Ettore Sarzotti[1], Nicola Gabellieri[2], Paolo Zatelli[3], Angelo Besana[4]

From paper to GIS: an example of historical maps vectorization

Keywords: historical maps, GIS, segmentation, historical cadastre

Summary: Historical maps provide crucial insights into past landscapes, yet their traditional analogue format restricts accessibility and analysis. Digitization and vectorization of these maps can enhance their usability by integrating them with Geographic Information Systems (GIS). This paper presents the result of the ongoing study for the vectorization of a 19th-century Austro-Hungarian cadastral map of Trentino, Italy, using Object-Based Image Analysis (OBIA) and other image processing techniques. Initial OBIA methods faced challenges with segmentation sensitivity due to the colour variations and scan resolution. Adjustments in threshold parameters led to over-segmentation and under-segmentation issues. Other approaches like zero crossing and colour filtering were tried but none of the tested approach has yield satisfactory results. The paper highlights the need for tailored vectorization approaches and continued methodological advancements to effectively process historical maps for large-scale digitization projects. Despite ongoing challenges, this study underscores the potential of combining various techniques to improve the accuracy and efficiency of transforming historical maps into valuable geospatial datasets.

1. Introduction: from the map to the Historical GIS

Historical maps are invaluable resources for understanding the spatial and temporal dynamics of past landscapes and territories. They can offer insights into land use changes, urban development, and environmental transformations. However, the traditional use of these maps is often limited by their analogue format, which restricts accessibility, searchability, and integration with modern geospatial datasets. The digitization and vectorization of historical maps overcomes some of these limitations by transforming them into digital formats that can be easily analysed and integrated with Geographic Information Systems (GIS).

Vectorization, in particular, involves converting raster images of historical maps into vector data, where geographical features are represented by points, lines, and polygons. This process enhances the utility of historical maps by enabling precise geospatial analyses, facilitating overlay with contemporary datasets, and supporting dynamic visualization. The transformation of historical maps into vector formats is not only a technical endeavour but also a critical step in preserving and leveraging our cartographic heritage. Data collected by historical maps could be organized and managed in order to build a Historical GIS, namely a dataset with georeferenced historical information (Grava et al. 2020). The importance of Historical GIS has been highlighted both for their usefulness in research and as an informational basis to support analysis and provide information for territorial governance (Gregory and Ell 2007). For a long time, data collection from historical cartography in a GIS environment was carried out manually. Despite the results achieved, this process is considered time-consuming and resource-intensive in terms of human effort (Iosifescu et al. 2016). Not surprisingly, many studies based on manual vectorization of historical cartography are conducted at the local scale or on limited areas of space.

[1] Research fellow, University of Trento, ettore.sarzotti@unitn.it
[2] Professor, University of Trento, nicola.gabellieri@unitn.it
[3] Professor, University of Trento, paolo.zatelli@unitn.it
[4] Assistant Professor, University of Trento, angelo.besana@unitn.it

In recent years, advancements in machine learning and image processing have significantly improved the efficiency and accuracy of map vectorization. Automated and semi-automated vectorization techniques have emerged, allowing for large-scale digitization projects that were previously impractical. Despite these advancements, challenges remain, including the handling of diverse map styles, varying levels of preservation, and the integration of heterogeneous data sources. Literature has repeatedly emphasized the need to increase the number of case studies and sources used, in order to discuss challenges and potentialities and to refine methodologies (Pavelková et al. 2016; Besana et al. 2019; Gobbi et al. 2019; Asokan et al. 2020; Gede et al. 2020; Iras 2020; Zatelli et al. 2022; Schlegel 2023).

In response to these insights, this paper presents a method applied to the vectorization of a historical map, specifically a sheet from the Austro-Hungarian cadastre created in Trentino (Italy) in the 19th century. While many attempts to work on this source have been conducted in various European regions, the application of an automatic processing method is still ongoing (Timár and Biszak 2010; Pavelková et al. 2016; Dolejš and Forejt 2019; Bacior 2023).

In particular, the current work, aims to develop the vectorization of the map not only obtaining a vector layer with the land use and land cover (LULC) but also by identifying each individual cadastral parcel and converting it into a distinct polygon. This layer would be highly significant because it could be integrated with datasets on individual land uses, parcel values, and the property system, retrieved from the cadastral registers.

The paper presents the ongoing work, discussing problem and potentialities of applied methods using a single sheet (Dajano municipality, sheet number 7) as pilot project. Firstly, the Object-Based Image Analysis (OBIA) has been tested. Secondly, a two-stage approach based on black lines recognition and simpler pixelwise maximum likelihood classification is being developed. Although in this article, about this second approach, the focus is on the extraction of boundaries and the approaches taken to achieve it, since it poses the greatest difficulties; while the retrieval of the LULC classes by a pixelwise maximum likelihood classification is a simpler procedure and hence omitted. While the work is still ongoing, first result shows significant questions, challenges and perspective that are discussed in the following paragraphs.

2. Materials and sources

The Austrian Land Registry, also known as the Hapsburgs or *Franceschino* cadastre, is one of the most prominent fiscal documents produced in Europe during the 19th century. In Trentino (Italy), which was part of the Empire at the time, it was created between 1853 and 1862 (Dai Prà 2013).

The cadastral documents include, in addition to the registers, detailed top-down maps at the scale of 1:2.880 of individual cadastral parcels. These maps were drawn up using a projection system centred at the Pfarrturm in Innsbruck. The current Province of Trento territory corresponds to 13,297 map sheets each one covering an area of 288 ha, and measuring approximately 52.68 x 65.85 cm.

The maps can be downloaded in JPEG format from the OPENkat website (Provincia autonoma di Trento n.d.) and are released under the Creative Commons Attribution 4.0 license; although to have a better image quality, the original scans in TIF format were retrieved directly from the *Servizio libro fondiario e catasto* of the Autonomous Province of Trento. These files have the scanned resolution of 230 PPI, 24-bit colour depth, compared to the compressed JPEG file that have a resolution of 96 PPI, 24-bit colour depth. The digital maps are georeferenced in the ETRS89/UTM 32 N (EPSG 25832) reference system and have a ground resolution of around 0.32 m.

Due to their nature as fiscal documents and the scale used, these records provide detailed information on land use and cover, settlement systems, and property structure. Additionally, unlike other contemporary cadastres, these maps feature rich iconography, with colours and symbols representing different cultivations, making them easily readable without the textual registers (Buffoni et al. 2015; Gilardi 2013). Consequently, the *Franceschino* cadastre has been widely used for studies on landscape and territorial history in the Alpine context, focusing on agricultural (Sarabia et al. 2016; Gabellieri 2019), wooded (Tattoni et al. 2010; Dai Prà and Gabellieri 2021), and urban areas (Diamantini and Franceschini 2015). Although various methods for automatic vectorization of the sheets of this document have been successfully tested in the past (Gobbi et al. 2019; Zatelli et al. 2019), they have not yet been applied at the parcel scale.

3. Methods

The vectorization of the map to retrieve the cadastre parcels with LULC information is performed through the Object-based Image Analysis (OBIA). As described in Gobbi et al. (2019), this method combines image segmentation and machine-learning classification. The approach chosen for the segmentation is the region-growing algorithm which requires the settings of two parameters: a threshold value and a minimum size value. The threshold parameter varies between 0 and 1 and it defines the grade of sensitivity to radiometric changes between pixels: to low values corresponds a higher sensitivity and is more likely for different pixels to be placed in different segments. The minimum size parameter represents the lowest number of pixels that a single segment should have. Smaller segments are merged to the neighbouring segment with most similar radiometric values.
While the approach is effective for creating a LULC map from the cadastre map (Gobbi et al. 2019), it is very difficult to extract the lines corresponding to the parcel boundaries. This is due to the nature of the original map, which is hand-drawn with lines varying in thickness and colours and, most importantly, where the lines are discontinuous, presenting unintentional gaps which make it difficult to identify them as lines. Additionally, the digitalization of the maps with limited resolution creates pixels with mixed colours from the (black) lines and the background, making the classification more difficult.
Furthermore, it is difficult to separate handwritten lines and text, since they often have similar colour, size and shape features. For these reason two different approaches have been used for the LULC map creation and the boundaries extraction.

4. Results

4.1. Object-Based Image Analysis

Ideally, Object-Based Image Analysis (OBIA) should be able to identify all the different type of objects in a map by a two-steps process of segmentation, where pixels belonging to the same object are grouped in a segment, and classification, where object are assigned to a class. Therefore, it should be possible to identify both the different parcels and their LULC class and the boundary lines as distinct objects. In reality the choice of the segmentation parameters, threshold and minimum number of pixel segments, is always the result of a trade-off between the need to identify small and often non uniformly coloured objects and the necessity of avoiding over-segmentation, i.e. artificially dividing large objects into small segments, whose geometry has no relation to the geometry of the object and therefore can be misleading during the classification process.

Figure 1 shows the portion of one of the maps that was used for the tests. Main colours correspond to different land use and coverage. Brown corresponds to the woodland, bright green to the meadows and grasslands, and light green to the pastures. Other colours refer to buildings, watercourses, or arable land. The parcels may also have small symbols that denote secondary crops and complicate the work on the map. Figure 2 presents the result of the segmentation process with the threshold set to 0.2 and the minimum size set to 1,000.

Figure 1: portion of the sheet number 7 of the Dajano Municipality, part of the Hapsburgs cadastre, on which elaboration was made.

Figure 2: segmentation of the map in figure 1 with threshold equal to 0.2 and minimum size equal to 1000.

Figure 3 represents an enlargement in correspondence to the large woodland parcel in the centre of the original map. In figure 3b it can be noted how the threshold is too sensitive to the radiometric (colour) changes in the pixels of the same object and creates different segments for the same single parcel.

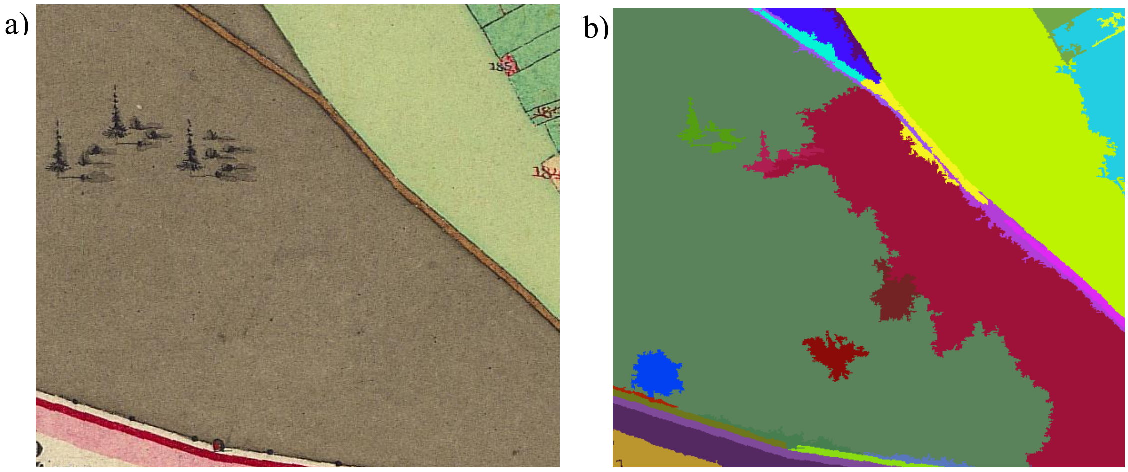

Figure 3: a) enlargement of figure 1 and b) relative segmentation.

In other cases, the threshold is not sensitive enough to detect radiometric changes in pixel and the result is that different parcels are merged with other adjacent parcels. This can be seen in the enlargement of figure 4 where parcels 1799, 1800 and 1805 are recognized as a single segment. The same can be said for the parcels 1779, 1759, 1762, 1768, 1769 and others.

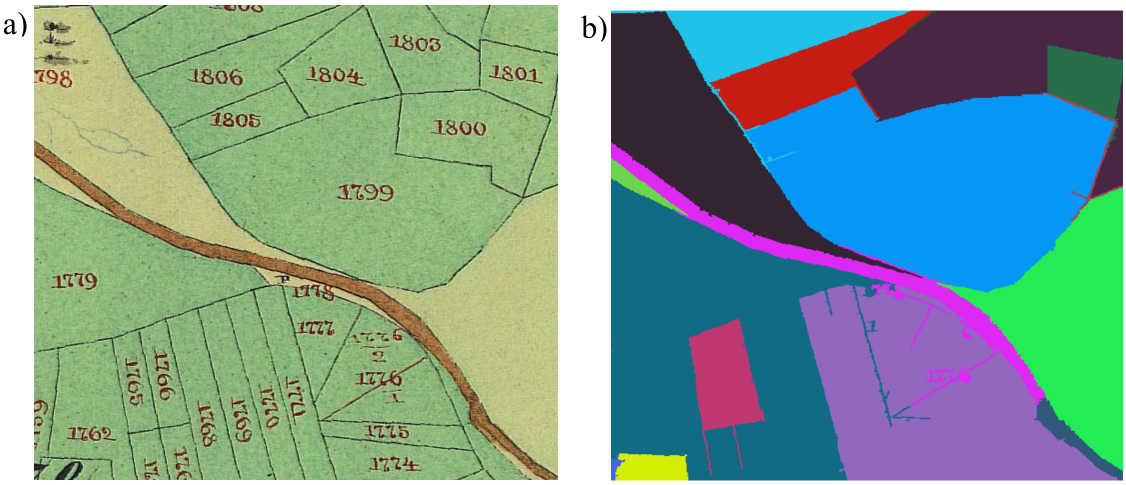

Figure 4: a) enlargement of figure 1 and b) relative segmentation.

4.2. Boundaries extraction

4.2.1 Zero crossing

The application of a "zero-crossing" filter (Krishnamoorthi and Bhattacharya 1998) assigns an image pixel to an edge by: i) applying the Fourier transformation to the image, ii) filtering the transformed image using a Fourier transform of the Laplacian of a two-dimensional Gaussian function,

iii) applying the inverse Fourier transform to the result, iv) searching the pixels corresponding to a sign change in the values, and v) selecting as edges the cells where a sign change occurs and the absolute value of the change is greater than a given threshold.
The application of this approach to the cadastre maps correctly identifies the colour discontinuities but each boundary line corresponds to two discontinuities: one between the parcel colour and the boundary line colour and a second one between the boundary line colour and the parcel colour on the other side; as seen in figures 5.

a) b)

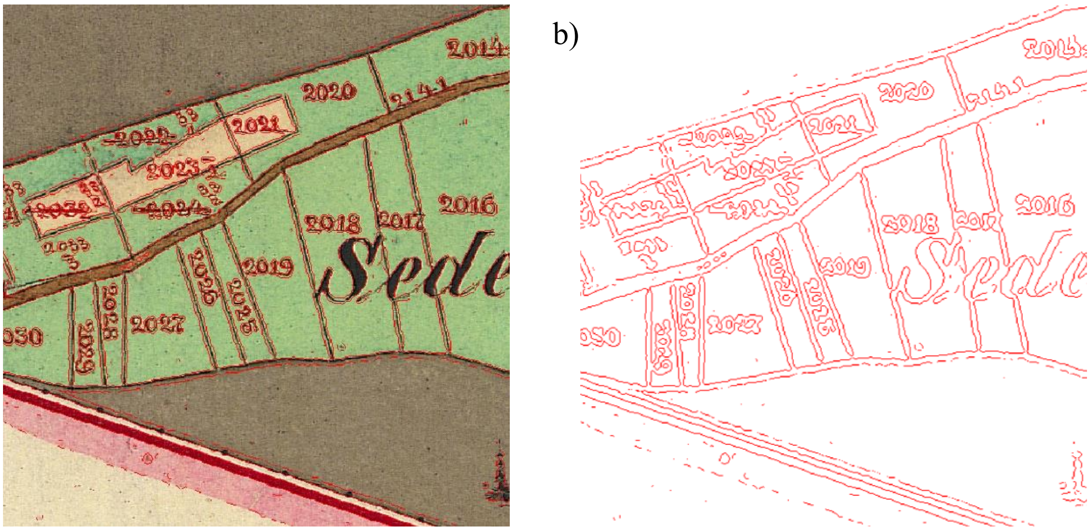

Figure 5: application of the "zero crossing" algorithm to the cadastre map (particular) a) with and b) without original map in background. Double lines for each boundary are present.

Different approaches have been tried to extract a single line. On the raster maps the median line could be found by intersecting two buffers around the two parallel lines: the problem with this approach is that the distance between the lines, which depends on the thickness of the line on the map, is different for each line. Therefore, it is impossible to find a buffer width suitable for every line. Another possibility is to use an algorithm creating the median lines after transforming the map into a vector map, but this approach cannot be used in this case because it assumes that the pair of lines whose median line must be found are uniquely identifiable by their category or by an attribute. In the case of the application of the "zero-crossing" algorithm all the lines have the same category.

4.2.2 Colour filtering

The boundaries of parcels have a black/dark grey colour on the map. The colour varies because the map is handwritten, and it was more than one century old at the time that the digital copy has been created. Moreover, the low-resolution scan mixes the colour of the of the lines and of the surrounding areas.
The image is encoded using 8 bit per colour; therefore the black colour corresponds to the RGB triplet (0,0,0) while white corresponds to (255,255,255). However, because of the colour variations described above, the thresholds for extracting the boundary lines have been set to (130,155,130).
Selecting all the pixels with RGB values below these thresholds identifies clusters of pixels only in part corresponding to the boundary lines (figure 6a). Therefore, a thinning algorithm has been applied to uniquely identify the position of the boundary line (Jang and Chin 1990) as shown in figure 6b.

a) b)

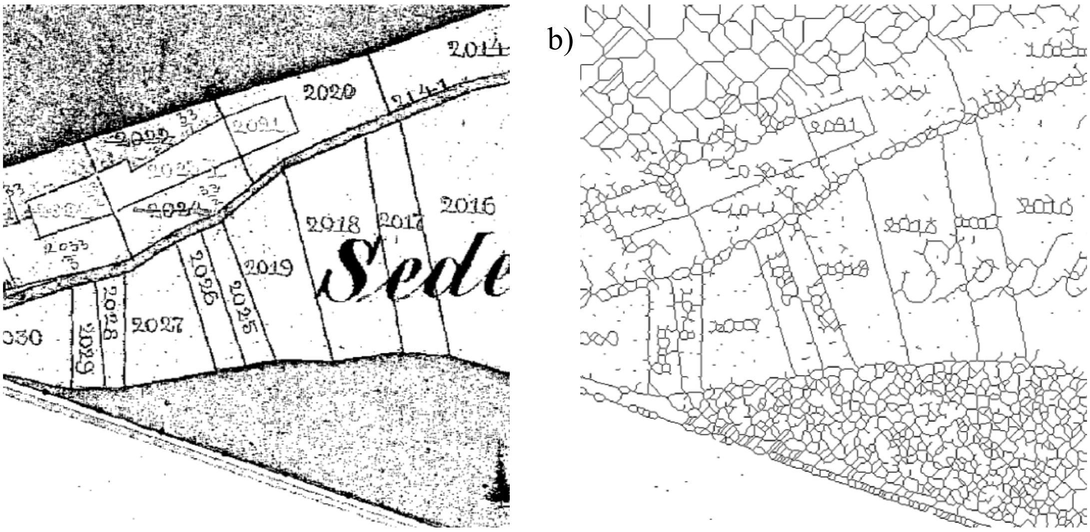

Figure 6: a) pixels remaining after selecting only values below the thresholds and b) map after thinning.

It is then possible to create a vector map containing the lines corresponding to the boundaries. The vector lines are filtered (figure 7) to remove artefacts using both the geometry, by removing very short lines and snapping vertices within a threshold, and the topology, deleting the so called “dangles” (i.e. small lines protruding from other lines); while these operations clean the map and get rid of lines resulting from the noise on the image, the problem of the gaps in the original lines remains. In principle, it should be possible to close the gaps by snapping the vertices using a suitable distance threshold, however some gaps are very large, so much so that the closest vertices to the initial o ending points are not the vertices on the other side of the gap.

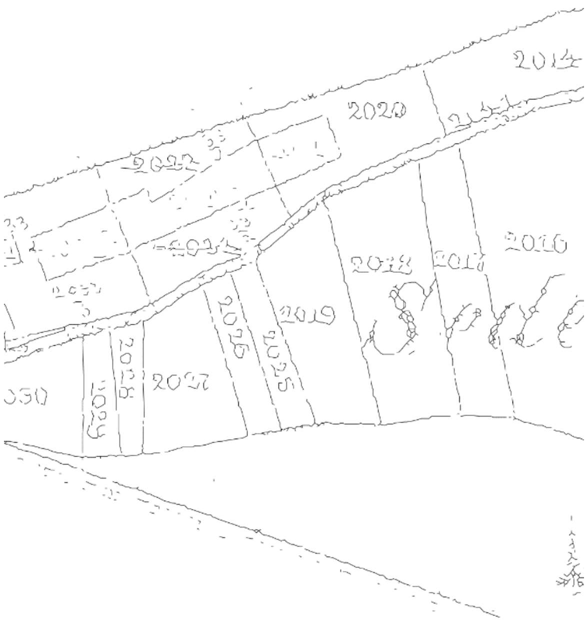

Figure 7: vector map after filtering.

To solve this problem a new approach is under study, which would have the additional benefit of regularizing the boundary line, therefore helping in reducing the effect of the noise on the image.

So far none of the tested approach has yield satisfactory results, however an approach using energy density and the shear transform by Miao et al. (2013) seems to be promising.

5. Discussion

As said earlier, this paper presents the current status of an ongoing work. As shown in the result section, the work encountered various issues with the vectorization process. In the first place the OBIA method is problematic due to the difficulties in setting the threshold parameter. Setting a low value would result in over-segmentation, meaning that in a single parcel more than one segment are created. Setting a higher value would resolve this problem in these areas, but it would also result in under-segmentation in other areas where the separation lines are not distinctly defined, merging different parcels in the same segment. This problem is due to the fact that the map is hand-coloured, making it harder to identify the single features, whereas in printed map colours are more uniform and it is easier to select an optima threshold. Another issue is the resolution of the scan, being at only 230 PPI is not enough to properly represent on the digital map the geometric feature present in the original map, especially when it comes to the separation lines of parcels. As it can be noted in the red circles in figure 8, there are gaps in the lines in the original map and the colour of the pixels that should represent the line is more green than black, making it almost impossible to detect for the algorithm. This represents a difficult obstacle to overcome. At the time of the writing, some solutions are explored based on the work of Spinello and Guitton (2004) and Miao et al. (2013).

a) b)

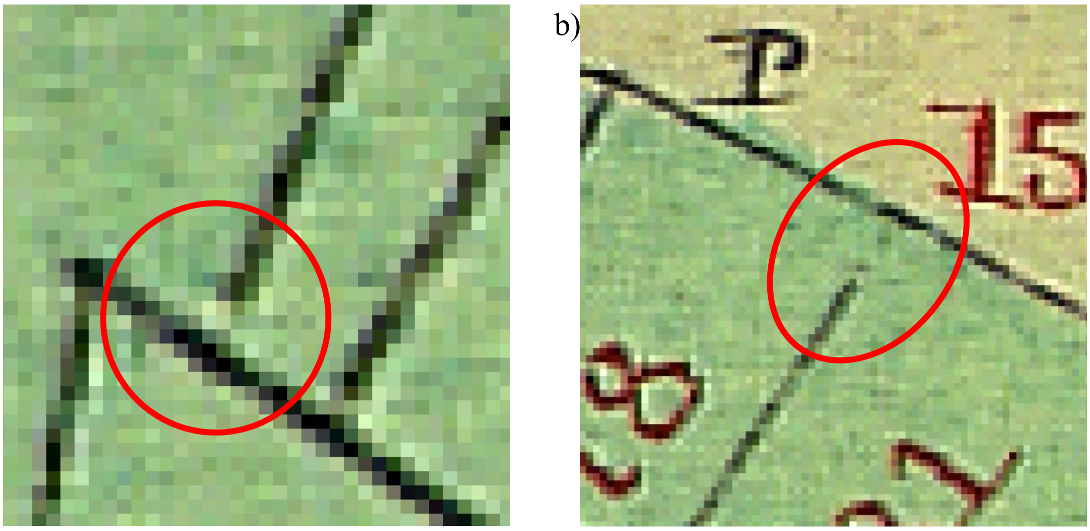

Figure 8: a) and b) examples of gaps in boundary lines in the original map.

6. Final remarks and perspectives

Although many methods are available for the vectorization of historical maps, not all maps are suitable for a standard approach, applied as it is described in scientific literature. As it is shown in this paper, to obtain a satisfying result, it is often necessary to combine different techniques, and sometimes to explore new ways in order to succeed. This article is an example of how vectorization of historical maps can still be a complicated task on which new studies and researches are needed. This is particularly relevant for hand-drawn maps, in which shapes and colours can be different for the same class of object on a single map sheet. Moreover, small imperfections which do not hinder the ability of a human to read correctly the map, can have a huge impact on the ability of algorithms to extract the relevant information preserving its consistency.

The final goal of the project is to create a land use and woodland cover map, which will also include information on land structure and property systems. This source map could be profitably compared with the current one both to verify changes in land use and to assess the impact of property systems on these changes. Therefore, digitizing each individual parcel is crucial. The work will continue on other sheets from the same source to refine the process and address open issues.

Acknowledgments

This work is funded by the European Union – Next Generation EU, under the PRIN 2022 call, project "Bridging geography and history of woodlands: analysing mountain wooded landscapes through multiple sources and historical GIS" (2022EKECST) – CUP E53D23010170006.

References

Asokan A., J. Anitha, M. Ciobanu, A. Gabor, A. Naaji, D. J. Hemanth (2020). Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification-An Overview. *Applied Sciences* 10 (12): 4207.

Bacior, S. (2023). Austrian Cadastre still in use – Example proceedings to determine the legal status of land property in southern Poland. *Land Use Policy* 131.

Besana A., M. Ciolli, S. Gobbi, P. Zatelli (2019). La cartografia di Cesare Battisti: sperimentazione di una procedura di digitalizzazione automatica. Proceedings of the XXIII Conferenza Nazionale ASITA. Milano: Federazione delle Associazioni Scientifiche per le Informazioni Territoriali e Ambientali.

Buffoni D., S. Endrizzi, T. Gilardi (2015). La mappa catastale asburgica ottocentesca: interpretazione di colori, segni e simboli nel paesaggio rurale trentino. In E. Dai Prà (ed.), *Approcci geostorici e governo del territorio*. Milano: Franco Angeli, 56-65.

Dai Prà, E. (ed.) (2013). *APSAT 9: cartografia storica e paesaggi in Trentino: approcci geostorici*. Mantova: SAP Società archeologica.

Dai Prà E. and N. Gabellieri (2021). Geostoria applicata del bosco e fonti cartografiche geodetiche e pre-geodetiche: casi studio di foreste trentine colpite dalla tempesta Vaia. In *Bollettino dell'Associazione Italiana di Cartografia*. Trieste: EUT Edizioni Università di Trieste, 16-30.

Diamantini C. and A. Franceschini (2015). L'evoluzione del modello insediativo del Trentino: dal catasto asburgico alle immagini aerofotogrammetriche. In E. Dai Prà (ed.), *Approcci geostorici e governo del territorio: 1.Alpi Orientali*. Milano: Franco Angeli, 254-262.

Dolejš M. and M. Forejt (2019). Franziscean Cadastre in Landscape Structure Research: A Systematic Review. *Quaestiones Geographicae* 38 (1): 131-144.

Gabellieri, N. (2019). Il paesaggio scomparso della gelsicoltura in Vallagarina: un approccio geografico-storico a fonti integrate. *Geostorie* 27 (1): 57-78.

Gede M., V. Árcai, G. Vassányi, Z. Supka, E. Szabó, A. Bordács, C. G. Varga, K. Irás (2020). Automatic Vectorisation of Old Maps using QGIS -Tools, Possibilities and Challenges. In *Automatic Vectorisation of Historical Maps: International workshop organized by the ICA Commission on Cartographic Heritage into the Digital*. Department of Cartography and Geoinformatics ELTE, 39-45.

Gilardi, T. (2013). La mappa catastale asburgica ottocentesca: breve guida semiologica per il paesaggio rurale trentino. In E. Dai Prà (ed.), *APSAT 9: cartografia storica e paesaggi in Trentino: approcci geostorici*. Mantova: SAP - Società Archeologica.

Gobbi S., M. Ciolli, N. La Porta, D. Rocchini, C. Tattoni, P. Zatelli (2019). New Tools for the Classification and Filtering of Historical Maps. *ISPRS International Journal of Geo-Information* 8 (10): 455.

Grava M., C. Berti, N. Gabellieri (2020). *Historical GIS: strumenti digitali per la geografia storica in Italia*. Trieste: EUT, Edizioni Università Trieste.

Gregory I. N. and P. S. Ell (2007). *Historical GIS: Technologies, Methodologies, and Scholarship*. Cambridge: Cambridge University Press.

Jang B.-K. and R. T. Chin (1990). Analysis of thinning algorithms using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (6): 541-551.

Iosifescu I., Tsorlini, A., Hurni, L. (2016). Towards a comprehensive methodology for automatic vectorization of raster historical maps. *e-Perimetron* 11 (2): 57-76.

Iras K. (ed.) (2020). Automatic vectorisation of Historical Maps. Proceeding of the *International workshop organized by the ICA Commission on Cartographic Heritage into the Digital*.

Krishnamoorthi R. and P. Bhattacharya (1998). Color edge extraction using orthogonal polynomials based zero crossings scheme. *Information Sciences* 112 (1-4): 51-65.

Miao Q., P. Xu, T. Liu, Y. Yang, J. Zhang, W. Li (2013). Linear Feature Separation From Topographic Maps Using Energy Density and the Shear Transform. *IEEE Transactions on Image Processing* 22 (4): 1548-1558.

Pavelková R., J. Frajer, M. Havlíček, P. Netopil, M. Rozkošný, V. David, M. Dzuráková, B. Šarapatka (2016). Historical ponds of the Czech Republic: an example of the interpretation of historic maps. *Journal of Maps* 12 (sup1): 551-559.

Provincia autonoma di Trento. (n.d.). Mappe storiche d'impianto. *https://catastotn.tndigit.it/mappe-storiche-impianto/it/index.html*

Sarabia J., G. Galeazzi, A. Busnardo (2016). I paesaggi dell'incolto. In J. Sarabia and G. P. Brogiolo (eds.), *Drena: Insediamenti e Paesaggi Dai Longobardi ai Nostri Giorni*. Mantova: SAP Società Archeologica, 145-171.

Schlegel, I. (2023). A Holistic Workflow for Semi-automated Object Extraction from Large-Scale Historical Maps. *KN - Journal of Cartography and Geographic Information* 73 (1): 3-18.

Spinello S. and P. Guitton (2004). Contour Line Recognition From Scanned Topographic Maps. *Journal of WSCG* 12 (1-3): 419-426.

Tattoni C., M. Ciolli, F. Ferretti, M. Cantiani (2010). Monitoring spatial and temporal pattern of Paneveggio forest (northern Italy) from 1859 to 2006. *iForest - Biogeosciences and Forestry* 3: 72-80.

Timár G. and S. Biszak (2010). Digitizing and georeferencing of the historical cadastral maps (1856-60) of Hungary. Proceeding of the *5th International Workshop on Digital Approaches in Cartographic Heritage*, 559-564.

Zatelli P., N. Gabellieri, A. Besana (2022). Digitalization and Classification of Cesare Battisti's Atlas of 1915. *ISPRS International Journal of Geo-Information* 11 (4): 238.

Zatelli P., S. Gobbi, C. Tattoni, N. La Porta, M. Ciolli (2019). Object-based Image Analysis for HistoISTORIC MAPS CLASSIFICATION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4/W14: 247–254.