Survey Review

ISSN: 0039-6265 (Print) 1752-2706 (Online) Journal homepage: https://www.tandfonline.com/loi/ysre20

Method for an automatic alignment of imagery and vector data applied to cadastral information in Poland

J. J. Ruiz-Lendínez, B. Maćkiewicz, P. Motek & T. Stryjakiewicz

To cite this article: J. J. Ruiz-Lendínez, B. Maćkiewicz, P. Motek & T. Stryjakiewicz (2019) Method for an automatic alignment of imagery and vector data applied to cadastral information in Poland, Survey Review, 51:365, 123-134, DOI: 10.1080/00396265.2017.1388959

To link to this article: https://doi.org/10.1080/00396265.2017.1388959

Published online: 20 Oct 2017.

Method for an automatic alignment of imagery and vector data applied to cadastral information in Poland

J. J. Ruiz-Lendínez[1], B. Maćkiewicz[2], P. Motek[2] and T. Stryjakiewicz[2]

Nowadays, an important problem in combining vector data and imagery is that they rarely align. This problem can become particularly acute in the case of cadastral systems. In this study, and as part of the partnership between the Universities of Jaén and Adam Mickiewicz (Poznań), we provide a methodological proposal to assess the conflation procedures between cadastral vector data and imagery, improving the alignment between both data sets. To do this, we use an automatic alignment algorithm which detects road intersections from both data sets as control points by using image texture characterisation. With this method, we first train the system on the imagery to learn the road texture distribution, then we can obtain its segmentation according to its texture, and finally the system locates road intersection points. The last step is to align vector data and imagery by using different techniques. This algorithm is based on an earlier one, detailed in [Ruiz, J.J., Rubio, T.J., and Ureña, M.A., 2011b. Automatic extraction of road intersections from images in conflation processes based on texture characterization. *Survey review*, 43 (321), 212–225.]. However, in the updated version we have solved the problem of not-well-defined intersection points, resulting in a substantial increase in the number of intersection points employed for the final adjustment to align both products and in a reduction of the computation time. On the other hand, the positional uncertainty assessment of parcel boundary lines both before and after applying our alignment procedure between them is provided. With regard to the experimental results, in the case of Polish cadastral data this procedure allows for significant improvement in the alignment between imagery and cadastral parcels boundaries.

Keywords: Land register, Conflation, Image, Alignment, Vector data, Texture characterisation

Introduction

Nowadays, easy online accessibility and computational power have been the cause of an ever growing availability of geospatial information. Moreover, there is an increasing necessity to share this information between different users with the purpose of creating new products which are the result of integrating geospatial data sets from heterogeneous sources. Among the most significant examples (Fig. 1) of this new geospatial paradigm are the popularisation of crowdsourced data (also referred to as volunteered geographic information (Goodchild 2007), the development of web applications like Google Earth, and the creation of Spatial Data Infrastructures (Masó *et al.* 2012). As a result, the current trend in the geospatial domain requires more data integration and consequently more precise matching and alignment procedures (Xavier *et al.* 2016).

In this socio-technological context, cadastral cartography has not remained on the margins of the technical and conceptual evolution. Thus cadastral mapping systems are becoming more complex, since they integrate the structuring and modelling of vector geospatial databases (GDB) and the semantic component of imagery (Fig. 1*c*). This integration process between vector GDB and imagery is what some authors call vector-imagery conflation (VIC) (see Ruiz *et al.* 2011a). As these authors point out, VIC processes are a particular case of geometric alignment between two spatial data sets where the problem is defined as how to transform the features of one data set (vector GDB) onto another (imagery), minimising the geometric differences between them (Casado 2006). These differences are presented as a

[1]Dpto. de Ingeniería Cartográfica, Geodésica y Fotogrametría. Escuela Politécnica Superior de Jaén, University of Jaén, Jaén, Spain
[2]Department of Regional Policy and European Integration, Faculty of Geographical and Geological Sciences, Institute of Socio-Economic Geography and Spatial Management of Adam Mickiewicz University of Poznań, Poznań, Poland

Corresponding author, email lendinez@ujaen.es; jjruilen@gmail.com

Received 22 September 2016; accepted 28 September 2017
DOI 10.1080/00396265.2017.1388959

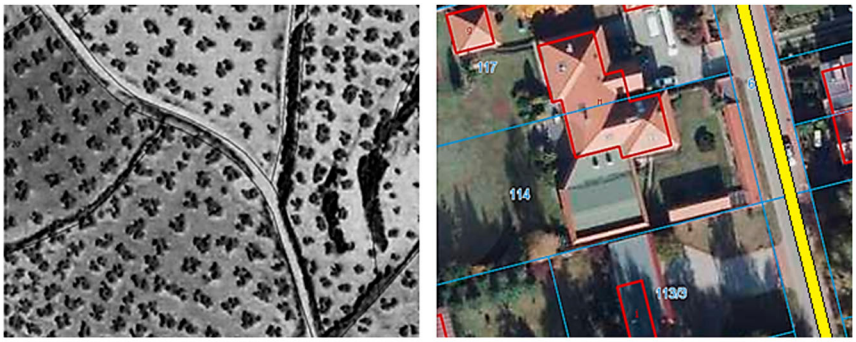
1 Examples of new cartographic products resulting from geospatial data integration (City of Poznań). *a* OpenStreetMap as an example of crowdsourced data, *b* Google Earth as an example of a web application, and *c* Cadastral parcel mapping as an example of integration between vector data and imagery (Source: http://poznan.podgik.pl/).

loss of positional interoperability between elements and are mainly due to several sources of error which lower the quality of the final product (denoted by the lack of geometric coherence). A highly representative example of this lack of geometric coherence between vector GDB and imagery in the web is the representation of cadastral parcel mapping. In many instances, by overlaying both data sets several elements such as roadsides, parcel boundary lines or buildings appear misaligned (Fig. 2). From our point of view, there are two basic reasons for this misalignment:

- The automated alignment tools employed have inaccurate matching mechanisms which are unable to establish robust correspondences between elements from both data sets.
- Alignment processes between vector GDB and imagery have certain characteristics (mainly related to the very nature of the imagery) that increase the complexity of the matching process.

Recent trends in VIC processes research have been focused on solving the two above-mentioned difficulties, trying to increase the robustness of the correspondences between data sets and thereby facilitating their subsequent alignment by means of several geometric adjustment operations. These operations are based on dimensional transformations, such as the Helmert transformation (Watson 2006), the affine transformation (Töbler 1994), the rubber-sheeting method (Gillman 1985, Doytsher 2000, Shimizu and Fuse 2003), the conformal transformations based on analytical functions (Ward-Brown and Churchill 2004) and the special transformation functions called 'multiresolution spline' (Brovelli and Zambroni 2004).

A key aspect to take into account in relation to VIC processes is the choice of the most suitable set of features for performing the matching procedure, because this choice will influence the subsequent capacity for establishing robust correspondences between the other features present in the vector GDB and imagery (for example, parcel boundaries in the case of cadastral information). In this sense, road intersections are the features most frequently used as candidates for matching points (Chen *et al.* 2006, Chiang *et al.* 2009). This is because, despite the fact that extracting roads from images is a complex problem principally due to noise leading to the inaccurate classification of road regions, road geometry around intersections is often well defined (Chiang *et al.* 2009). This characteristic allows their extraction to be carried out more efficiently than in the case of other features. In addition, the application of localised spatial filters such as localised template matching (LTM) (Chen *et al.* 2006) allows work on reduced areas around intersection roads and therefore the use of matching techniques with one-to-one correspondences. This is the simplest and



2 Examples of superposition of a vector GDB on an image. In both cases, we can appreciate the leak of geometric coherence between the vector elements (border of the path, parcel boundary line, or buildings) and the digital image. *a* Spanish case Source: Virtual Cadastral Office. https://ovc.catastro.meh.es and *b* Polish case Source: http://poznan.podgik.pl/.

most efficient approach to matching geospatial data (Xavier *et al.* 2016).

In recent years, many approaches have been developed to extract roads from images for urban areas. Following Alshehhi and Reddy (2017), road extraction methodologies can be mainly classified on the basis of two criteria: (i) they can be divided into either road-area extraction or road-centreline extraction, and (ii) extraction methods can be semi-automatic or automatic. With regard to the road-area extraction, it depends on image segmentation (Unsalan and Sirmacek 2012, Cheng *et al.* 2014, Li *et al.* 2014), while road-centreline extraction methods concentrate on detecting road-skeletons (Cao and Sun 2014, Shi *et al.* 2014, Sironi *et al.* 2014, Cheng *et al.* 2016). On the other and, semi-automatic approaches require some additional geographical information (Chaudhuri *et al.* 2012, Liu *et al.* 2015), while automatic methods do not require this kind of information (Mnih and Hinton 2010, Maurya *et al.* 2011, Wegner *et al.* 2013).

In this paper, we develop an extraction method that could be classified as an automatic road-area extraction method (according to the classification proposed by Alshehhi and Reddy (2017)). The objective of our approach is to improve the performance of VIC mechanisms by means of a non-parametric approach to the textural segmentation of images. As a consequence of this, our algorithm is able to achieve significant improvements in the alignment between images and parcel boundary lines from vector GDB on the basis of the establishment of robust correspondences between road intersections extracted from both data sets. As described below, this is possible thanks to the inclusion of roads (by means of their axes) in the model of cadastre studied. In addition, and although our method can be considered as a road-area extraction method, it overcomes some disadvantages which characterise these kinds of methods. These disadvantages are as follows:

- Many of these methods fail since they utilise techniques based on pixel-based classification. Therefore, they cannot discriminate between contextual structures (Alshehhi and Reddy 2017).
- Many of these methods analyse the shape of the grey-scale histogram and classify the histogram clusters based on their sizes splitting the original image into a set of regions which are visually distinct and uniform with respect to certain statistical properties (Ojala *et al.* 1996, Ojala and Pietikäinen 1999). However, in the case of real imagery regions do not display statistical uniformity (due to the presence of noise or extensive shady zones).

The effectiveness of our algorithm has been tested and proved through several studies carried out on different Spanish cartographic products (PNOA orthophotos at a resolution of 0.5 m, Quick Bird (QB) Satellite Imagery at a resolution of 0.6 m, and IECA orthophotos at a resolution of 1 m), (see Ruiz *et al.* 2007, 2011b). Thus, and although a detailed description of these products is not provided here, some of the results achieved are summarised in the section 'Results' in order to highlight the effectiveness of our approach.

This paper is based on an earlier one, Ruiz *et al.* (2011b), and includes three important novelties: (i) the application of the methodology developed in it to cadastral data in Poland, (ii) the resolution of the problem of not well-defined intersection points, and (iii) the positional uncertainty assessment of certain cadastral vector data linked to imagery both before and after applying our alignment procedure between them. It is organised as follows: the next section describes the main characteristics of the cadastral system in Poland and discrepancies, henceforth referred to as *inconsistencies*, with regard to the lack of an alignment between a vector GDB and imagery detected in it. In the section 'Methodological improvement approach', and as a reminder, we briefly describe our methodological approach to improving this alignment and address in greater detail the last two novelties listed above. The 'Results' section shows the experimental results. Finally, in the last section 'Conclusions' are presented.

Cadastral system in Poland and detection of inconsistencies

The change of the social and economic order in Eastern Europe has caused important changes in the perception of real property ownership and real property transactions (Cetl *et al.* 2012). These changes have particularly affected the everyday lives of citizens and their rights, and Poland has been no exception to this situation. Thus, Polish official registers have been subjected to a permanent modernisation process that began in 1989. Since then, three basic models of register have been developed here: a Real Property Register, a land and buildings cadastre, and the spatial registration of utility infrastructure.

- The *Real Property Register* is a public register carried out by the relevant departments of district courts. Its basic objective is to establish who has the rights to property, ensuring safe trade in real estate. From a technical point of view, this register is digitalised so we can access it through a web site. However, it has a clear inconvenience: the data stored in it lack spatial information.
- The *Land and Buildings Cadastre* is an official register which basically contains detailed information on land and buildings and their owners (or entities who dispose of this property). It is founded and maintained by governmental agencies for the whole country and managed by local authorities at the county level. Although the most common cadastral objects are polygonal features such as buildings and land parcels (for which location, boundaries, area, land use, and soil classification are provided), the cadastre model also includes information regarding the spatial location of other objects such as linear features. Thus, in the case of Poznań (our study case) roads are features explicitly defined as cadastral objects (by means of their axes), and not as a spatial extent of the boundaries of parcels. This is key aspect to our approach. From a technical perspective, the software tools used for maintaining the Cadastre are based on relational GDB (MS, SQL, and ORACLE) (Siejka *et al.* 2014).
- The *Spatial Registration of Infrastructure* contains a set of technical data regarding the spatial location of land infrastructure. As in the previous case, the software systems are based on a georelational data model.

The updating and maintenance of these models, and therefore their efficiency, depends on different

organisations. Thus, governmental administration is responsible for maintaining the consistency of the Spatial Registration of Infrastructure GDB and the Cadastral GDB, while the courts are responsible for the Real Property Register. However, in the case of Poznań, the Cadastre also incorporates information regarding the Real Property Register and the Spatial Registration of Infrastructure. So it includes data concerning the rights to property as well as the dates of changes to those rights by registering several transactions related to the same ownership in a defined period of time. In relation to certain objects (such as land plots, buildings, and utility infrastructure), the cadastre gathers the dates of all changes in terms of spatial management. As a result, the cadastral management model developed for Poznań is highly efficient and competitive because it allows the building of a spatial cadastre of municipal networks with sufficient precision, and which can be used for the purposes of managing the rights to use certain facilities, as well as for design, taxation, and other purposes (Siejka *et al.* 2014).

However, despite all this the responsibilities resulting from the obligation to update the GDB are not fully respected. This causes the appearance of some positional inconsistencies related to the lack of a geometric coherence between images and their associated cadastral data. As mentioned in the previous section, this lack of geometric coherence results in an alignment problem between them. Figure 3 shows three examples of this positional inconsistency. Although just those three examples have been employed for checking graphically our approach (because of space limitations), in the case of parcel boundary lines (Fig. 3*c*) the graphic results have been supplemented by an analytical survey of their positional uncertainty both before and after applying our alignment procedure.

(a)

(b)

(c)

3 Representative examples of geometric inconsistencies detected. *a* Lack of correspondence between road intersections (Strzeszyn neighbourhood), *b* lack of correspondence between road intersections (Ogrody neighbourhood), and c lack of alignment between parcel boundary lines and buildings (Przezmierowo neighbourhood)
Source: http://poznan.podgik.pl/.

Methodological improvement approach

As mentioned in the'Introduction' section, the main objective of our methodological improvement approach is to achieve significant improvements in the alignment between images and parcel boundary lines from vector GDB on the basis of the establishment of robust correspondences between road intersections extracted from both data sets. For this, the inconsistencies detected (Fig. 3*a* and *b*) must be removed and the alignment between parcel boundary lines, roadsides and buildings (Fig. 3*c*) must be improved. This paper is based on our earlier paper, Ruiz *et al.* (2011b). Hence only a brief description of the algorithm is given here. However, we present two new methodological contributions to the issue: (i) the resolution of the not-well-defined points of intersection and (ii) the positional uncertainty assessment of the parcel boundary lines.

Automatic alignment algorithm

There are nine major steps in our automatic alignment algorithm which can be summarised as follows (see Ruiz *et al.* (2011b) for further information):

(1) In a preliminary step, we used the LTM (Chen *et al.* 2006) to work on reduced areas of the original image. These areas, in turn, were decomposed into a grid, because texture cannot be associated with a single pixel.
(2) *Texture characterisation*. The texture is locally sampled by the joint distribution of two properties associated with pixels: (i) a local binary pattern (LBP) and (ii) a contrast measure (C). To extract the road regions from the image we had to train the system on a small area of the image to learn the road texture parameters LBP, C. Their computation is addressed in detail in Ruiz *et al.* (2011b). The LBP/C distribution for each cell of the grid was approximated by a discrete two-dimensional histogram of $256 \times b$ pixels, where b is the number of bins for C. In order to compare LBP/C

distributions between two histograms (A, B), a log-likelihood-ratio G statistic was used. In Equation (1), N is the number of bins and f_i is the frequency at bin i.

$$G = 2\left(\left[\sum_{A,B}\sum_{i=1}^{N} f_i \log f_i\right] - \left[\sum_{A,B}\left(\sum_{i=1}^{N} f_i\right)\log\left(\sum_{i=1}^{N} f_i\right)\right] - \left[\sum_{i=1}^{N}\left(\sum_{A,B} f_i\right)\log\left(\sum_{A,B} f_i\right)\right] + \left[\left(\sum_{A,B}\sum_{i=1}^{N} f_i\right)\log\left(\sum_{A,B}\sum_{i=1}^{N} f_i\right)\right]\right) \quad (1)$$

(3) *Hierarchical structure generation.* A pyramidal representation was adapted to texture segmentation (Fig. 4). The LBP/C distribution of the grid cells formed the base of this pyramidal structure (Fig. 4*a*). Each level l of the pyramid was a reduced map with one-fourth of the cells of the level immediately below (Fig. 4*b*).

Each pyramid cell, denoted by (x, y, l), had the following parameters associated with it:

- *Homogeneity.* $H(x, y, l)$ ranged from 1 (if the four cells immediately underneath had the same texture) to 0. The setting of H was based on a uniformity test. Thus, the four cells had the same texture if a measure of relative dissimilarity within that region was lower than a certain threshold U, ($G\max/G\min < U$). U must be set in such a way as to ensure the detection and differentiation of textures. For this reason, it is advisable to choose a small value close to one for this threshold.
- *Texture.* If the cell was homogeneous, $T(x, y, l)$ was equal to the sum of the LBP/C distributions of the four cells immediately underneath. Otherwise, it was set to a fixed value (T_{NH}).
- *Parent link.* $(X, Y)_{(x, y, l)}$. If $H(x, y, l)$ was equal to 1, the values of the parent links of the four cells immediately underneath were set to (x, y). Otherwise, these four parent links were set to null.
- *Centroid.* $C(x, y, l)$. The centre of mass of the base region associated with (x, y, l).
- *Histogram.* Each parent link stored a two-dimensional histogram which characterised the texture of the image region represented by this node.

When this step finished, the cells of the hierarchical structure which presented a homogeneity value equal to 1 and had no parent were included in the set. These cells were linked to homogeneous regions at the base, defining initial image segmentation.

(1) *Growth of homogeneous cells.* The algorithm linked cells whose parent link values were null. Thus, a cell (x, y, l) was linked to the parent of neighbours $(xp, yp, l+1)$ when two cells had the same texture.

(2) *Fusion of homogeneous cells.* The neighbouring cells, $(x1, y1, l)$ and $(x2, y2, l)$, were merged if the following four conditions were true:

- (X, Y) $(x1, y1, l)$ = null. Therefore, the cell had no parent.
- (X, Y) $(x2, y2, l)$ = null. Therefore, the cell had no parent.
- The cells had a homogeneous texture. $H(x1, y1, l) = 1$ & $H(x2, y2, l) = 1$.
- The cells had the same texture.

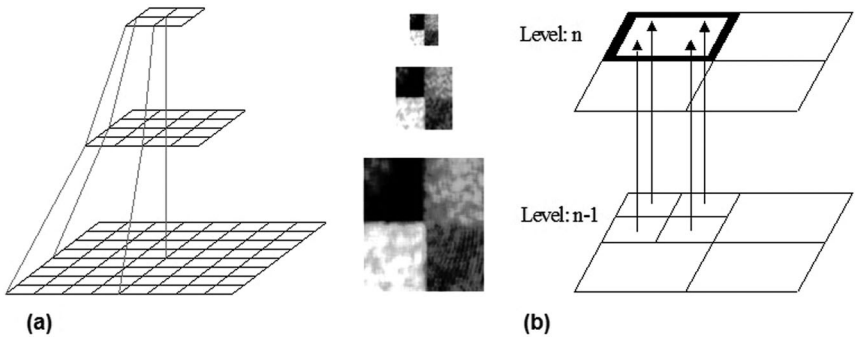
(3) *Pixel-wise.* In order to soften the segmented image, in a post-processing step the resolution of all blocks in the texture region boundaries was increased until those boundaries were one pixel wide.

(4) *Road extraction.* To identify and separate road zones from the rest, we trained the system to learn road texture parameters (LBP, C). Thus, a set of thresholds for these texture parameters was generated.

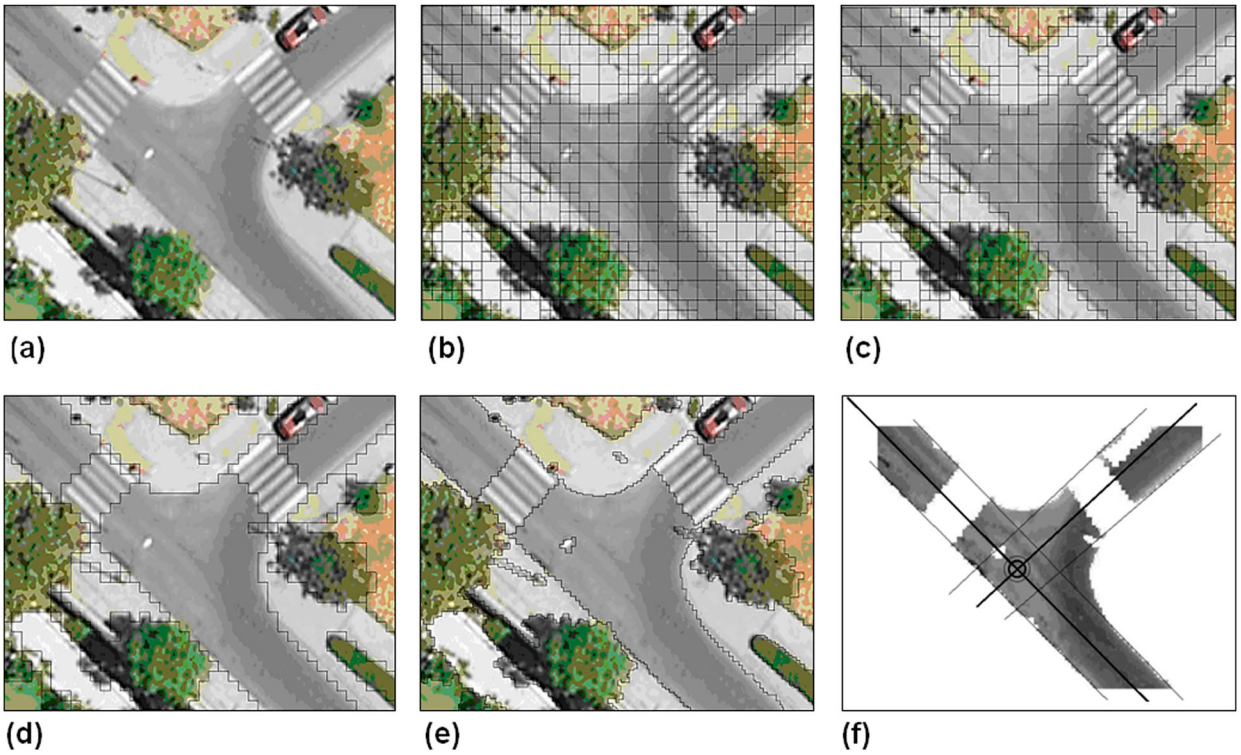
(5) *Determination of road intersection points.* First, we computed two outside lines of each road by means of an interpolation method so that the centre axis could be derived from them. Finally, the road intersection point was calculated by means of the intersection of both centre axes.

(6) *Alignment of imagery and vector GDB.* After obtaining the intersection points in images and their homologous points in vector GDB, we applied a set of geometric adjustment operations to align both products (Saalfeld 1985, Doytsher 2000).

For a better understanding, Fig. 5 shows the main segmentation stages of the proposed algorithm. Figure 5*a* shows the original image obtained from the template inferred from the vector information. Results of the initial hierarchical segmentation are presented in Fig. 5*b* and *c* shows the resulting regions after the growth of homogeneous cells. Figure 5*d* shows the final regions after the fusion procedure. Results of the smoothing phase are



4 Hierarchical splitting used to divide the image into regions of uniform texture. *a* Set of pyramidal structures from texture and *b* Parent link
Source: Ruiz *et al*. (2011b).



5 Sequence of the proposed segmentation algorithm with regard to its main segmentation stages: *a* original image after applying LTM, *b* base regions generated after the splitting step, *c* base regions generated after homogeneous cells growing, *d* base regions generated after homogeneous cells fusion, *e* boundaries after smoothing phase, and *f* road regions extracted Source: Ruiz *et al.* (2011b).

presented in Fig. 5*e* and finally, Fig. 5*f* shows the road regions extracted from the image, the outside lines of each road and its centre axis, and the road intersection point.

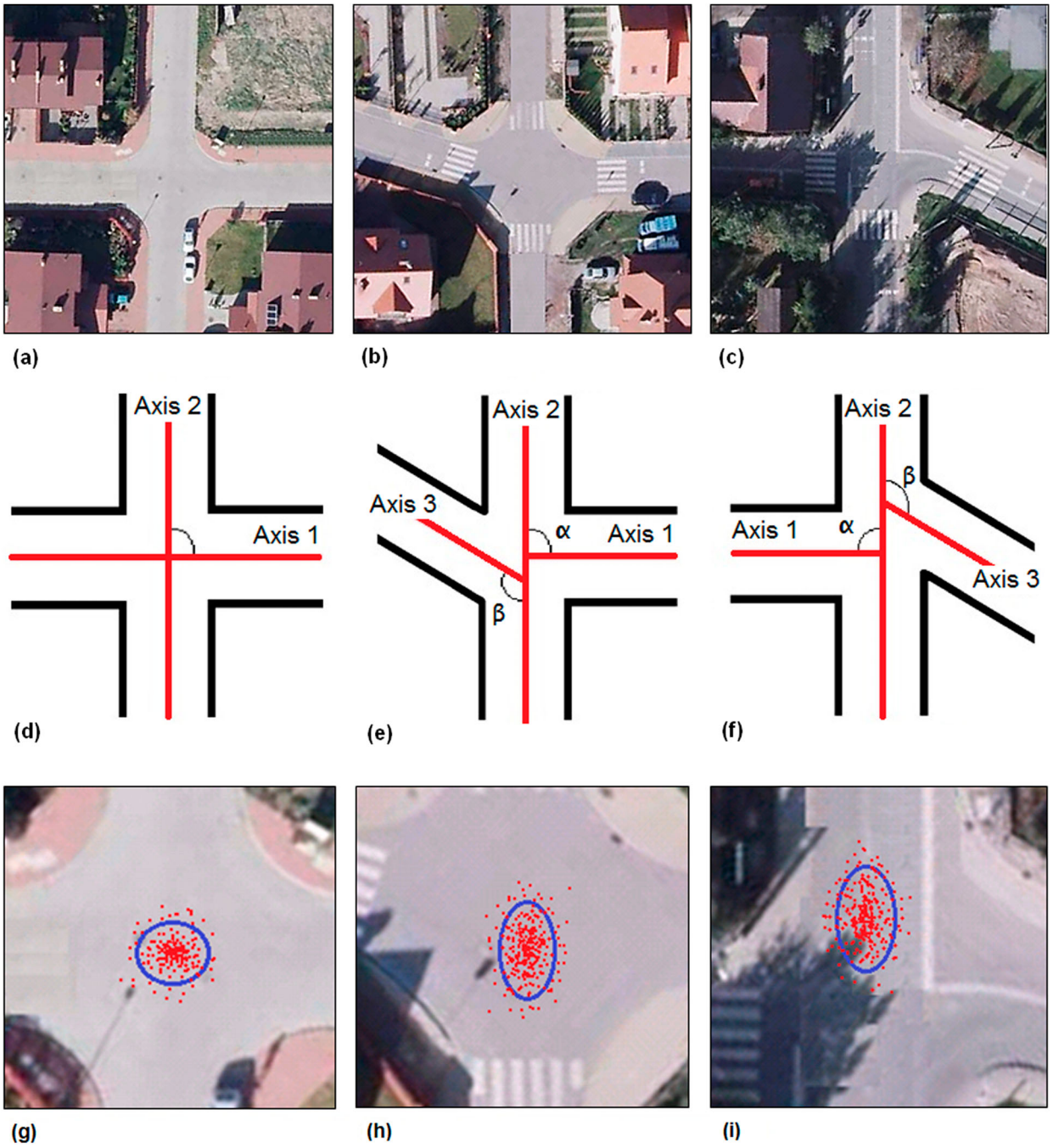
Resolution of the not-well-defined points of intersection

In our first version of the algorithm (Ruiz *et al.* 2011b) not all the intersection points were useful because some of them were not well defined. Therefore, we had to filter out candidates in order to identify and remove this type of points. In the afore mentioned study, two options were considered: (i) the intersection of two axes, and (ii) the intersection of three axes. Thus, in the case of only two axes there is a unique intersection point determination, calculated as the intersection of both axes (Fig. 6*a* and *d*). However, in the examples shown in Fig. 6*b–f* (the intersection of three axes) there is more than one candidate for the intersection point, for example, the intersection between axis 1 and 2, or that between axis 2 and 3. This case was initially discarded because the extraction algorithm had no information about the quality of each intersection point.

However, we have overcome this problem in our updated version of the algorithm, adding a significant new value to it. Thus, in the case of the intersection of three axes (Fig. 6*e* and *f*), our strategy was to compute the angles formed between the axes on the basis of metadata (orientation) provided by the vector GDB. Specifically, two angles were computed: α as a result of the intersection between axes 1 and 2, and β as a result of the intersection between axes 2 and 3. After computing these angles, we selected the intersection with the angle closest to 90°.The reason for this choice is the fact that, according to the *Guide to the Expression of Uncertainty in Measurement* (JCGM 2008), the positional uncertainty of a point derived from the intersection between two lines increases when the angle of intersection between them differs from this value. In our case, and taking into account the law of the propagation of uncertainty, the main source of positional uncertainty for determining an intersection point is the uncertainty with which we computed (by means of an interpolation method) the two outside lines of each road because it is propagated to the derived centre axes, and therefore to their intersection.

In order to test these assumptions, we used a statistical simulation process to verify whether the spatial distribution of intersection points followed the pattern we expected, analysing the variability of the estimated position. Simulation can be defined as the construction of a mathematical model to reproduce the characteristics of a phenomenon, system, or process, in order to obtain information or solve problems (Ríos *et al.* 1997). Specifically, for this process we applied the Monte Carlo method (Robert and Casella 2004). So we replicated the interpolation method by which the outside lines are generated under controlled circumstances by means of synthetic normal populations of positional errors with parameters ($\mu_p = 0$, $\sigma_p^2 = 1$ and, following the discussion of Heuvelink (1998) on the number of runs (NR), with NR = 200.

Taking into account our aim, the simulation results achieved for the selected images are shown from a graphic



6 Examples of road intersection points. *a*, *d* Well-defined intersection point, *b*, *c*, *e*, *f* multiple intersection cases *g*, *h*, *i*) graphic results of the simulation process

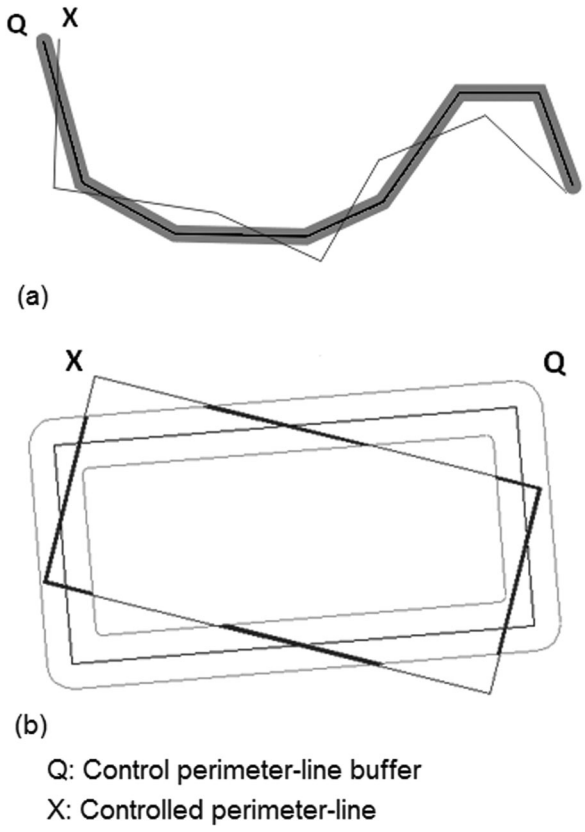
perspective in Fig. 6*g–i*. They show variations of behaviour in the three cases analysed, the first being the one with a lower variability (with the intersection between the axes at an angle closest to 90°). On the other hand, some analytical results regarding this variability are shown in the next section.

Finally, we must note that this solution contributes substantially to the improvement of the algorithm, and also therefore of our alignment methodology, in two key aspects

- First, a substantial increase in the number of intersection points which will be employed for the final adjustment to align both products.
- Second, the reduction of the computation time due to (i) a decrease in the segmentation area in the image; thus, the area from the axis discarded was not employed for the computation; and (ii) the deletion of the filtering procedure used for the detection of not-well-defined points.

Positional uncertainty assessment of the parcel boundary lines

In order to supplement the graphic results with regard to the improvement in the alignment between images and parcel boundary lines after applying the new version of our algorithm, we have used a technique which allows



7 ***a*** **The SBOM, and (b) adaptation of the SBOM to our framework (line-closed case)**
Source: Ruiz-Lendinez *et al.* (2013).

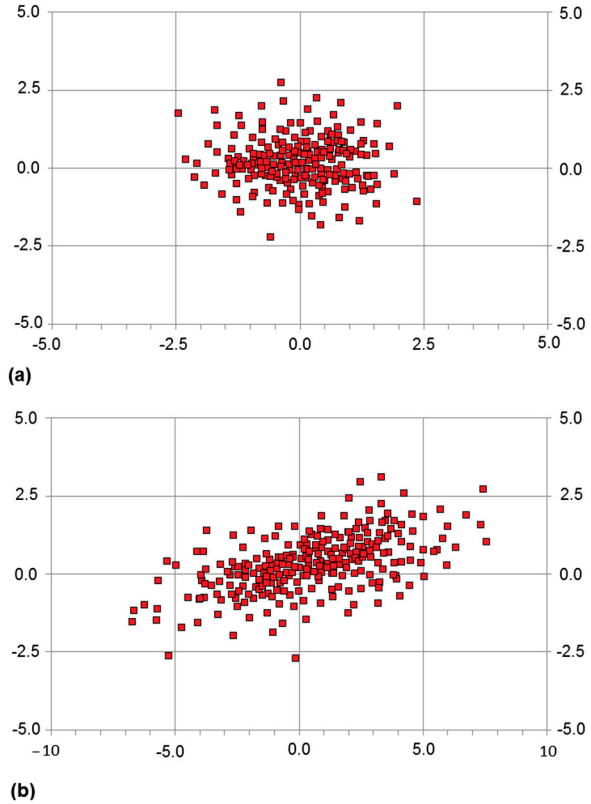
us to analyse the displacement between two polygonal features and, therefore, to quantify this improvement in numerical terms. Specifically, we have used the simple buffer overlay method (SBOM) (Goodchild and Hunter 1997). There are many examples of its application in the field of positional assessment of GDB (Ariza-López *et al.* 2008, Mozas-Calvache and Ariza-López 2010, Ruiz-Lendinez *et al.* 2013). Based on buffer generation on the line of the source of greater accuracy (Q), this method determines the percentage of the controlled line (X) which is within this buffer (Fig. 7*a*). By increasing the width of the buffer, we obtain a probability distribution of inclusion of the controlled line inside the buffer of the source of greater accuracy (Ruiz-Lendinez *et al.* 2013). In general, it can be said that the displacement between two polygonal features can be defined through measures of the positional discrepancies between the apparent location of a spatial entity (X) recorded in a GDB, and its true location (real world) (Q).

For this purpose, and together with the two data sets to be assessed (parcel boundary lines from the initial vector GDB and parcel boundary lines from the vector GDB resulting from the alignment procedure), we have selected a control sample of 40 parcels whose geometry was defined by means of GPS data acquisition (true location). In addition, and in order to increase as far as possible the robustness of the results, these parcels must be located near to points employed in the imagery georeferencing process. In our case, the set of metrics applied by the proposed method, described above, must be adapted to the line-closed case (Fig. 7*b*). In addition, in order to implement the SBOM we have employed all the parcels present in the control sample, obtaining an aggregated distribution curve. This allows us to obtain the uncertainty originated by the inclusion of the parcels boundary lines from the two vector GDB into the parcels control (from the GPS acquisition) perimeter-line buffer (SBOM) and therefore to compute the positional uncertainty both before and after applying our alignment procedure.

Results

The main parameters used to assess the efficiency in software applications such as that developed here are: (i) the computation time and (ii) the accuracy achieved by the operator which allows making the key decisions. Hence although the overall performance of our algorithm depends on the accuracy for each component, in our case the dominant component is the LBP operator. In addition, its runtime is the one that presents a highest degree of dependence with regard to (i) the imagery properties (size and resolution) and (ii) the execution platform properties. These are key aspects to our approach.

In view of the above, in the following two subsections we have focused on discuss the computation time and analyse the accuracy achieved by the LBP operator during the resolution procedure of graphic inconsistences. Finally, in the section 'Methodological improvement approach' some results with regard to the positional uncertainty of the parcel boundary lines both before and after applying our approach are shown.



8 Analytical results from simulation process with regard to the positional variability of intersection points (multiple intersection case): *a* road intersection angle of 91° and *b* road intersection angle of 117°

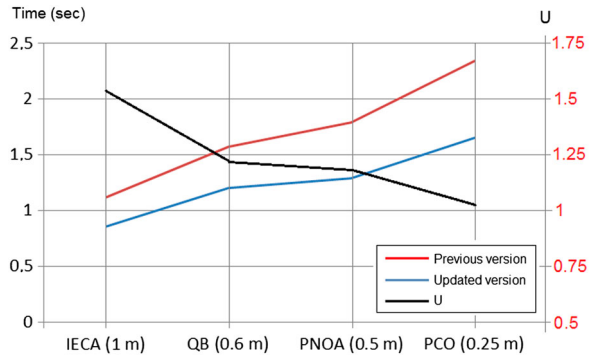
Results of simulation analysis and their effects on the computation time

In order to supplement the graphic results obtained by the Monte Carlo method in the cases studied (multiple intersection cases shown in Fig. 6), we report the variability achieved from an analytical point of view. Thus, Fig. 8 shows the variability results achieved for the intersection angles α (the intersection of axes 1 and 2) and β (the intersection of axes 2 and 3) related to the case shown in Fig. 6*b,e* and *h*. Specifically, α takes the value of 91° and β takes the value of 117°. As can be seen, when the intersection angle is closest to 90° (Fig. 8*a*) the variability is lower (approximately 4 m) than when the angle differs from this value (Fig. 8*b*). In this last case, the variability is about 8 m.

In addition, as mentioned above, the resolution of the not-well-defined points of intersection caused a substantial increase in the number of points employed for the final adjustment to align both products. Specifically, and for our working area, the number of intersection points computed increased by 47% (employing the new version of the algorithm) with regard to the results achieved for the same working area when we employed the previous version developed in Ruiz *et al.* (2011b).

With regard to the computation process, both the decrease in the segmentation area in the image (and therefore in its number of textural regions) and the deletion of the filtering procedure used to detect not-well-defined points considerably reduced the time needed to complete the alignment procedure. Our test platform was an Intel Core i5 dual-core 2.7 GHz with 8 GB memory. In this case, the average computation time (ACT) needed to complete the process (for the same working area) decreased by around 24% with respect to the time taken to complete the alignment procedure when the previous version of the algorithm was employed. Obviously, the computation time also decreased significantly in studies carried out on different Spanish cartographic products, not only due to the improvement of computational power, but also the improvement of our algorithm, despite using the same value for U. In this sense, we must note that the computational efficiency of our algorithm depends on (i) the threshold U (see the section 'Automatic alignment algorithm') and (ii) the number of textural regions in the image. Thus, the computation time increases when (i) the threshold 'U' decreases and (ii) images contain a high number of textural regions. The ACT results achieved from the application of both algorithm versions to the tested image sources are shown in Table 1.

As can be seen, to achieve the alignment procedure of higher resolution images the algorithm required more computation time. This is because, as mentioned above, the computation time increases when the threshold 'U' decreases, and to achieve the segmentation of high resolution images a small value for threshold 'U' is required. Figure 9 shows the relationship between computation time, image resolution, and threshold 'U'.



9 Computation time vs. image resolution and U

Resolution of graphic inconsistencies

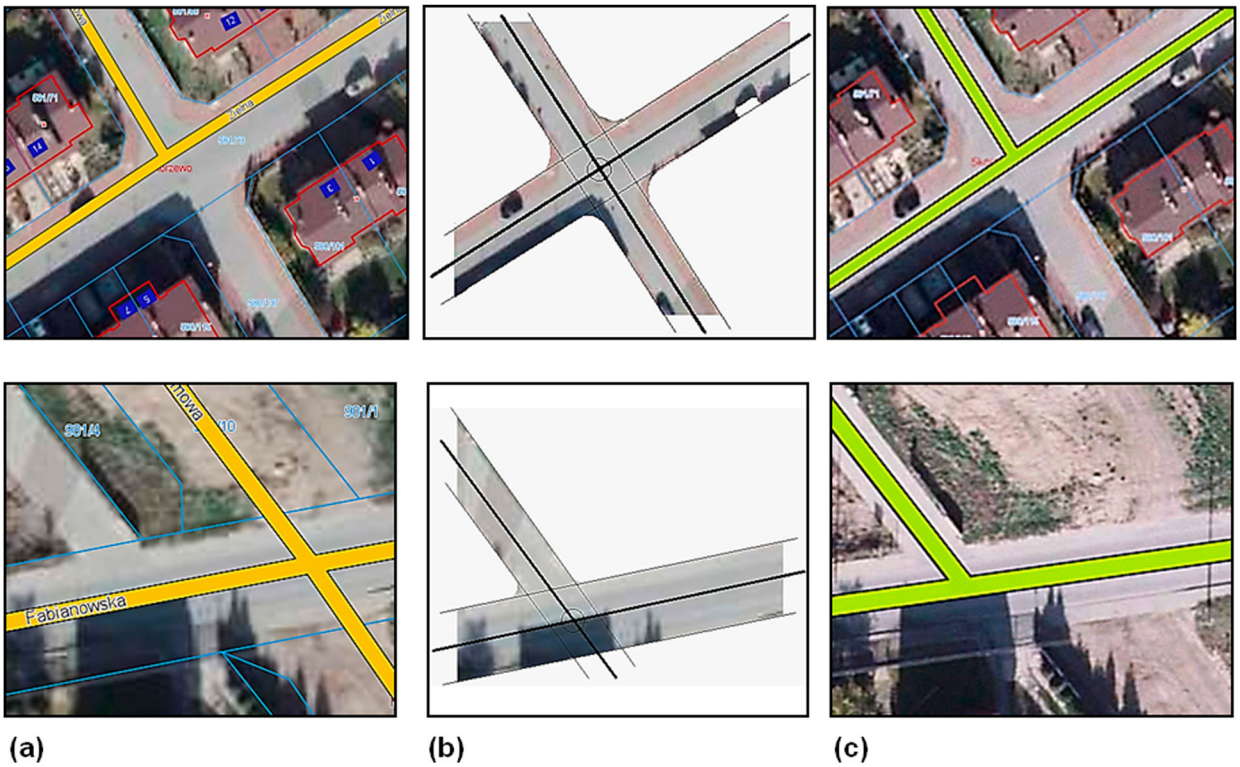
The test carried out on different Spanish cartographic products (PNOA orthophotos at a resolution of 0.5 m, QB Satellite Imagery at a resolution of 0.6 m and IECA orthophotos at a resolution of 1 m) proved that, although the textures of real images were less uniform than the homogeneous textures of artificial images and the spatial pattern changed smoothly from one texture to another, in the specific case of roads adjacent textured regions were generally separated by well-defined boundaries and for this reason we were able to obtain good results for the segmentation process.

In the case of cadastral information from Poland, we obtained similar results with regard to graphic segmentation (Fig. 10*b*). However, the accuracies of the LBP operator were significantly better than those achieved with the Spanish products due to their greater quality and resolution. Following the official specifications provided by governmental agencies, orthophotos are usually developed in the following pixel sizes: 1.0, 0.5, 0.25, and 0.1 m (geoportal.gov.pl 2016). Specifically, we employed images at a 0.25-m resolution. Table 2 shows the average accuracy of the LBP operator in our case study (Poznań) with respect to the values obtained for the Spanish image sources employed in our previous study (Ruiz *et al.* 2011b).

Finally, Fig. 10*c* shows the results (from a graphic point of view) with regard to the matching between road intersections extracted from both data sets. As

Table 1 ACT with respect to various sources

Image source	Image resolution (m)	Number of working areas	ACT (previous version) (s)	ACT (updated version) (s)
QB satellite imagery	0.6	6	1.572	1.201
IECA orthophoto	1	8	1.117	0.858
PNOA orthophoto	0.5	10	1.693	1.289
Polish cadastral Orthophoto (PCO)	0.25	10	2.445	1.798



10 Graphic results after applying our algorithm to the detected inconsistencies: *a* initial stage *b* image segmentation and computation of intersection points, and *c* final alignment between road intersections

Table 2 Average accuracy of LBP operator with respect to various sources

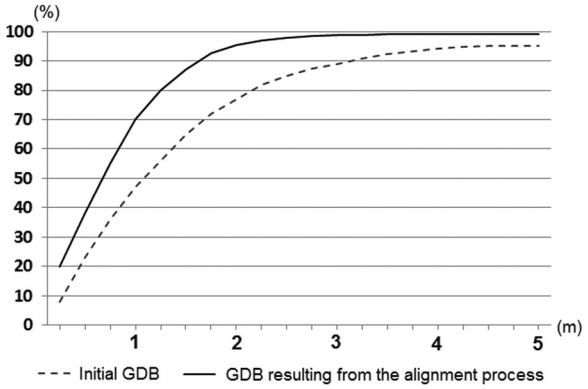
Image source	Image resolution (m)	Number of test areas	Average accuracy of LBP operator (%)
QB satellite imagery	0.6	6	93.7
IECA orthophoto	1	8	84.3
PNOA orthophoto	0.5	10	95.3
Polish cadastral orthophoto	0.25	10	97.5

can be seen, our algorithm has been able to establish robust correspondences for the detected inconsistencies (Fig. 10*a*). In addition, significant improvements in the alignment between images and the rest of the features present in the vector GDB (parcel boundary lines, roadsides, and some buildings) have been achieved (Fig. 11).

Positional uncertainty assessment of the parcel boundary lines

Figure 12 presents the aggregated distribution functions resulting by applying the SBOM (Fig. 7*b*) on both data sets (parcel boundary lines from the initial vector GDB and parcel boundary lines from the vector GDB resulting

11 Graphic results after applying our algorithm to the detected inconsistencies: *a* initial stage and *b* final alignment between roadsides, parcel boundary lines, and some buildings



12 SBOM results for the initial GDB and for the GDB resulting from the alignment process

from the alignment procedure) using buffers with widths from 0.5 to 5 m. The aggregated curves obtained with this method show a distribution function of the uncertainty of each GDB for several levels of confidence. The figure indicates a value around 1.8 m for a 95% level of confidence in the case of the GDB resulting from the alignment process and a value around 4.6 m for a 95% level of confidence in the case of the initial GDB. These values imply that (i) the resulting GDB has less uncertainty than the initial GDB and (ii) the graphic results with regard to the improvement in the alignment procedure are therefore confirmed.

Conclusions

This paper presents the results of the first stage of an continuing research project in which the University of Jaén (Spain) and the Institute of Socio-Economic Geography and Spatial Management at Adam Mickiewicz University of Poznań (Poland) participate, and whose main objective is to assess the VIC procedures affecting cadastral data in Poland.

To that end, we employed an automatic alignment algorithm based on the use of texture as the main radiometric parameter in the segmentation of the original image. This algorithm, initially developed in Ruiz *et al.* (2011b), was significantly improved with a partial resolution of the problem of not-well-defined intersection points (three-axes intersections), thus increasing the number of intersection points computed by 47%. In addition, the updated version achieved a reduction in computation time of up to 25%.

On the other hand, the algorithm provided excellent results in both texture classification and segmentation problems on our work images, achieving 97.5% average accuracy of the LBP operator. This value is in agreement with the results achieved for Spanish cartographic products with which we worked previously, particularly given the fact that the resolution of the images used was higher than those mentioned above.

Finally, and with regard to the improvement of the alignment between vector data and imagery, the values resulting from the application of SBOM have allowed us to quantify the accuracy of these vector data both before and after applying our alignment procedure. These values prove that the resulting GDB has less uncertainty than the initial one confirming the graphic results previously obtained. Therefore, we can conclude that there are reasonable grounds for extrapolating our approach to cadastral data in Poland, which would mark a considerable improvement in their positional quality.

However, this study represents only a further step towards the improvement of the alignment between vector GDB and imagery. Thus, although the updated version of our algorithm has solved the problem of not-well-defined intersection points (three-axes intersections), this solution remains partial because there are still cases (multiple intersections with more than three axes) which have to be clarified. This challenge must be addressed in future works.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work has been partially funded by the Ministry of Education, Culture and sports of Spain under Grant No. CAS16/00034 (Mobility stays abroad 'José Castillejo').

References

Alshehhi, R. and Reddy, M., 2017. Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. *ISPRS journal of photogrammetry and remote sensing*, 126, 245–260.

Ariza-López, F., Atkinson-Gordo, A., and Rodríguez-Avi, J., 2008. Acceptance curves for the positional control of geographic data bases. *Journal of surveying engineering*, 134 (1), 26–32.

Brovelli, M. and Zambroni, G., 2004. Adaptive transformations of cartographic bases by means of multiresolution spline interpolation. *In: M.O. Altan, ed. Proceedings of XXth ISPRS congress*, 12–23 July 2004 Istanbul, Turkey.

Cao, C. and Sun, Y., 2014. Automatic road centerline extraction from imagery using road gps data. *Remote sensing*, 6 (9), 9014–9033.

Casado, M., 2006. Some basic mathematical constraints for the geometric conflation problem. *In:* M. Caetano and M. Painho, eds. *7th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, 5–7 July 2006. Lisbon, Portugal.

Cetl, V., Roic, M., and Mastelic, S., 2012. Towards a real property cadastre in Croatia. *Survey review*, 44 (324), 17–22.

Chaudhuri, D., Kushwaha, N., and Samal, A., 2012. Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques. *IEEE journal of selected topic in applied earth observations and remote sensing*, 5 (5), 1538–1544.

Chen, C., Knoblock, C., and Shahabi, C., 2006. Automatically conflating road vector data with orthoimagery. *Geoinformatica*, 10 (4), 495–530.

Cheng, G., *et al.*, 2014. Urban road extraction via graph cuts based probability propagation. *In: IEEE Computer Society, ed. IEEE international conference on image processing (ICIP)*, 27–30 October 2014. Paris, France: IEEE Computer Society, 5072–5076.

Cheng, G., *et al.*, 2016. Accurate urban road centerline extraction from VHR imagery via multiscale segmentation and tensor voting. *Neurocomputing*, 205, 407–420.

Chiang, Y., *et al.*, 2009. Automatic and accurate extraction of road intersections from raster maps. *Geoinformatica*, 13 (2), 121–157.

Doytsher, Y., 2000. A rubber sheeting algorithm for non-rectangular maps. *Computers & geosciences*, 26 (9/10), 1001–1010.

Geoportal.gov.pl. 2016. Available from: http://www.geoportal.gov.pl/dane/ortofotomapa [Accessed 1 June 2017].

Gillman, D., 1985. Triangulations for rubber sheeting. *In: the American Society of Photogrammetry and the American Congress on Surveying and Mapping, eds. Proceedings of the seventh international symposium on computer-assisted cartography*, 11–14 March 1985, Washington, USA.

Goodchild, M., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69 (4), 211–221.

Goodchild, M. and Hunter, G., 1997. A simple positional accuracy measure for linear features. *International journal of geographical information science*, 11 (3), 299–306.

Heuvelink, G., 1998. *Error propagation in environmental modelling*. London: Taylor and Francis.

Joint Committee for Guides in Metrology. 2008. JCGM 100:2008. Evaluation of measurement data - Guide to the Expression of Uncertainty in Measurement (GUM). Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG1). Available from: http://www.bipm.org/en/publications/guides/gum.html [Accessed 1 June 2017].

Li, J., *et al.*, 2014. Robust urban road image segmentation. *In: IEEE Robotics and Automation Society, ed. 11th world congress on intelligent control and automation (WCICA)*, 2923–2928.

Liu, B., *et al.*, 2015. Main road extraction from zy-3 grayscale imagery based on directional mathematical morphology and vgi prior knowledge in urban areas. *PLoS one*, 10 (9), 1–16.

Masó, J., Pons, X., and Zabala, A., 2012. Tuning the second-generation SDI: Theoretical aspects and real use cases. *International journal of geographical information science*, 26 (6), 983–1014.

Maurya, R., Gupta, P., and Shukla, A., 2011. Road extraction using k-means clustering and morphological operations. *In: IEEE Computer Society, ed. IEEE international conference on image information processing (ICIIP)*, 3–5 November 2011. Shimla, India: IEEE Computer Society, 1–6.

Mnih, V. and Hinton, G.E., 2010. Learning to detect roads in high-resolution aerial images. *In:* K. Daniilidis, P. Maragos, N. Paragios, eds. *11th European conference of computer vision-ECCV 2010*, 5–11 September 2010. Crete, Greece: Lecture Notes in Computer Science, 6316, 210–223.

Mozas-Calvache, A. and Ariza-López, F., 2010. Methodology for positional quality control in cartography using linear features. *The cartographic journal*, 47 (4), 371–378.

Ojala, T. and Pietikäinen, M., 1999. Unsupervised texture segmentation using feature distributions. *Pattern recognition*, 32 (1), 477–486.

Ojala, T., Pietikäinen, M., and Harwood, D., 1996. A comparative study of texture measures with classification based on feature distributions. *Pattern recognition*, 29 (1), 51–59.

Ríos, D., Ríos, S., and Martín, J., 1997. *Simulación, métodos y aplicaciones.* Madrid: Ra-Ma.

Robert, C. and Casella, G., 2004. *Monte carlo statistical methods.* 2nd ed. New York: Springer.

Ruiz, J.J., Rubio, T.J., and Lara, J.M., 2007. Extracción de puntos de control mediante caracterización textural en procesos de conflación vector-ortoimagen. *Mapping*, 122, 34–37.

Ruiz, J.J., *et al.*, 2011a. Digital map conflation: a review of the process and a proposal of classification. *International journal of geographical information science*, 25 (9), 1439–1466.

Ruiz, J.J., Rubio, T.J., and Ureña, M.A., 2011b. Automatic extraction of road intersections from images in conflation processes based on texture characterization. *Survey review*, 43 (321), 212–225.

Ruiz-Lendínez, J., Ariza-López, F., and Ureña-Cámara, M., 2013. Automatic positional accuracy assessment of geospatial databases using line-based methods. *Survey review*, 45 (332), 332–342.

Saalfeld, A., 1985. A fast rubber-sheeting transformation using simplical coordinates. *The American cartographer*, 12 (2), 169–173.

Shi, W., Miao, Z., and Debayle, J., 2014. An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE transaction on geoscience and remote sensing*, 52 (6), 3359–3372.

Shimizu, E. and Fuse, T., 2003. Rubber-sheeting of historical maps in GIS and its application to landscape visualization of old-time cities: focussing on Tokyo of the past. *In: K. Miyamoto, ed. Proceedings of the 8th international conference on computers in urban planning and urban management*, 27–29 May 2003, Sendai, Japan.

Siejka, M., Ślusarski, M., and Zygmunt, M., 2014. 3D+time Cadastre, possibility of implementation in Poland. *Survey review*, 46 (335), 79–89.

Sironi, A., Lepetit, V., and Fua, P., 2014. Multiscale centerline detection by learning a scale-space distance transform. *In: IEEE Computer Society, ed. IEEE conference on computer vision and pattern recognition*, 23–28 June 2014. Columbus, USA: IEEE Computer Society, 2697–2704.

Tobler, W., 1994. Bidimensional regression. *Geographical analysis*, 26 (3), 187–212.

Unsalan, C. and Sirmacek, B., 2012. Road network detection using probabilistic and graph theoretical methods. *IEEE transaction on geoscience and remote sensing*, 50 (11), 4441–4453.

Ward-Brown, J. and Churchill, V., 2004. *Variable Compleja y Aplicaciones.* Madrid: McGraw-Hill.

Watson, G., 2006. Computing Helmert transformations. *Journal of computational and applied mathematics*, 197 (2), 387–394.

Wegner, J.D., Montoya-Zegarra, J.A., and Schindler, K., 2013. A higher-order crf model for road network extraction. *In: IEEE Computer Society, ed. IEEE conference on computer vision and pattern recognition*, 23–28 June 2013. Portland, USA: IEEE Computer Society, 1698–1705.

Xavier, E., Ariza-López, F., and Ureña-Cámara, M., 2016. A survey of measures and methods for matching geospatial vector datasets. *ACM computing survey*, 49 (2), 1–34.