge of the positional accuracy of a GDB is a critical aspect. Thus, a GDB with an insufficient level of positional accuracy for a particular purpose can be useful by integrating with other GDB with higher accuracy. Therefore, and regardless of this level of accuracy, it is necessary to know it. Positional accuracy has traditionally been evaluated by means of statistical methods based on measuring errors in points between the real world and the

dataset (Figure 1a). From a statistical point of view, one of the most controversial aspects of these methods is the number of points that must be used to ensure, with a given level of confidence, that a GDB with an unacceptable quality level will not be acquired [7]. Thus, although recommendations always suggest at least 20 points [8,9], other studies [10] and institutions [11,12] suggest that this size (n = 20) seems to be very small, and larger sizes are proposed [13]. In this sense, the new capabilities offered by Global Navigation Satellite Systems (GNSS) in the last few decades have allowed a great improvement in the process of field data acquisition; however, their use still presents the inconvenience of having to use a limited number of points. This factor, combined with the very high cost of field surveying, represents the main weakness of the traditional points-based methods (Figure 1a).
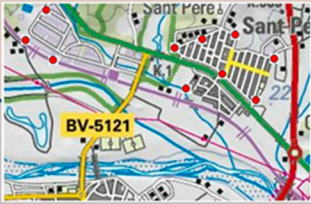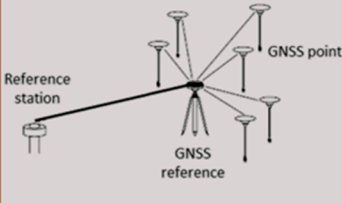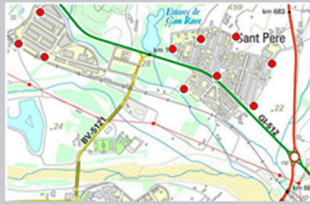

Figure 1. (**a**) Traditional points-based methods of positional quality assessment and (**b**) the new approach based on automated positional assessment. GBD, Geospatial Databases.

Having taken into account these considerations, and in order to overcome the difficulties mentioned previously, a new assumption was developed that positional accuracy could also be defined by measuring the differences between the location stored in a GDB (tested source) and a location determined by another GDB (reference source) of higher accuracy (Figure 1b). Thus, if the accuracy of the second source is high enough, the unmeasured difference between it and the real one can be ignored [14]. Under this new approach to the overall process of positional assessment, the first and most important stage is to identify the homologous elements between both GDB, quickly and reliably. Thus, we consider that now is the right time for developing new methodologies based on the automated assessment of the positional and geometric components of geographic information. In this sense, the authors developed a methodology to analyze the assessment of the positional components of lines based on a previous matching and a set of descriptors [15]. Latterly, they continued their study on positional accuracy by means of point-based methodologies [16]. Both studies are based on a two-step matching process that is extensively described in this paper. Thus, we can consider that the assessment process is essentially dealing with a matching problem where there are two key issues [2]: (1) how to measure the similarity between elements from both datasets (in our case, GDB) in order to match them (Figure 2a) and (2) how to achieve the optimal vertex-to-vertex correspondence between the previously-matched elements (Figure 2b). Following these ideas, in this paper, we show the matching process in a two-step process (inter-elements and intra-elements) that we have been developing in order to allow quality control assessment of two GDBs.
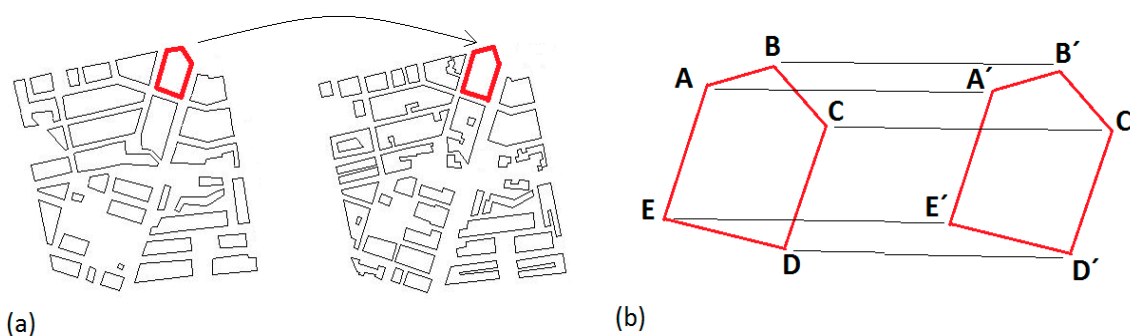
Figure 2. (**a**) Similarity between elements from both GDB and (**b**) vertex-to-vertex correspondence between the elements matched previously.

1.1. Inter-Elements Matching (Inter-Polygons)

With regard to the measurement of similarity in order to match elements, this is generally evaluated by means of a quantitative variable, which measures the distance between the elements compared and analyzes them by means of a set of characteristics. In the field of geographic information, there are several studies regarding the classification of these measures, among them Tong et al. [17], who defined three different categories:

- Geometric measures: considers all those aspects arising from the coordinates that define a spatial object: (i) positional aspect and (ii) shape, which is represented by means of the geometry.
- Spatial relationship measures: considers geometric aspects related to the relation between two features: (i) distance between features, (ii) direction and (iii) topological.
- Attribute measures: considers thematic aspects of the geographic information.

The issue examined in this paper is related both to the geometric measures used to determine a geometric similarity and the spatial relationship measures employed in order to determine distances from matching features. In our case, the position and the shape adopted by cartographic elements that represent real-world phenomena are the most significant aspects taken into consideration in measuring the similarity between them and their subsequent matching.

Over the years, some matching techniques have been developed. However, the majority of the current approaches concentrate on network matching and mainly on road networks [18–27]. In contrast to network matching, there are few studies concerning matching area; nevertheless, despite this shortage of research, there are relevant area-based matching algorithms that must be analyzed.

Overall, it can be said that the techniques employed to match objects using this type of algorithm are based on: the percentage of overlapped area [28–30], the context by means of a Delaunay triangulation [31,32], the belief theory on position and orientation [33] and probability classifiers over a set of "evidence", both geometric and attribute-based [34]. A review of these studies shows that even though the overall performance in matching between features is satisfactory, there are some aspects that need to be improved. In this sense, the main problems are usually related to: the generation of false matching pairs in the cases of *1:n* or *n:m* correspondences with polygons with relatively complex contours [29,30], the matching between networks with different levels of detail [33,34] and the process´s low level of automatization [32]. Especially interesting are the ideas presented by Mortara and Spagnolo [31]. In their study, the correspondences between polygons are found by taking into account the morphological characterization of the shapes by means of their approximate skeleton. Thus, the correspondence problem is essentially reduced to a question of a skeleton matching process. It could be said, therefore, that this approach represents a combined solution (area-based and network-based) to the matching problem. On the other hand, Zhang et al. [35] compare different matching algorithms like the Classification And Regression Tree (CART) [36], C4.5 algorithms [37], the naive Bayes classifier (a probabilistic model) and Support Vector Machines (SVM) over the same set of measures, area

relationship, shape similarity and an adapted wall statistical weighting. The conclusions of these authors do not mark a difference from the results of all these classifiers.

Another type of matching algorithm is inspired by the concept of shape blending [38,39]. Based on the morphological characterization of objects, this process is related to the gradual transformation of one shape into another and requires, as a first step, finding correspondences between features from the two shapes. Although the main and the most numerous practical applications of shape blending are related to computer animation, scientific visualization and solid modeling, some authors have developed methods applied to spatial data. Among these methods, we must highlight the approach developed by Prakash and Robles-Kelly [40]. These last authors measure the similarity (or dissimilarity) between objects by counting the number of edit operations required to transform one object into another, and although this approach is closely related to the problem of measuring geometrical similarities, it does not seem to provide an optimal solution when the number of correspondences is too high. For this reason, recent efforts have been focused on comparing an input shape with various template shapes (patterns), which represent potential solutions to the matching problem, and reporting the template with the highest similarity (or lowest dissimilarity) to the input shape. In our case, and because of the need to match a large amount of shapes, it makes sense to use this last mechanism based on pattern comparison. In addition, there are various similarity/dissimilarity measures to be considered for this task [41].

1.2. Intra-Elements Matching (Vertex-to-Vertex)

With regard to achieving an optimal vertex-to-vertex correspondence between previously matched elements, the background information available in the field of spatial data is once again very sparse. Thus, the main contributions center on applications related to graph matching, such as shape analysis, face recognition and structural pattern recognition. Among the proposed methods. we must highlight the studies of Gösseln and Sester [42] and Butenuth et al. [43]. These authors match polygon datasets by using an iterative closest point algorithm that detects corresponding point pairs for two point sets derived from each contour of corresponding objects. Zhang and Meng [44] integrated two spatial datasets by means of an iterative matching approach, and Safra et al. [45] developed a location-based join algorithm for searching for corresponding objects based on the distance proximity between their representative points. On the other hand, Huh et al. [46] develop a method for detecting a corresponding point pair between a polygonal object pair with a string matching method based on a confidence region model of a line segment. Another common approach is that of isometric point pattern matching [47]. According to these authors, in such a case, we assume that a "copy" of an object G (the "template") appears isometrically transformed within G´ (the "target"). Song et al. (2011) [6] develop a procedure based on a relaxation of the probability of matching between different junctures inside a traffic network.

However, these methods have some limitations in detecting vertex-to-vertex correspondences between previously matched elements. Some of them require topological properties such as a valence of each node or the directions of edges connected at a node [6,44], while others consider the vertices or locations of objects as isolated points, not as geometric primitives that form the shape of an object [42,43,45], and finally, others are restricted to a low density of objects to be matched [42,46]. Following Fan et al. [30], in this last case when neighboring polygons are located immediately next to each other and are similar in shape and size, there will be error matching.

In order to overcome the aforementioned limitations, the contour-matching algorithms have been developed. These algorithms take into account both the order and location of an object's vertices and match the contours of an object pair by comparing their vertex strings [48]. Nevertheless, these methods, as in those described above, present some disadvantages, which in this case, are related mainly to the comparison between vertex strings, which belong to objects with relatively different contours. In addition, they cannot be applied to a *1:n* or an *m:n* corresponding object pair [29].

A different interpretation of the contour-matching algorithms is proposed by Arkin et al. [49], who compare two polygonal curves by means of their turn angle representations. Specifically, this last approach allows us to solve the point matching problem more efficiently than in the previous cases due to the characteristics of the geographic information with which we work.

1.3. Paper Outline

Taking into account all of the above, we propose an approach for the matching of objects from GDBs, which tries to overcome as far as possible the limitations of the methods discussed above. Constrained to a unique spatial phenomenon like urban zones and a specific type of element like polygonal shapes, our method considers the first two measures mentioned by Veltkamp and Hagedoorn (2000) [41] applied in two major steps:

- Quantification of similarity (inter-elements): The matching of polygonal shapes is effected by means of a weight-based classification methodology using the polygon's low-level feature descriptors (edge-based features such as area, perimeter, number of angles, etc.). This classification allows for the categorization of the matching quality from a quantification of the similarity. Among all the existent classifiers, we have chosen a Genetic Algorithm (GA). The GA belongs to a family of stochastic global optimization methods based on the concept of Darwinian evolution in populations. Their performance has been sufficiently studied and developed by several authors [50–54], and all their possible variants share the main mechanism that operates on an initial population of randomly-generated chromosomes or individuals, who represent possible solutions to the problem, and consists of three operations: evaluation of individual fitness, development of an intermediate population through selection mechanisms and recombination through crossover and mutation operators. The main reason for employing a GA is that, compared with other matching methods based on descriptors such as Fourier descriptors [55], these classifiers have proven their applicability not only to solve matching problems for all those approaches that need a power search procedure for problems with large, complex and poorly understood search spaces [56], but also to produce quality values from this matching [57,58]. These quality values in turn are closely related to the geometric similarity measures and are very relevant for our work since they allow us to select *1:1* corresponding objects pairs among all the possible correspondences. This is the simplest and most efficient approach to matching geospatial data [59] and avoids the disadvantages described in Section 1.1 (false matching pairs in the cases of *1:n* or *n:m* correspondences).

- Threshold of dissimilarity (intra-element): After having matched the polygonal shapes from the two GDBs, the second and last step involves using a metric to compute homologous points, which represent a vertex-to-vertex matching. Specifically, we used a contour-matching metric based on the method defined by Arkin et al. [49] for comparing two polygonal shapes, which essentially consists of comparing their turning functions and checking whether the distance between these two functions in certain areas of their graphic representations is smaller than a previously fixed threshold. In this case, the use of this metric applied to *1:1* correspondences allowed us to overcome the disadvantages described in Section 1.2. In addition, this metric has also allowed us to define a powerful new edge-based shape descriptor, which takes into account the relationships between adjacent boundaries of polygons.

The rest of this paper is organized as follows: Section 2 defines the two geospatial datasets and the polygonal features used. Sections 3 and 4 focus on the matching methodology. Specifically in Section 3, the matching of polygonal shapes is addressed, and in Section 4, we examine the issue of vertex-to-vertex matching. Finally, experimental results and discussions are presented in Section 5, followed by conclusions in Section 6.

2. The Two Geospatial Databases and Polygonal Features Used

The two GDBs selected for our study needed to meet certain requirements in order to assert that both GDBs must be independently produced and that neither of these two GDBs, in turn, can be derived from another cartographic product of a larger scale through any process. All this implies that none of them have undergone any action involving a degradation of their quality. In addition, both datasets should be interoperable, which means that we must ensure that they will be comparable both in terms of reference system and cartographic projection [60]. Therefore, we used two official cartographic databases in Andalusia (southern Spain) that meet the aforementioned requirements: the BCN25 (from "*Base Cartográfica Numérica E25k*") and the MTA10 (from "*Mapa Topográfico de Andaluc selectfontía E10k*").

The BCN25 dataset is the digital version of the "Mapa Topográfico Nacional" at scale 1:25,000, (MTN25K map series), produced by the "Instituto Geográfico Nacional" (National Geographic Institute of Spain) in 2010 [61]. Its planimetric accuracy has been estimated at between 2.5 and 7.5 m, depending on the type of entity considered. However, several authors estimate its planimetric accuracy at 4.0 m [62]. The MTA10 dataset, known as the "Mapa Topográfico de Andalucía E10K" (Andalusian Topographic Map at scale 1:10,000), is produced by the "Instituto de Estadística y Cartografía de Andalucía" (IECA), and more precisely, the GDB used was produced in 2009. The declared positional accuracy of the MTA10 is RMSE = 3 m [63]. The MTA10 would play the role of "reference source" in the subsequent process of positional accuracy assessment of the BCN25 database.

After describing the GDB, we must define the polygonal features used in our matching methodology. In this case, we selected buildings from urban zones. The use of this type of object is based on the following: (1) buildings are a huge set of polygonal features in any GDB, at least in the number of elements, although perhaps not in area, and (2) the buildings included in any GDB contain a sufficient quantity of geometrical information (vertexes and the edges that join them) for our purpose.

Owing to the lack of relevant previous studies, in the initial stages of this study, we realized the need to apply the proposed methodology to a wider population sample. For this reason, the sample comprises nine urban areas selected to meet as far as possible all the variability requirements with regard to the typologies of polygons. These requirements include a wide geographical distribution and an ample variety in the type of buildings and topographical terrain. The urban areas selected are included in three sheets of the MTN50k (National Topographic Map of Spain at scale 1:50,000). More specifically, we used the following urban areas: Mairena de Alcor, Alcalá de Guadaira and Carmona (0985 Sheet); Fuente Vaqueros, Santa Fe and Granada (1009 Sheet); and Coín, Alhaurín de la Torre and Fuengirola (1066 Sheet) (Figure 3a). Figure 3b shows two examples of polygonal features corresponding to buildings belonging to MTA10 and BCN25 (Carmona, 0985 sheet).
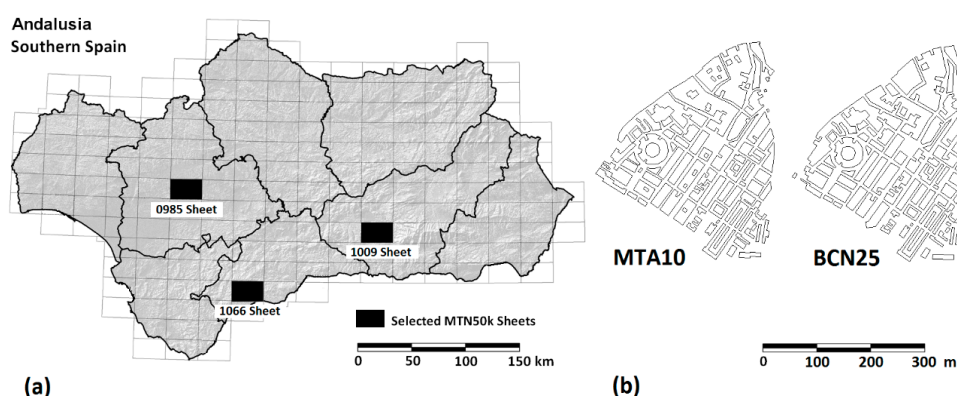

Figure 3. (**a**) Selected sheets of the Mapa Topográfico Nacional (MTN) at the scale of 1:50,000 of the region of Andalusia (southern Spain) and its administrative boundaries and (**b**) examples of polygonal features belonging to Mapa Topográfico de Andalucía E10K (MTA10) and Base Cartográfica Numérica E25k (BCN25) (Carmona, 0985 sheet).

3. Inter-Elements Matching

3.1. Polygon Low-Level Feature Descriptors

The main purpose of object descriptors involves finding discriminative features that describe the object of interest [64]. These features can represent different information levels. In our case, due to the characteristics of the geographic information and to the specific type of objects (two-dimensional polygonal shapes) we worked with, the features used as descriptors to encode the shape of an object belong to low-level information. Following Shih et al. [65], low-level feature descriptors are usually extracted to describe the geometric properties [66], spatial properties [67] and shape distributions [68] of two- and three-dimensional objects. In addition, the similarity between two 2D objects can be measured by comparing their features. In our approach, the GDB used must be able to compute these features efficiently and effectively. For this reason, the descriptors selected were edge-based features and included the following:

- Number of convex and concave angles: Both features are conditioned by edge orientation. This descriptor informs us about the complexity of the polygon's perimeter.
- Perimeter: This is a feature that is not conditioned by edge orientation, but depends on their lengths. This characteristic should indicate both the complexity of the polygon and some basic shape description in relation to the area.
- Area: This depends both on the orientation of the edges and on their lengths. Despite not being a particularly interesting descriptor with respect to the shape of the element, it has a very important visual weight in the GDB to be considered in the matching process.
- Minimum moment of inertia of second order: This can be computed easily by means of the polygon's total area, centroid and vertices, which are connected by straight lines in a 2D system. It depends both on the orientation and the lengths of the edges.
- Arkin Graph Area (AGA): For a polygon A, we define the AGA descriptor as the area of the region below the turning function θ_A. In order to avoid differences of AGA based on the starting points, all AGA are computed starting from the northwestern point in a-clockwise direction, and the perimeter distance is normalized to one. In addition, as mentioned in Section 1, this descriptor takes into account the relationships between adjacent boundaries of polygons and depends both on the orientation and the lengths of the edges. Its computation is addressed in Section 4.
- Minimum Bounding Rectangle (MBR): This is specified by coordinate pairs (X_{min}, X_{max}, Y_{min}, Y_{max}). This feature represents the absolute spatial position of the polygon and in the case of geographic information is a key aspect in object characterization.

These descriptors are used as the attributes of each polygon from the first GDB that tries to match another polygon of the second GDB. All of them are used in a normalized form from their differences values except the minimum bounding rectangle, which is determined as the total area of normalized overlapping rectangles.

Finally, we must note that all of these features aim to define, as much as possible, the parameters that humans tend to apply in a manual match of the objects. For example, the minimum bounding rectangle represents the first attempt of a cartographer at identifying homologous objects, that is objects that share a similar position in 2D space.

3.2. GA Classification of Polygonal Shapes

After characterizing the polygonal shapes, a set of homologous polygons from both GDBs was determined in order to match them. For this purpose, a GA based on real number representation (called Real Coded GA (RCGA)) was used. In an RCGA, a chromosome is a vector of floating point numbers whose size is the same as the proposed descriptors, which is the solution to the problem. This type of codification has been sufficiently developed by several authors [53,69–72], and for this reason, this issue is not described here.

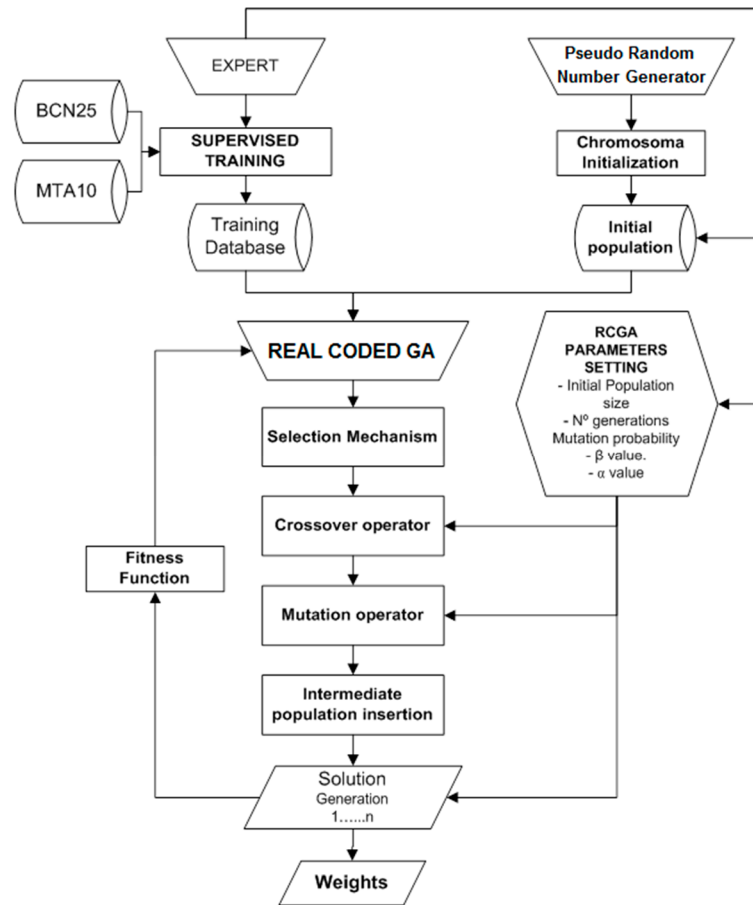The conceptual framework for the proposed methodology is shown in Figure 4.

Figure 4. Conceptual framework for the classification process. RCGA, Real Coded GA.

The classification methodology is based on computing a set of weights that must be assigned to the features sets and then using these weights to confront each object from both GDBs. To achieve this goal, it is necessary not only to establish the capability of the RCGA as a classification tool, but also to demonstrate that it can produce sufficiently good quality matches by means of a fitness function. The fitness function evaluates the quality of each solution (chromosome) [54]. In our approach, this function evaluates each of the weights assigned to each of the descriptors used to identify pairs of objects, the genes included in the chromosome. The weights evaluate the similarity of the polygon matching over a supervised training database, and the value obtained represents the quality of this solution. Obtaining this database using a supervised training procedure was the first stage of the process. The training database was composed of 805 polygons, with their respective homologous polygons, that were manually obtained by three different cartographers. The set of polygons obtained by these three cartographers was 265, 265 and 275 homologous polygons. These three sets were independent of each other and different from the final GDB used for the matching. However, the polygons were extracted from GDBs similar to those used in this paper. These two hypotheses, using different cartographers and a different set of GDB polygons, were used to avoid the influence of the person selecting homologous polygons and the polygons to be used. In addition, the supervised training process was externally evaluated by fifteen reviewers selected from a pool of internationally recognized experts. Their participation in the procedure reveals the robustness of the training process and the quality of the training dataset. Both the details relating to this procedure and the software implemented specifically for carrying it out can be found in [73,74].

The initial population, the set of weights for each descriptor, represents possible solutions to the matching problem and is generated through a process called chromosome initialization [75]. Paying special attention to this process may have a positive effect on the success of the GA. Although there are several alternatives such as quasi random sequences and the simple sequential inhibition process, in our case, the chromosome initialization was carried out by means of a Pseudo-random Number Generator Algorithm (PNGA). This is because it is particularly well suited to cases where the speed of convergence is high [76], which may happen with the specific type of spatial objects with which we work. In addition, it is fast and easy to use (a detailed description of this process is addressed in Section 5.1). The initial population evolves toward the final solution (iterative process) by developing an intermediate population through mechanisms of selection, crossover and mutation. These mechanisms or requirements also change during the execution of the algorithm. Thus, we prevent the premature convergence of RCGA to suboptimal solutions. In addition, multiple runs of the RCGA using different initial populations were carried out in order to assure that the initial selection did not affect the proposed solution. With regard to the size of the initial population, this is related to: (i) crossover and mutation probabilities (inversely related) and (ii) the number of iterations (called generations). In this last aspect, some authors suggest using a ratio value (size of population/number of generations) close to 1/5 [77]. In our approach, the initial population was integrated by *m* chromosomes, which included seven genes, each treated as the weight assigned to each of the descriptors.

Defining the main mechanism, which operates on an initial population, was the next stage in our approach. In order to develop an intermediate population we employed the proportional selection calculation [78], the Blend Alpha (BLX-α) crossover operator [69] and the non-uniform mutation [79]. This choice is justified because the experiments carried out highlight these genetic operators as the most suitable ones for building RCGAs [53]. On the other hand, both the *BLX-α* crossover operator and non-uniform mutation are controlled by two parameters (*α* and *β*), which determine the exploitation interval in the search space. The values assigned to these parameters depend in large part on the population size.

The weight computation was the last stage of our process. This is the final solution to the problem and represents the chromosome that has achieved the best results regarding the evaluation. Having computed the weights (after finalizing the training phase of the RCGA), a classification of the remaining features can be achieved. As mentioned in Section 1.3, this classification allows for the categorization of the matching quality from a quantification of the similarity and consists of weights assigned to each of the polygon's descriptors described above. Keeping this in mind, we consider two matched polygons (P_A, P_B) exactly equal if the Match Accuracy Value (MAV) achieved is equal to one (100%). This value is obtained as a linear combination of the *n* polygons descriptors ($At(P_A)$, $At(P_B)$) and the *n* weights ($W(At)$) computed in the training phase (Equation (1)).

$$MAV(P_A, P_B) = MAX \sum_{i=1}^{n} \left(At_i(P_A) / At_i(P_B) \cdot W_{At_i} \right), P_A \in GDB_{BCN25}, \forall P_B \in GDB_{MTA10} \qquad (1)$$

Equation (1) defines two characteristics: the first allows for determining the matching between polygons (P_A is selected, then we can choose P_B from the second GDB based on the maximum value of the linear combination); the second is the value of the linear combination that represents the accuracy of the matching, or in other words the shape similarity measure between two polygons. Thus, polygons having different degrees of similarity will provide non-equal values of MAV ranging from zero to one. Another important consideration in order to be able to achieve good results is that of avoiding the acceptance of erroneously-matched polygons (false positive), because this acceptance would also imply the subsequent introduction of mismatched points. For this reason, two matched polygons will be used to compute the matching between points depending on the MAV value achieved and only if this value exceeds a previously fixed threshold.

Finally, we must note that an extensive description of the codification of the RCGA employed can be found in [80]. The most powerful feature of this software is that control parameters of selection, crossover and mutation operators can be defined in order to find the most adequate solution for our approach.

4. Intra-Elements Matching (Vertex-to-Vertex Matching)

After having matched the polygonal shapes (buildings) from both GDBs, and in order to achieve a vertex-to-vertex matching, a metric derived from the method defined by Arkin et al. [49] was used. Following these authors, a common representation of the boundary of a simple polygon P_A can be obtained by the turning function $\theta_{P_A}(s)$. However, we used this function, which measures the angle of the counterclockwise tangent, as a function of the arc-length s in the clockwise direction, measured from some reference point *0* on *A*'s boundary. Thus, $\theta_{P_A}(0)$ is the angle *v* that the tangent at the reference point *0* makes with some reference orientation associated with the polygon (such as the x-axis). $\theta_{P_A}(s)$ keeps track of the turning that takes place, increasing with left-hand turns and decreasing with right-hand turns. If we assume that each polygon is rescaled so that the total perimeter length is one, then $\theta_{P_A}(s)$ is a function from (0, 1) to $(-4\pi, 4\pi)$ (Figure 5). The function $\theta_{P_A}(s)$ has several properties, which make it especially suitable for identifying and extracting homologous points between two polygons. As mentioned in [49], it is piecewise-constant for polygons, making computations particularly easy and fast. By definition, the function $\theta_{P_A}(s)$ is invariant under translation and scaling of the polygon *A*. Rotation of *A* corresponds to a simple shift of $\theta_{P_A}(s)$ in the θ direction. In addition, we used the area of the region below $\theta_{P_A}(s)$ in order to define what we denominated as an Arkin Graph Area (AGA). This area can be easily computed by the value of the integral (Equation (2)) and was used as another polygon's shape descriptor in the classification stage. This integral can be obtained numerically by a sum of prisms because the changes in height only take place in each vertex.
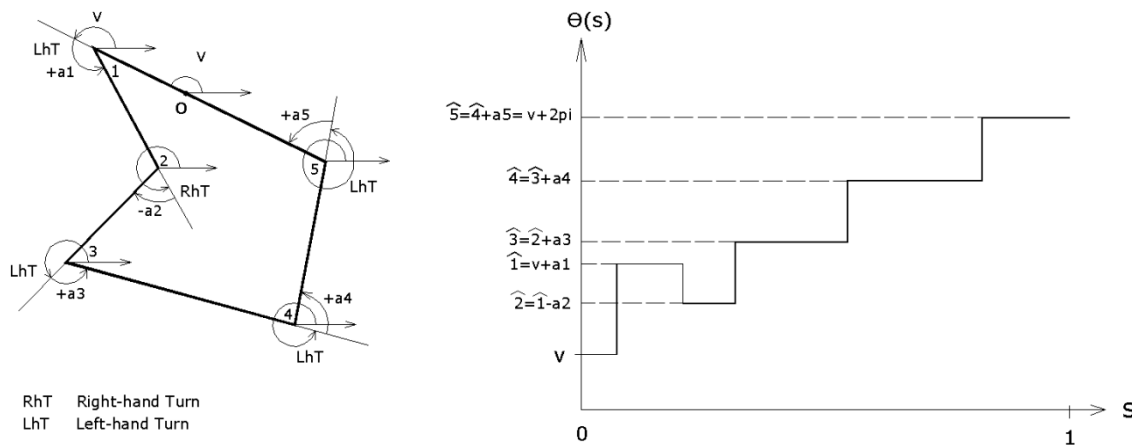
$$AGA(P_A) = \int_0^1 \theta_{P_A}(s)ds \tag{2}$$

Figure 5. Defining the turning function $\theta_{P_A}(s)$.

In our approach, we used the angle formed by any side of a polygon and the extension of its adjacent side (exterior angles of a polygon) to compute the turning functions. We must note that the exterior angles of a simple polygon always add up to 4π. As mentioned above, if each polygon is rescaled so that the total perimeter length is one and the exterior angles always add up to 4π, then the turning functions of two matched polygons can be overlapped and compared (Figure 6, Column b). Taking this idea in mind, we used this formulation to compute a distance function for comparing two simple polygons (P_A and P_B) by looking at natural notions of distances between their turning

functions $\theta_{P_A}(s)$ and $\theta_{P_B}(s)$. Thus, each discontinuity is defined by a pair of points, which represents a vertex of the first polygon that can be matched with another vertex of the second one (homologous vertex) (Figure 6, Column c).
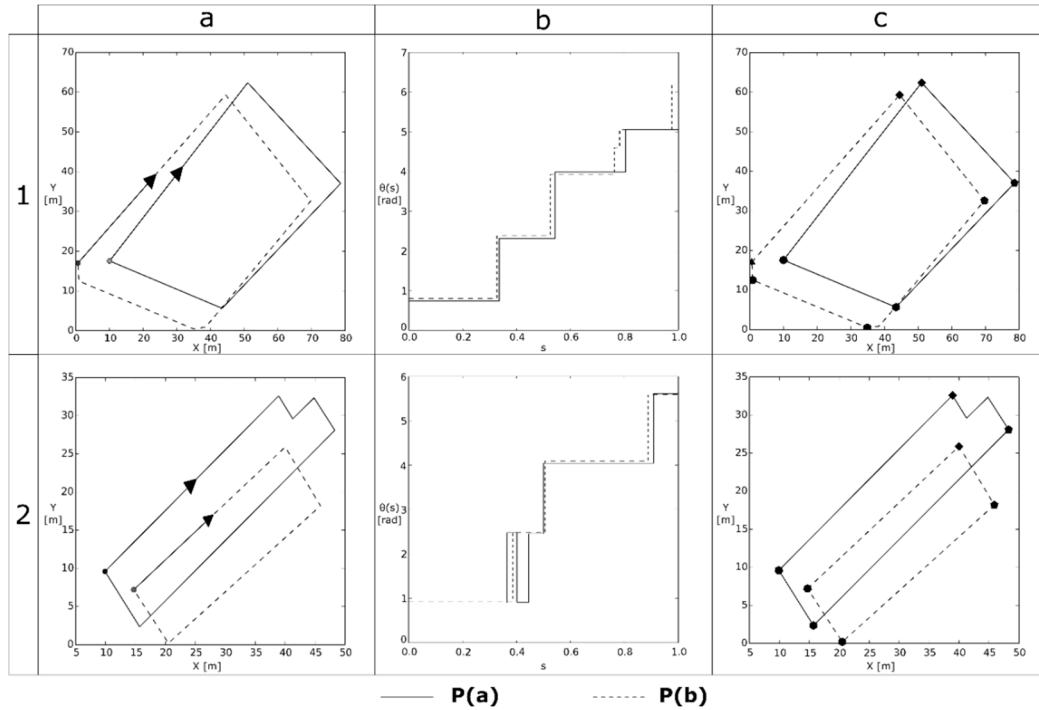

Figure 6. θ(s). Matching of homologous points. (**a**) Matched polygons, (**b**) turning functions overlapping and (**c**) matching of homologous points; homologous points are represented by a different number of polygon sides of the marker.

The distance between two pairs of vertices, denoted by $L_V(A, B)$, depends on (Equation (3)): length shifts (variations on the s axis due to length changes of the polygon sides) and angular shifts (variations on the θ axis due to direction changes of the polygon sides) of the turning functions. In order to avoid matching non-homologous points, two Threshold values were fixed for these Length and Angular Shifts (denoted by TLS and TAS respectively). Thus, only when these shifts are lower than the previously-defined threshold values will the matching between these points be achieved.

$$
\begin{aligned}
&L_V(A, B) = \sqrt{\left(\theta_{P_A}(s_A) - \theta_{P_B}(s_B)\right)^2 + \left((s_A/s_{P_A} - s_B/s_{P_B})\max(s_{P_A}, s_{P_B})\right)^2}, \ where: \\
&\left|\theta_{P_A}(s_A) - \theta_{P_B}(s_B)\right| < TAS \ \& \ \left|(s_A/s_{P_A} - s_B/s_{P_B})\max(s_{P_A}, s_{P_B})\right| < TLS \ \& \\
&s_B(0) = s_B(i)/\min(AGA(P_A - P_B(s_i))), \ i \in [0, 1], \ \left|\theta_{P_B}(s_i) - \theta_{P_B}(s_i + \Delta)\right| > 0
\end{aligned}
\tag{3}
$$

In addition, the possible residual anomalous values of $L_V(A, B)$ distance computed for each pair of points that belong to homologous polygons were removed by means of computing two additional parameters: the average value of $L_V(A, B)$ ($\mu_{Lv(A,B)}$) and the root mean squared error of $L_V(A, B)$ ($RMSE_{Lv(A,B)}$). These threshold values also acted as control parameters to prevent the acceptance of unpaired points derived from the difference in the complexity of the shape of polygons due to a possible different level of generalization and scale of representation.

Figure 6, Row 2, shows a pair of matched polygons. The polygon P_A belonging to MTA10 (larger scale representation GDB) is more complex and has a more detailed shape than its counterpart B belonging to the BCN25 (GDB with a smaller scale representation and higher level of generalization). In this polygon P_A, there are two vertexes with no counterpart on P_B polygon; therefore, the differences included in the turning function $\theta_{P_A}(s)$ are not present in the turning function $\theta_{P_B}(s)$. If we enlarge

the threshold values, these vertexes could be matched with one vertex from Polygon B; however, the minimum *Lv* distance removes these homologous points.

5. Experimental Results

5.1. RCGA Parameters Setting

The experimental parameters that define the RCGA behavior are mainly five: the size of the initial population, number of generations, α parameter (crossover operator), β parameter and probability of mutation (mutation operators). The values assigned to the different parameters are included in Table 1. In addition, the relation between the size of the initial population and number of generations has also been highlighted in Section 3.2, so these two parameters are usually defined together using RCGA speed convergence as the criterion for their definition. In order to determine their values, an experimental test was carried out. The only objective of this test was to determine the effects that different initial populations have on the convergence (number of generations required to reach the stability of the fitness function) and the final fitness function value. To do this, a supervised database of homologous polygons is needed. The supervised database is composed of 800 pairs of polygons matched by specialized cartographers. This database was fragmented into five samples composed of 60, 70, 80, 90 and 100 pairs of polygons each, and 15 chromosomes, different possible solutions, was the initial value selected as the starting population. The experimental results show that the RCGA convergence is reached after 45 or 50 generations (Figure 7b) (when there is no change in the fitness function value (95%), Figure 7a). Therefore, 45–50 generations should be sufficient to reach an adequate solution for the mentioned size of the initial population. However, it is advisable to exceed the number of generations in order to ensure convergence results, so the final value given to the number of generations was 100. Taking all this into account, and using the ratio value (size of population/number of generations) of 1/5 suggested by Correa et al. (2009) [77], the ultimate size of the initial population was 20 chromosomes.

Table 1. Parameter values assigned to the RCGA.

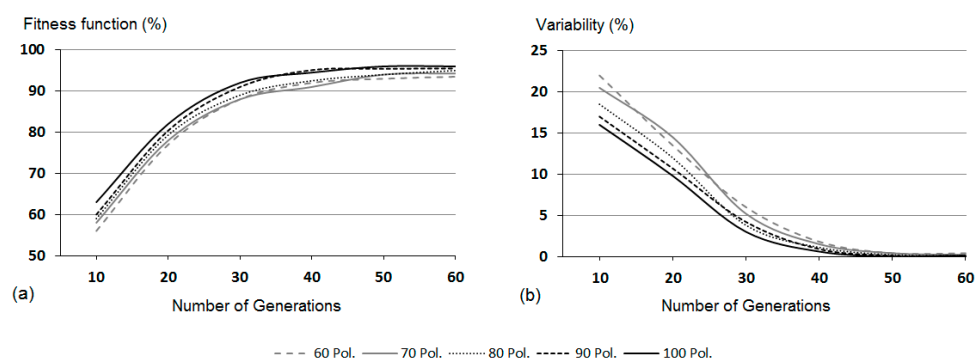| Parameter	Assigned Value
Size of initial population	20
Number of generations	100
α parameter	0.5
probability of mutation	0.005
β parameter	0.3

Figure 7. RCGA convergence with different initial populations.

Finally, the parameter values for the crossover and mutation operators needs to be specified. We considered $\alpha = 0.5$, a choice allowing us to balance exploration and exploitation [53], since the new gene has the same probability of lying inside or outside the interval defined by its parents. With regard to β

and the probability of mutation, the choice of their values (0.3 and 0.005, respectively) was suggested by the literature [80–82].

5.2. Resulting Weights

The resulting weights from the RCGA are shown in Table 2. These values correspond to the optimal solution computed in the sense of the maximum value of the fitness function for all the evolutionary paths explored. As shown in Table 2, all the characteristics of a polygon are perfectly represented by just three descriptors: MBR, moment of inertia and AGA. This result shows that the first idea of the cartographer in order to match two polygons features is: (i) the spatial position of both polygons (MBR); (ii) then their general orientation (moment of inertia); and finally, (iii) the differences between vertex and the orientation of the sides (AGA). The other descriptors have lesser significance (less than 8%) and represent redundant information for the RCGA.

Table 2. Resulting weights from the RCGA for the descriptors shown in Section 3.1. AGA, Arkin Graph Area.

Descriptor	Value
MBR	0.520
Perimeter	0.069
Area	0.006
Number of concave angles	0
Number of convex angles	0
Moment of inertia	0.285
AGA	0.119
Fitness (%)	0.946

The RCGA weights shown in Table 2 could be a little surprising because some standard descriptors, like area or number of concave and convex angles, have almost no influence on the matching indicator, MAV. First, we must remember that both datasets were removed from external influences on coordinates, that is both datasets were in the same coordinate reference system. Second, the datasets have few temporal differences. With these considerations in mind, MBR includes area differences, and AGA includes variation in angles.

5.3. Shape Similarity Measure and Matching of Polygonal Shapes

Once the supervised training stage of the GA algorithm was carried out (over the supervised training database), a classification of the other features from the proposed GDB was achieved by applying Equation (1) to each pair of polygons from MTA10 and BCN25 and keeping that pair of maximum MAV. With this classification, 8863 polygons from BCN25 and 9067 polygons from MTA10 were categorized from a quantification of their shape similarity. The results obtained (Table 3) show the percentage distribution of matched polygons for each interval of accuracy (MAV) classified into bad matching results (MAV < 0.5), intermediate matching results ($0.5 \leq$ MAV < 0.8) and good matching results (MAV $\geq$ 0.8). We must emphasize that this value represents the shape similarity measure between two matched polygons. With regards to matching 1:0, unmatched polygons, this type of cardinality in the matching process has no sense in our approach. Because all polygons have several MAV values inside a configurable radius of search, we assume that the maximum MAV value is the *1:1* matching. For this reason, the unmatched polygons are included in the bad matching results.

Table 3. Percentage distribution of polygons according to Match Accuracy Value (MAV) value.

Sheet	Urban Area Denomination	Number of Polygons BCN25/MTA10	Polygons Matched	Percent Distribution of Polygons According to MAV Value		
				Bad	Intermediate	Good
0985	Mairena del Alcor	847/875	772	33.41	34.09	32.5
0985	Alcalá de Guadaira	1070/1096	1061	21.97	37.88	40.15
0985	Carmona	851/870	810	23.08	25.81	51.11

Table 3. *Cont.*

Sheet	Urban Area Denomination	Number of Polygons BCN25/MTA10	Polygons Matched	Percent Distribution of Polygons According to MAV Value		
				Bad	Intermediate	Good
1009	Fuente Vaqueros	568/591	563	28.43	38.73	32.84
1009	Santa Fe	649/670	642	31.93	44.42	23.65
1009	Granada	2250/2301	2223	24.83	30.1	45.07
1066	Coín	438/454	433	30.25	40.87	28.88
1066	Alhaurín de la Torre	342/349	342	27.47	42.99	29.54
1066	Fuengirola	1848/1861	1830	20.65	41.09	38.26

In addition, it also sets out the results of the percentage distribution of matched polygons for each interval of accuracy (MAV) grouped by number of vertexes. This is because, although AGA is a powerful shape descriptor (as shown in Table 2), its value is not very intuitive compared to the number of vertexes of a polygon. Figure 8a–i displays these data in the form of a bar chart for each urban area under analysis.
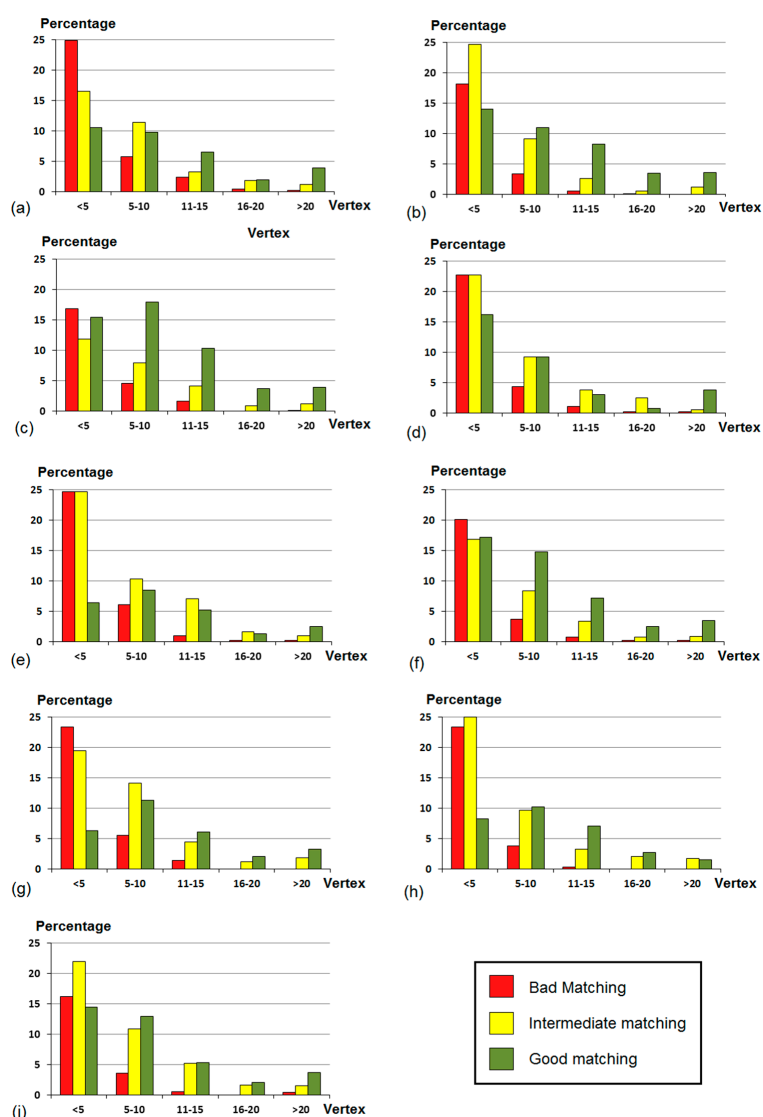

Figure 8. Percentage distribution of matched polygons for each interval of accuracy (MAV) grouped by number of vertexes of the BTN25 for each urban area under analysis: (**a**) Mairena del Alcor, (**b**) Alcalá de Guadaira, (**c**) Carmona, (**d**) F. Vaqueros, (**e**) Santa Fe, (**f**) Granada, (**g**) Coín, (**h**) Alhaurín de la Torre, (**i**) Fuengirola

The graphics from Figure 8 can help give a better understanding of these results. Thus, we note that for high values of MAV, the total number of matched polygons with a high number of vertex increases, as well. However, for low values of MAV, the number of matched polygons with fewer than five vertexes is very high. This result shows that it is more difficult to match similar shape polygons, which have fewer vertexes than more representative polygons with high numbers of vertexes. In order to corroborate these remarks statistically, an assessment of the normality of data was carried out using the mean value of MAV and the standard deviation as comparing parameters. Although the results do not show significant differences between the mean values of MAV, they show some differences between polygons with fewer than 10 vertexes and the other polygons. This corroborates the idea that for high values of MAV, the total number of polygons with a high number of vertexes increases, as well. In addition, the results show homogeneous behavior between the sheets with regard to MAV values.

As mentioned in Section 3.2, not all the matched polygons were used to compute the matching between points. Thus, only those pairs of polygons with a MAV value higher than a fixed threshold were used. The threshold was set at 0.8, which has been categorized as demonstrating good matching results. The choice of this threshold value was made in order to avoid the acceptance of erroneously-matched polygons (false positives or error of commission) and was fixed on the basis of the confusion matrix computed from the supervised training database.

Finally, Figure 9a shows an example of the spatial distribution of polygons according to the MAV value achieved, or in other words the shape similarity measure between polygons. More specifically, it shows the spatial distribution corresponding to a part of the urban area of Alhaurín de la Torre (1066 sheet). In this figure, the matched polygons are distributed throughout the whole area, which is a good indicator that the matching algorithm is not affected by spatial position. With regards to polygons with intermediate MAV values, these represent high vertex and AGA differences between MTA10 and BCN25 (e.g., the polygons with "c" shapes with no counterpart in MTA10, marked as 1 and enlarged in Figure 9b). Finally, well-matched polygons include both polygons that can be considered equal with low differences in AGA, position and with the same number of vertexes to polygons with vertex differences and AGA. The first include small rectangular polygons on the bottom left of the image, marked as 2 and enlarged in Figure 9c. The second include polygons with more differences, e.g., the green polygon with a red one inside (marked as 3 and enlarged in Figure 9d).
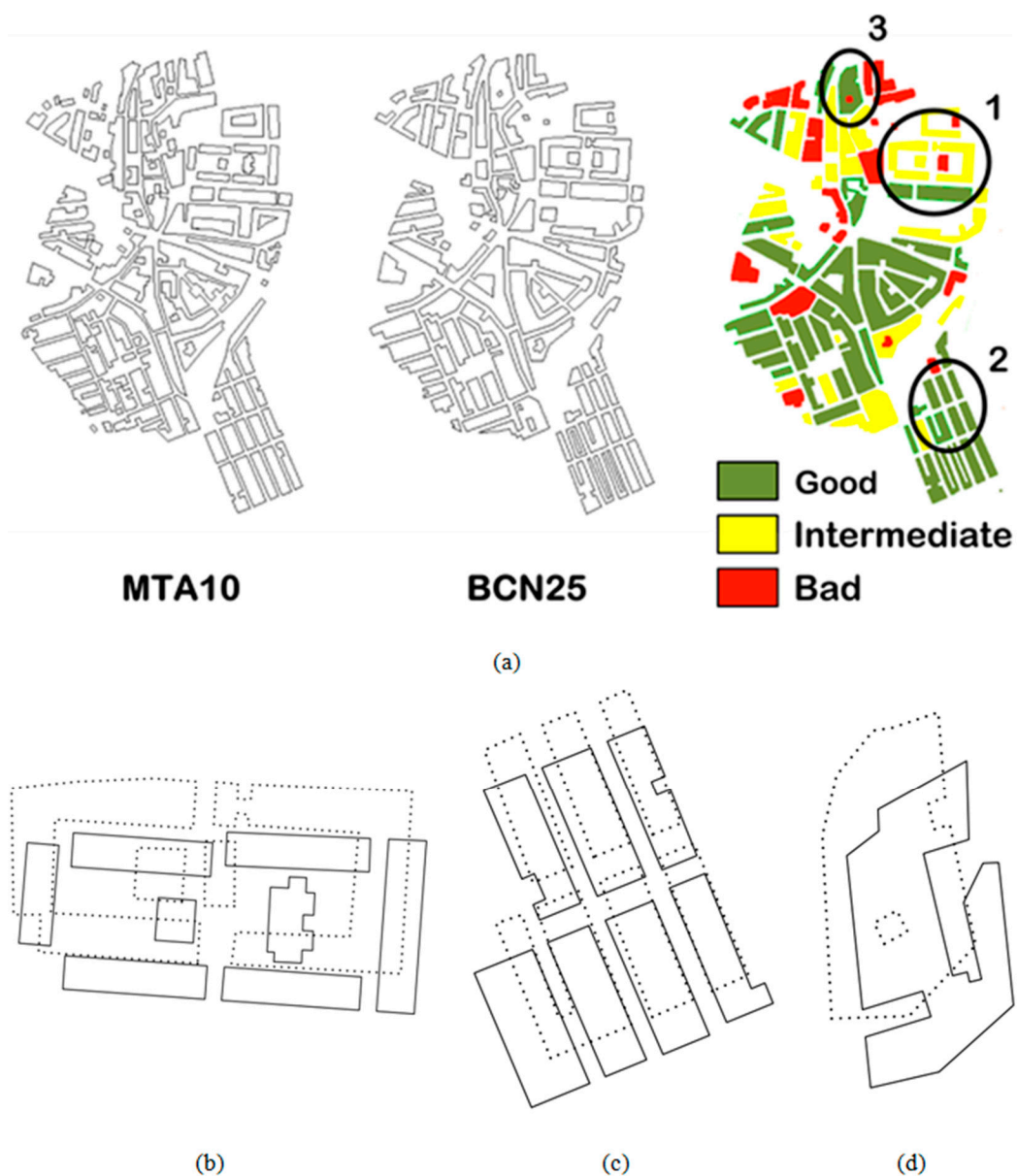
Figure 9. (**a**) Spatial distribution of matched polygons according to MAV values (Alhaurín de la Torre, 1066 sheet). (**b–d**) Enlargement of Zones 1, 2 and 3, respectively (not at a defined scale). Continuous lines represent MTA10, and dot lines represent BCN25.

5.4. Vertex-to-Vertex Matching

5.4.1. Thresholds Setting

As mentioned in Section 4, the thresholds used in this stage include: (1) a threshold fixed for length shifts (TLS), which represent variations on the *s* axis of the turning functions due to length changes of the polygon sides; and (2) a Threshold fixed for Angular Shifts (TAS), which represent variations on the θ axis of the turning functions due to rotation changes of the polygon sides (see Figure 5). The impact of these threshold values on the final matching process is significant. Therefore, in order to set the TLS and TAS values, a sample of 32 pairs of polygons belonging to the supervised training database with a good matching (MAV ≥ 80%) and with a number of vertices ranging from 3–20 was used. Over this sample, pairs of homologous points were identified, taking into account all possible combinations of TAS and TLS including values within the following ranges: TAS range from 0–1 rad

(with Δ = 0.2 rad), and TLS range from 0–10 m (with Δ = 2 m) (this last value must be rescaled for each case). Finally, for each setting of threshold values, a confusion matrix was computed [83]. In this matrix, false negatives (errors of omission) were defined by the rejection of rightly matched pairs of points, and errors of commission were defined by the acceptance of erroneously-matched pairs of points. The last (commission) is the type of error that must be avoided as much as possible because it is one that adds noise to the matching process. With this last consideration in mind, the criterion for setting the thresholds was to maximize the number of well-matched points and to minimize the number of erroneously-matched points that were accepted. Table 4 shows the results for the 25 possible combinations of TLS and TAS and the values of well-matched points versus commissions and the ratio between both of them. This table includes the optimal combination that we were looking for.

Table 4. Percentage distribution of matched points according to the Threshold fixed for Angular Shifts (TAS) and Threshold fixed for Length Shifts (TLS) values.

m	0.2 Rad		0.4 Rad		0.6 Rad		0.8 Rad		1.0 Rad	
	Well-Matched/ Commission	Ratio	Well-Matched/ Commission	Ratio	Well-Matched/ Commission	Ratio	Well-Matched/ Commission	Ratio	Well-Matched/ Commission	Ratio
2	8.5 1.6	**5.31**	12.1 2.9	4.17	17.8 9.3	1.91	18.8 14.1	1.33	12.3 19.9	0.62
4	11.5 2.8	4.11	30.7 3.4	**9.03**	28.3 12.2	2.32	28.6 18.7	1.53	14.2 23.6	0.60
6	15.7 7.5	2.09	32.1 8.2	3.91	35.5 15.3	**2.32**	39.1 25.6	1.53	21.6 31	0.70
8	18.9 12.2	1.55	35.8 14.3	2.50	37.7 20.1	1.88	38.1 27.7	**1.38**	23.7 32.8	0.72
10	21.3 19.8	1.08	33.6 21.35	1.57	37.1 25.9	1.43	24.9 33.3	0.75	27.4 35.3	**0.78**

The first option (0.2 rad/2 m) presents the lowest percentage of errors of commission. However, this option was discarded due to the low percentage of correct matches provided. In this sense, we can conclude that the values assigned to the parameters are too restrictive for our purposes. The opposite occurs in the last three options (0.6 rad/6 m, 0.8 rad/8 m and 1 rad/10 m). Thus, although significant percentages of correct matches were obtained, they were achieved at the cost of also including high error rates. Therefore, these options were also discarded. Therefore, the option chosen was the second (with values assigned to the parameters of 0.4 rad and 4 m, respectively). This option was selected because it presents the maximum ratio (Well-matched/commissions) and the most suitable option according to the criteria mentioned above.

5.4.2. Vertex Matching

Table 5 shows the results of the matching between points for each urban area and under the TLS and TAS values previously discussed. In this table, the total number of vertexes refers to the sum of all the vertexes of the polygons belonging to each BDG, while the number of matched points relates to those that are detected on pairs of polygons matched with an MAV higher than 0.8 (as mentioned in Section 5.3).

Table 5. Percentage distribution of matched points for each urban area.

Sheet	Urban Area Denomination	Number of Polygons BCN25/MTA10	All MAV				MAV > 80%		
			Polygons Matched	Total Number of Vertexes		% of Matched Polygons	Number of Matched Vertex	% of Matched Vertex	
				BCN25	MTA10			BCN25	MTA10
0985	Mairena del Alcor	847/875	772	4421	6179	32.5	452	10.22	7.31
0985	Alcalá de Guadaira	1070/1096	1061	6411	7549	40.15	324	5.05	4.29
0985	Carmona	851/870	810	5276	6606	51.11	579	10.97	8.76
1009	Fuente Vaqueros	568/591	563	2374	3787	32.84	624	26.28	16.47
1009	Santa Fe	649/670	642	2696	4543	23.65	352	13.05	7.74
1009	Granada	2250/2301	2223	8999	16,144	45.07	3378	37.53	20.92
1066	Coín	438/454	433	2124	3153	28.88	355	16.71	11.25
1066	Alhaurín de la Torre	342/349	342	1563	2462	29.54	275	17.59	11.17
1066	Fuengirola	1848/1861	1830	10,946	15,169	38.26	650	5.93	4.28
	TOTAL	8863/9067	8676	44,810	65,592	38.67	6989	15.60	10.66

The total number of matched vertexes was 6989 from the set of matched polygons with an MAV higher than 0.8. These represent 15.6% of the vertex total of BCN25 (45.7% of vertexes from well-matched polygons) and 10.7% of MTA10 (37.2% of vertexes from well matched polygons). The percentage results achieved are greatly conditioned not only by the restrictive values of TAS and TLS, which act as control parameters to prevent the acceptance of unpaired points derived from the difference in the complexity of the shape of polygons due to a possible different level of generalization and scale of representation, but also by the number of polygons used, which is limited by the threshold value of MAV fixed in the process. With regards to individual pairs of GDB, dense zones have a higher percentage of matched polygons. However, this will not represent a high percentage of vertex matching, e.g., 45% of matched polygons and 37% (21%) of matched vertexes in Granada versus Fuengirola with 38% of matched polygons and only 6% (4%) of matched vertexes. This is because vertex matching is a completely different process that depends on the complexity of the shape, which is different in MTA10 and BCN25.

Finally, Figure 10 shows an example of the spatial distribution of matched vertexes corresponding in this case to a part of the urban area of Granada (sheet 1009). As shown in this figure, our method determines a high number of points with high spatial accuracy. However, there are still some matched polygons without matched vertexes; this may be due to differences in the precision of matching between both algorithms, which in turn indicates a more flexible inter-element matching and a more restrictive intra-element matching. This is not a defect, but a design in this two-step matching approach, because our interest is that the matching should be enough to detect different polygon shapes as the same one, but not allow high differences in distances or angles of the vertex in order to maintain positional precision. This can be observed in the polygons with complex shapes in MTA10 and simplified forms in BCN25 (marked with a circle in Figure 10).
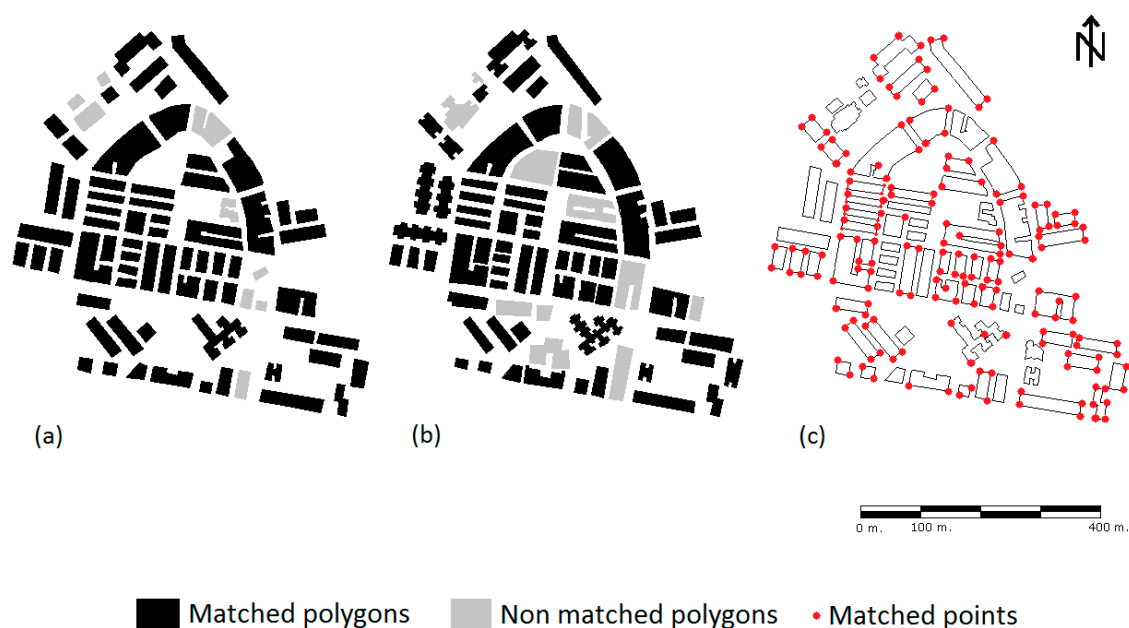

Figure 10. (**a**) Spatial distribution of matched and non-matched polygons (BCN25). (**b**) Spatial distribution of matched and non-matched polygons (MTA10). (**c**) Spatial distribution of matched point (Granada, Sheet 1009).

5.4.3. Computation Time

Finally, an important advantage of the application of the metrics described above in computing the correspondences between homologous vertex (vertex matching) is the low computational time required compared to traditional methodologies (especially when applied to a large number of GDBs, and if we

consider field work for GNSS data acquisition). This is particularly significant measuring the efficiency of our approach by means of the ratio (number of points/minutes) achieved: (20–25/60) applying this methodology in a manual way (traditional method); and (452/1.5), (324/1.2), (579/1.8), (624/1.9), (352/1.2), (3378/11.2), (355/1.2), (275/1.7), (650/2.1) applying this methodology automatically for the eight cases presented: Mairena del Alcor, Alcalá de Guadaira, Carmona, Fuente Vaqueros, Santa Fe, Granada, Coín, Alhaurín de la Torre and Fuengirola, respectively. Logically, the efficiency of our approach depends on the size of the GDBs employed (number of polygons of both datasets: tested GDB and reference GDB, and total number of vertices). Finally, we must note that our test platform was an Intel Core i5 dual-core 2.7 GHz with 8 GB memory.

6. Conclusions

An efficient methodology for matching geospatial features of geographic information, based on homologous vertexes, has been presented. The key feature of our approach is being able to solve the problems that characterize traditional positional assessment methods based on points, which are associated with the use of a limited number of points. In order to achieve a greater number of highly accurate points, two main tasks have been proposed and successfully performed. First, polygonal shapes from two datasets (in our case, GDBs) have been matched on the basis of their similarity (inter-element matching). Second, a vertex-to-vertex correspondence between these previously matched shapes has been achieved (intra-element matching).

With regard to the first task, RCGA has shown itself to be an appropriate and efficient search method on matching problems where spatial variables are implied. Rather than identifying pairs of homologous objects, the proposed framework appeals directly to the low-level descriptors measurement in order to classify data features as similar or non-similar to model features and to assign probabilities to those classes. Therefore, RCGA provides a quality match measure. Using these descriptors, a new parameter has been proposed, MAV, which has achieved 38.7% of total polygons with a good matching quality (more than 80% similarity).

With regard to the second task, the use of the metric defined by Arkin et al. [49] in a new way, that is for matching two set of vertexes, has provided us not only with the capability to obtain, in an unassisted way, a high quantity of well-matched vertexes for each polygon, but also a powerful edge-based shape descriptor. The proposed matching achieves more than 10% total homologous vertexes, with a high positional accuracy, which represents hundreds of matched vertexes for each GDB. This value is greater than the number of proposed points, near 20, in standard positional quality assessment based on points. In addition, a second advantage of the metrics proposed is the low computational time required compared to traditional methodologies (especially when applied to a large number of GDBs, and if we consider field work for GNSS data acquisition).

The process can however be further improved. For this reason, in future studies, we plan to explore several directions. Thus, a new set of GDBs with different density both in polygons and number of vertexes is being tested. Moreover, two improvements could be implemented: The first is to attain *1:n* relationships in inter-element matching instead of *1:1*. The second improvement, in intra-element matching, is planned in order to achieve vertex matching from polygons with great differences in AGA.

Author Contributions: The research was conducted by the main author, Juan J. Ruiz-Lendínez, under the supervision of the co-authors Manuel A. Ureña-Cámara and Francisco J. Ariza-López. All authors jointly drafted and critically revised the paper. All authors read and approved the final manuscript.

References

1. Tang, J.; Jiang, B.; Zheng, A.; Luo, B. Graph matching based on spectral embedding with missing value. *Pattern Recognit.* **2012**, *45*, 3768–3779. [CrossRef]

2. 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]

3. Oxera. What Is the Economic Impact of Geoservices? Prepared for Google. WWW Document. Available online: http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services_1.pdf (accessed on 3 December 2017).

4. 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.

5. 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.

6. Song, W.; Keller, J.; Haithcoat, T.; Davis, C. Relaxation-Based Point Feature Matching for Vector Map Conflation. *Trans. GIS* **2011**, *15*, 43–60. [CrossRef]

7. Ariza-López, F.J.; Atkinson-Gordo, A. Variability of NSSDA Estimations. *J. Surv. Eng.* **2008**, *134*, 39–44. [CrossRef]

8. Federal Geographic Data Committee (FGDC). *Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data Accuracy*; FGDC-STD-007; FGDC: Reston, VA, USA, 1998.

9. Minnesota Planning Land Management Information Center (MPLMIC). *Positional Accuracy Handbook*; MPLMIC: St. Paul, MN, USA, 1999.

10. Li, Z. Effects of check points on the reliability of DTM accuracy estimates obtained from experimental test. *ISPRS J. Photogramm. Remote Sens.* **1991**, *5710*, 1333–1340.

11. National Joint Utilities Group (NJUG). *Quality Control Procedure for Large Scale Ordnance Survey Maps Digitized to OS, Version 1*; Publication No. 13; NJUG: London, UK, 1988.

12. Newby, P. Quality management for surveying, photogrammetry and digital mapping at the ordnance survey. *Photogramm. Rec.* **1992**, *7914*, 45–58. [CrossRef]

13. Atkinson-Gordo, A. Control de Calidad Posicional en Cartografía: Análisis de los Principales Estándares y Propuesta de Mejora. Ph.D. Thesis, University of Jaén, Jaén, Spain, 2005, Unpublished.

14. Goodchild, M.; Hunter, G. A Simple Positional Accuracy Measure for Linear Features. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 299–306. [CrossRef]

15. Ruiz-Lendinez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A.E. Automatic positional accuracy assessment of geospatial databases using line-based methods. *Surv. Rev.* **2013**, *45*, 332–342. [CrossRef]

16. Ruiz-Lendinez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A.E. A point-based methodology for the automatic positional accuracy assessment of geospatial databases. *Surv. Rev.* **2016**, *48*, 269–277. [CrossRef]

17. Tong, X.; Shi, W.; Deng, S. A probability-based multi-measure feature matching method in map conflation. *Int. J. Remote Sens.* **2009**, *30*, 5453–5472. [CrossRef]

18. Walter, V.; Fritsch, D. Matching spatial data sets: A statistics approach. *Int. J. Geogr. Inf. Sci.* **1999**, *13*, 445–473. [CrossRef]

19. Samal, A.; Seth, S.; Cueto, K. A feature-based approach to conflation of geospatial sources. *Int. J. Geogr. Inf. Sci.* **2004**, *18*, 459–489. [CrossRef]

20. Xiong, D.; Sperling, J. Semiautomated matching for network database integration. *ISPRS J. Photogramm. Remote Sens.* **2004**, *59*, 35–46. [CrossRef]

21. Deng, M.; Li, Z.; Chen, X. Extended Hausdorff distance for spatial objects in GIS. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 459–475.

22. Mustière, S.; Devogele, T. Matching networks with different levels of detail. *Geoinformatica* **2007**, *12*, 435–453. [CrossRef]

23. Zhang, M. Methods and Implementations of Road-Network Matching. Ph.D. Thesis, Institute for Photogrammetry and Cartography, Technical University of Munich, Munich, Germany, 2009.

24. Kim, J.; Huh, Y.; Yu, K.; Kim, J.O. Automatic change detection based on areal feature matching in different network data-sets. *J. Korean Soc. Surv. Photogramm. Cartogr.* **2013**, *31*, 483–492. [CrossRef]

25. Li, L.; Goodchild, M. An optimisation model for linear feature matching in geographical data conflation. *Int. J. Image Data Fusion* **2011**, *2*, 309–328. [CrossRef]

26. Koukoletsos, T.; Haklay, M.; Ellul, C. Assessing data completeness of VGI through an automated matching procedure for linear data. *Trans. GIS* **2012**, *16*, 477–498. [CrossRef]

27. Yang, B.; Zhang, Y.; Luan, X. A probabilistic relaxation approach for matching road networks. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 319–338. [CrossRef]

28. Hastings, J.T. Automated conflation of digital gazetteer data. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 1109–1127. [CrossRef]

29. 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]

30. 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]

31. Mortara, M.; Spagnuolo, M. Similarity measures for blending polygonal shapes. *Comput. Graph.* **2001**, *25*, 13–27. [CrossRef]

32. 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]

33. 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]

34. 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.

35. Zhang, X.; Zhao, X.; Molenaar, M.; Stoter, J.; Kraak, M.; Ai, T. Pattern classification approaches to matching building polygons. In Proceedings of the XXII ISPRS Congress, Melbourne, Australia, 25 August–1 September 2012; pp. 19–24.

36. Breiman, L.; Friedman, J.; Stone, C.; Olshen, R. *Classification and Regression Trees*; Chapman & Hall: New York, NY, USA, 1984.

37. Quinlan, J. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Mateo, CA, USA, 1993.

38. Surazhsky, T.; Surazhsky, V.; Barequet, G.; Tal, A. Blending polygonal shapes with different topologies. *Comput. Graph.* **2001**, *25*, 29–39. [CrossRef]

39. Yang, W.; Feng, J. 2D Shape morphing via automatic feature matching and hierarchical interpolation. *Comput. Graph.* **2009**, *33*, 414–423. [CrossRef]

40. Prakash, S.; Robles-Kelly, A. Geometric graph comparison from an alignment view point. *Pattern Recognit.* **2012**, *45*, 3780–3794. [CrossRef]

41. Veltkamp, R.; Hagedoorn, M. Shape similarity measures, properties and constructions. In Proceedings of the Fourth International Conference on Advances in Visual Information Systems, Lyon, France, 2–4 November 2000.

42. Gösseln, G.; Sester, M. Semantic and geometric integration of Geoscientific data sets with ATKIS-applied to geo-objects from geology and soil science. In Proceedings of the ISPRS Commission IV Joint Workshop, Stuttgart, Germany, 8–9 September 2003; pp. 111–116.

43. Butenuth, M.; Gösseln, G.; Tiedge, M.; Heipke, C.; Lipeck, U.; Sester, M. Integration of heterogeneous geospatial data in a federated database. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 328–346. [CrossRef]

44. Zhang, M.; Meng, L. An iterative road-matching approach for the integration of postal data. *Comput. Environ. Urban Syst.* **2007**, *31*, 597–615. [CrossRef]

45. Safra, E.; Kanza, Y.; Sagiv, Y.; Beeri, C.; Doytsher, Y. Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets. *Int. J. Geogr. Inf. Syst.* **2010**, *24*, 69–106. [CrossRef]

46. Huh, Y.; Yang, S.; Ga, C.; Yu, K.; Shi, W. Line segment confidence region-based string matching method for map conflation. *ISPRS J. Photogramm. Remote Sens.* **2013**, *78*, 69–84. [CrossRef]

47. McAuley, J.; Caetano, T. Fast matching of large point sets under occlusions. *Pattern Recognit.* **2012**, *45*, 563–569. [CrossRef]

48. Kaygin, S.; Bulut, M. Shape recognition using attributed string matching with polygon vertices as the primitives. *Pattern Recognit. Lett.* **2002**, *23*, 287–294. [CrossRef]

49. Arkin, E.; Chew, L.; Huttenlocher, D.; Kedem, K.; Mitchell, J. An Efficiently Computable Metric for Computing Polygonal Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 209–216. [CrossRef]

50. Holland, J. *Adaptation in Natural, Artificial Systems*; MIT Press: Cambridge, MA, USA, 1975.

51. Goldberg, D. Real Coded Genetic Algorithms, Virtual Alphabets, and Blocking. *Complex Syst.* **1991**, *5*, 139–167.

52. Forrest, S. Genetic algorithms: Principles of natural selection applied to computation. *Science* **1993**, *261*, 872–878. [CrossRef] [PubMed]

53. Herrera, F.; Lozano, M.; Verdegay, J. Tackling Real-Coded Genetic Algorithms: Operators and Tools for Behavioral Analysis. *Artif. Intell. Rev.* **1998**, *12*, 265–319. [CrossRef]

54. 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]

55. Hsiao, H.; Wong, K. Automatic Matching of Buildings and Corners. *Photogramm. Eng. Remote Sens.* **1999**, *65*, 803–810.

56. Myers, E.; Hancock, E. Least-commitment Graph Matching with Genetic Algorithms. *Pattern Recognit.* **2001**, *34*, 375–394. [CrossRef]

57. Alba, E.; Dorronsoro, B. *Cellular Genetic Algorithms*; Springer: New York, NY, USA, 2008.

58. Sánchez, A.; Lozano, M.; Villar, P.; Herrera, F. Hybrid Crossover Operators with Multiple Descendents for Real-Coded Genetic Algorithms: Combining Neighborhood-Based Crossover Operators. *Int. J. Intell. Syst.* **2009**, *24*, 540–567. [CrossRef]

59. Xavier, E.; Ariza-López, F.; Ureña-Cámara, M. A survey of measures and methods for matching geospatial vector datasets. *ACM Comput. Surv.* **2016**, *49*, 39. [CrossRef]

60. Ruiz-Lendinez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A.; Blázquez, E. Digital Map Conflation: A Review of the Process and a Proposal for Classification. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1439–1466. [CrossRef]

61. National Geographic Institute of Spain (NGI). Cartographic Series: 1:25000 and 1:50000 Series. Available online: http://www.ign.es/ign/layout/series.do (accessed on 3 December 2017).

62. Casado, M. Some Basic Mathematical Constraints for the Geometric Conflation Problem. In Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lisbon, Portugal, 5–7 July 2006.

63. Cartographic Institute of Andalusia (ICA). Mapa Topográfico de Andalucía 1:10.000 Vectorial (MTA10v_2001). Available online: http://www.ideandalucia.es/catalogo/info.php (accessed on 3 December 2017).

64. Nguyen, D.; Ogunbona, P.; Li, W. A novel shape-based non-redundant local binary pattern descriptor for object detection. *Pattern Recognit.* **2013**, *46*, 1485–1500. [CrossRef]

65. Shih, J.; Lee, C.; Wang, J. A new 3D model retrieval approach based on the elevation descriptor. *Pattern Recognit.* **2007**, *40*, 283–295. [CrossRef]

66. Zhang, C.; Chen, T. Efficient feature extraction for 2D/3D objects in mesh representation. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Thessaloniki, Greece, 7–10 October 2001.

67. Vranic, D.; Saupe, D.; Richter, J. Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics. In Proceedings of the IEEE Workshop on Multimedia Signal Processing, Cannes, France, 3–5 October 2001.

68. Osada, R.; Funkhouser, T.; Chazelle, B.; Dobkin, D. Shape distributions. *ACM Trans. Gr.* **2002**, *21*, 807–832. [CrossRef]

69. Eshelman, L.; Schaffer, J. Real-coded genetic algorithms and interval-schemata. In *Foundations of Genetic Algorithms 2*; Whitley, L., Ed.; Morgan Kaufmann Publishers: Burlington, MA, USA, 1993; pp. 187–202.

70. Mühlenbein, H.; Schlierkamp, D. Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization. *Evol. Comput.* **1993**, *1*, 25–49. [CrossRef]

71. Chun, J.; Chiang, H.; Chih-Wei, T. Structure-specified Real Coded Genetic Algorithms with applications. In *Advanced Knowledge Based Systems: Model, Applications & Research*; Sajja, P., Akerkar, R., Eds.; TMRF e-Book: Kolhapur, India, 2010; pp. 160–187.

72. Zhang, G.; Liu, W. A Genetic Algorithm Based on a New Real Coding Approach. In Proceedings of the Second International Conference on Intelligent System Design and Engineering Application, Sanya, China, 6–7 January 2012.

73. Ruiz-Lendinez, J.J. Automatización del Control de Calidad Posicional de la Cartografía. Ph.D. Thesis, Escuela Politécnica Superior de Jaén, Universidad de Jaén, Jaén, Spain, 2012.

74. Ruiz-Lendinez, J.J.; Ureña-Cámara, M.A.; Ariza-López, F.J. Matching Viewer: Herramienta para la automatización del control de calidad posicional de la cartografía. *Rev. Colegio Oficial de Ingeniería Geomática y Topográfica* **2016**, *35*, 37–45.

75. Knuth, D. Extensive coverage of statistical tests for non-randomness. In *The Art of Computer Programming*; Addison-Wesley: Reading, MA, USA, 1997; pp. 1–193.

76. Maaranen, H.; Miettinen, K.; Penttinen, A. On initial populations of a genetic algorithm for continuous optimization problems. *J. Glob. Optim.* **2007**, *37*, 405–436. [CrossRef]

77. Correa, A.; Rodríguez, E.; Londoño, M. Diseño experimental para la evaluación de un Algoritmo Genético de Programación de Producción. In Proceedings of the Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática, Orlando, FL, USA, 10–13 July 2009.

78. Brindle, A. Genetic Algorithms for Function Optimization. Ph.D. Thesis, University of Alberta, Edmonton, AB, Canada, 1991.

79. Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs*; Springer-Verlag: Berlin/Heidelberg, Germany, 1992.

80. López, J.A.; Herrera, F. Aprendizaje de Multiclasificadores Evolutivos (AME). Ph.D. Thesis, University of Granada, Granada, Spain, 2002, Unpublished.

81. Back, T. Mutation parameters. In *Evolutionary Computation 2: Advanced Algorithms and Operators*; Back, T., Fogel, D., Michalewicz, Z., Eds.; Institute of Physics Press: Bristol, UK, 2000; pp. 142–151.

82. Vajda, P.; Eiben, A.; Hordijk, W. Parameter Control Methods for Selection Operators in Genetic Algorithms. In *Lecture Notes in Computer Science, Parallel Problem Solving from Nature—PPSN X, 13–17 September 2008*; Rudolph, G., Jansen, T., Lucas, S., Poloni, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 620–630.

83. Hardin, P.; Shumway, M. Statistical Significance of Normalized Confusion Matrices. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 735–740.