Intercontinental Geoinformation Days

igd.mersin.edu.tr

Generating temporal cadastral parcels with artificial intelligence algorithms within the scope of cadastre 2034

Fatma Betül YIDIZ*[1], İsmail Ercüment AYAZLI [2], Hidayet TAKCI [3]

[1]*Sivas Cumhuriyet University, Engineering Faculty, Department of Geomatics Engineering, Sivas, Turkey*
[2]*Sivas Cumhuriyet University, Engineering Faculty, Survey Engineering Department, Sivas, Turkey*
[3]*Sivas Cumhuriyet University, Engineering Faculty, Computer Engineering Department, Sivas, Turkey*

Keywords
Cadastre 2034
Land Management
Digital Image Processing
Artificial Intelligence
Deep Learning

ABSTRACT

Since real estates are a reliable investment tool, many changes occur in ownership and parcel geometry. Therefore, temporal cadastral data have a great importance in terms of sustainable development policies because cadastre can provide the main data of smart cities. The Cadastre 2034 Vision started by the International Federation of Surveyors (FIG) proposes to record the temporal dimension of the cadastral data. The temporal dimension of cadastral parcels are stored in the documents named as "fenni evraklar" (technical papers) in Turkey. The study aims to develop a new model in which the temporal dimension of cadastral parcels will be automatically digitized in accordance with the four-dimensional cadastral approach targeted in the Vision of Cadastre 2034. Therefore, an interface using artificial intelligence algorithms was created in the Python programming language and changes in cadastral parcels can be monitored at the beginning from the first facility cadastre on any cadastral parcel.

1. INTRODUCTION

Since the second half of the 20th century, due to the rapid increase in the world population, "new policy for sustainable development (land management) about the use and protection of land, and it is a tool for the implementation of these policies, and determining, registering and recording information about the ownership, value and use of real estate and publishing process" (Cagdas & Gur, 2003) has spawned the concepts of land administration. Since the early 2000s, cadastre has been defined as a spatial information system modeling human-land relation (Demir, 2001; Ayazlı, 2006) and evolved into land administration, which is the implementation tool of sustainable development policies (Grant & Williamson, 1999; Enemark, 2001; Cagdas ve Gur, 2003; Enemark, 2009). As it can be perceived from the definition, the cadastre is one of the most important tools necessary for the efficient use of natural and environmental resources, and it is a geospatial information system in which register the real estate properties guaranteed by the constitution. Ownership of the real estates in Turkey starts with the registration date and it may change or end in time by buying and selling and/or land regulations. Therefore, cadastre is a living entity, hence it has a temporal dimension. In this context, cadastre is defined as a four-dimensional (4D) that is a space-time dependent entity. (Stoter, 2004; Ayazli vd., 2011). To design the future of the cadastre after 2014, a paper was published in 2010, the Cadastral Future: Setting Up a New Vision for Nature and the Role of Cadastre (Bennett et al., 2010). In the paper, they discussed the factors affecting the developments in the cadastral field as globalization, urbanized population, good governance, climate change, environmental management, three-dimensional visualization/analysis technologies, wireless networks, standardization and interoperability (Bennett et al.,2010). Under the influence of the factors, the objectives of Cadastre 2034 carried out by FIG are specified as accurate, object-oriented, four dimensional, real-time and global cadastre. (Steudler, 2014; Bennett et al., 2010; LINZ, 2014; Polat and Alkan, 2015; ICSM, 2015). In this context, the cadastre maintains the responsibility to reconstruct the infrastructure that will form the basic data for smart cities and geographical information systems (GIS) by a contemporary perspective (Kaufmann and Steudler 2002; Bennett vd, 2010; (Anzlic Committee on Surveying & Mapping, 2015).

In Turkey, cadastral studies are carried out in accordance with the provisions of the "Cadastral Law" numbered 3402 and 5304. Temporal data in the cadastral system are kept in called technical documents, technical folders, change folders and application folders, and the land registers and the condominium rights. These sources contain over a hundred attribute in total

* **Corresponding Author**

*(20199249005@cumhuriyet.edu.tr) ORCID ID 0000-0002-2152-5661
(eayazli@cumhuriyet.edu.tr) ORCID ID 0000-0003-0782-5366
(htakci@cumhuriyet.edu.tr) ORCID ID 0000-0001-6995-5073

Cite this study

Yıldız F.B, Ayazlı İ.E & Takcı H. (2021). Generating temporal cadastral parcels with artificial intelligence algorithms within the Ssope of cadastre 2034. 2nd Intercontinental Geoinformation Days (IGD), 96-99, Mersin, Turkey

and considering them recorded in whole country, it is obvious that the data density is very high.

In Turkey, cadastral technical documents were scanned within the scope of the "Cadastral Data Consolidation" (CDC) project and converted into "Portable Document Format" (pdf). The aim of this study is to extract temporal data of cadastral parcels from technical documents as fully automatically using artificial intelligence algorithms. In this context, technical documents created within the scope of the CDS project are automatically processed with a prepared software and temporal cadastral parcel data are generated according to the detected coordinate data.

2. METHOD

Within the scope of this study, technical folders with .pdf extension produced in the scope of CDS project in Sancaktepe district of Istanbul were used. At the beginning of the study, the .pdf data were converted into an image format using the Python programming language, then textual expressions and numerical data on the image were separated and saved in a separate file. The data were filtered using digital image processing methods as they contain noises. They were filtered using noise pixel cleaning and "Threshold Gaussian (ATG)" methods.

A graphical user interface (GUI) has been designed. Simultaneous processing has been done in the software with automatic detection.

The process steps followed in the study are as follows:

- GUI design in Python and QT Designer environment.
- Import of libraries to be used (Opencv, PyQt5, Tesseract, NumPy, Matplot, Openpyxl) in Python environment.
- Determination of .exe file locations of the environment where the original textual expressions will be extracted.
- Adjusting the folders to add images in the GUI environment and determining the sizes of the folders.
- Performing image processing steps. These image processing methods are: OpenCv file operations, the control of pixel values in terms of color channels, grayscale image acquisition and thresholding filters.
- Configuration settings for extracting textual and numerical data
- Extracting and printing textual and numerical expressions.
- Performing file processing steps in Python. It covers cases such as opening and closing files in Excel and text document formats, processing values into files and placing them in a meaningful way.
- Final checks in GUI design.

2.1. GUI Design

GUI design was carried out on Python Qt Designer. Functions were created on the design by importing it in Python with PYQT5.

The image adding section, the reading area and the fields showing the separation of the characters are added to the design. The image was read with the "read" button and separation operations can be done. All functions were performed by connected to the button. The scope of functions constitutes image processing and textual expressions in the image. The GUI design is shown in Figure 1.

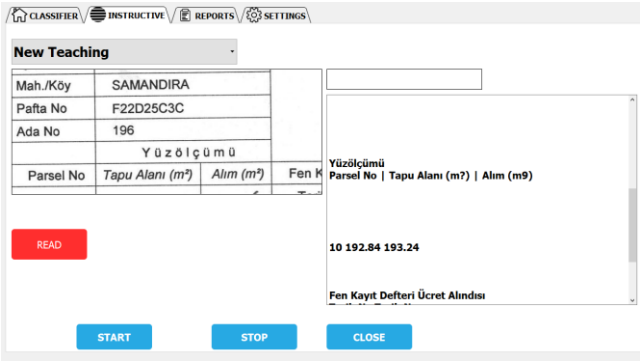

Figure 1. GUI design structure

2.2. Image processing

Image processing techniques have been used as the extraction of text on the image does not give an entirely acceptable result when raw images are used directly. Some image processing steps applied to the image can be listed as follows:

Removing the remaining areas outside the text, finding and sharpening text by converting it to an 8-bit image, resizing the image, clearing unnecessary and non-textual expressions on the text with adaptive thresholding method as images are obtained in different areas and under different lighting conditions and Gaussian Filter, one of the image filters that can make the text more readable, has been applied to soften the image. As a result, the results in Figure 2 were obtained.

Texts may not appear directly in .pdf format as a straight page. In this case, the images must be rotated. For this reason, these images were detected first and this problem was resolved.

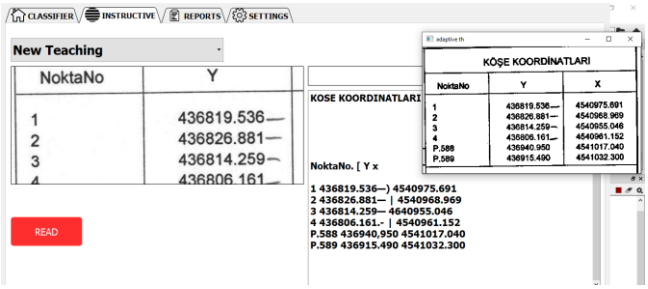

Figure 2. Results from image processing

As represented in the Figure 3, the image was first cleaned, then each character was perceived separately and character segmentation was made as in the Figure 4.

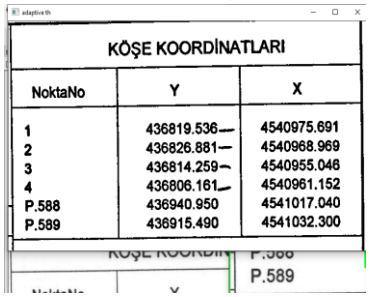

Figure 3. Pre-processed image

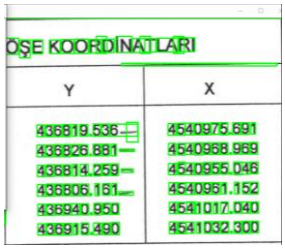

Figure 4. Separating characters from the image

2.2.1. Data processing and extraction

Python-tesseract and EAST deep learning text sensor, Ocular, OCRopus, SwiftOCR (deep learning), LSTM techniques used in text processing were tested in Python language. The best method has been determined.

The texts of the technical folder data with .pdf extension seen in the Figure 5 have been accessed. In the first stage, text files were created for the text extracted with letters, numbers and different signs. Since these need to be cleared, some adjustments have been made on the model to create meaningful text. As settings, fine-tuning has been made, such as language setting, numerical and textual expression detection, navigation setting, and page segmentation mode. Language setting is a setting used to detect different languages and parameter values are given accordingly. Page segmentation mode setting takes and accepts parameters by making a selection according to the numbered list. Thanks to these settings, more meaningful expressions were created in the text and printing operations were realized.

Python-tesseract used in the work is an optical character recognition (OCR) tool for Python. It is an auxiliary tool used to find embedded text on images. Another tool used is the EAST deep learning text detector tool. These two tools offer the best techniques and settings for text recognition. At the same time, these tools can work in real time and perform text finding processes simultaneously.

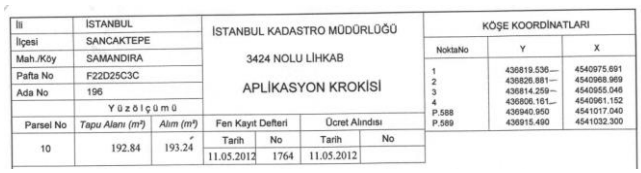

Figure 5. Attributes in technical folder

3. RESULTS

There are a number of problems such as ambiguous texts, light reflections, blurry images in the .pdf extension data created within the scope of the CDS project. To overcome these problems, Python-tesseract and EAST deep learning text detectors are used.

Noises on data used as input images were cleared by passing through filters such as Gaussian, grayscale and thresholding. The cleaned image was processed with the help of the text detector function algorithm. As the result data, the texts were extracted from the image simultaneously.

The Figure 6 shows the process step in which the results obtained as a result of the software are written on the file simultaneously and the accuracy checks are carried out.

The main parts of the software are schematized in the Figure 7.

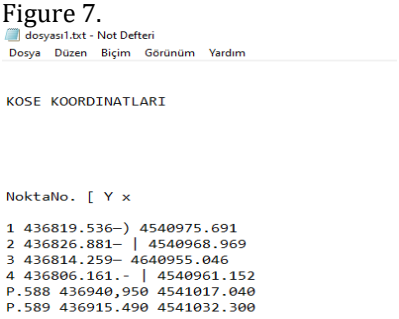

Figure 6. Outputs

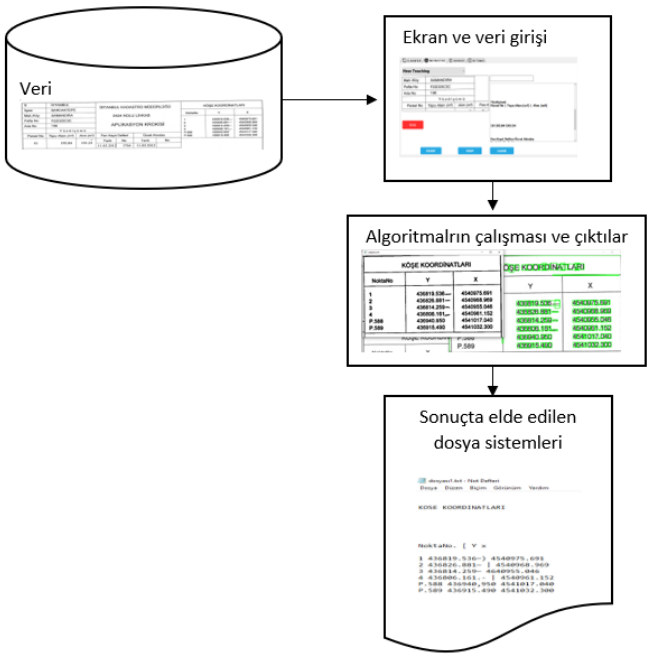

Figure 7. Overwriting files and checking accuracy

4. CONCLUSION

Since real estates are a reliable investment tool, the number of transactions carried out in the land registry and cadastre offices are increasing day by day. In this study, it was aimed to provide a more dynamic tracking system of cadastral records and a solution was presented.

According to the results, it has been possible to digitize the data kept in different file formats automatically, regardless of the format, thanks to the developments in computer technologies.

The proposed solution was encoded in Python programming language, and the texts were extracted and stored with artificial intelligence algorithms on the desktop application. In this way, the changes occurring from starting the first facility cadastre in a cadastral parcel can be monitored graphically.

REFERENCES

Adrian at PyImageSearch (2020). Start Here with Computer Vision, Deep Learning, and OpenCV. Mart 15, 2021- https://www.pyimagesearch.com /start-here/

Anzlic Committee on Surveying & Mapping. (2015). *Cadastre 2034: Powering Land & Real Property*. https://www.icsm.gov.au/sites/default/files/Cadastre2034_0.pdf

Ayazli, I. E., Helvaci, C., Batuk, F., & Stoter, J. (2011). Designing three dimensional property right database. African Journal of Business Management, 5(22), 9440-9447.

By Mathias Lemmens, Delft University of Technology, The Netherlands, Towards Cadastre 2034, September 7,2010,https://www.gim-international.com/content/article/towards-cadastre-2034

Bennett, R., Kalantari, M., & Rajabifard, A. (2010, 06 28). Beyond Cadastre 2014. 12 30, 2017 GIMInternational:https://www.gim-international .com/content/article/beyond-cadastre-2014

Cagdas, V., & Gur, M. (2003). Sustainable Development and Evoulation of the Cadastre. HKM Jeodezi, Jeoinformasyon ve Arazi Yönetimi Dergisi(89), 42-48. http://www.hkmo.org.tr/resimler/ekler/M6K2_f877f38657_ek.pdf

Demir, H. (2001). Kadastro Bilgisi Ders Notları. İstanbul

Enemark, S. (2001). Land Administration Syestems -a major challenge for the surveying profession. A Spatial Odyssey : Proceedings of the XVIII American Surveying Teachers Conference. Penn State University, Wilkes Barre, USA.

Grant, D., & Williamson, I. (1999). The Bathurst Declaration on Land Administration for Sustainable Development. FIG. Retrieved from http://www.fig.net/resources/publications/figpub/pub21/figpub21.asp

Firhan Maulana Rusli, Kevin Akbar Adhiguna, Hendy Irawan (2020). Indonesian ID Card Extractor Using Optical Character Recognition and Natural Language Post-Processing. arXiv:2101.05214v1 [cs.CV] 15 Dec 2020. https://arxiv.org/pdf/2101.05214.pdf

Gorsa Lakshmi Niharika; Shahana Bano; Pavuluri Shyam Kumar; Tinnavalli Deepika; Hampi Thumati (2020). Character Recognition using Tesseract enabling Multilingualism. ISBN:978-1-7281-6388-8 DOI: 10.1109/ICECA49313.2020.9297609

ICSM. (2015). Cadastre 2034: Powering Land and Real Property. 12 30, 2017-The Intergovernmental Committee on Surveying and Mapping Web Page: http://www.icsm.gov.au/cadastral/Cadastre2034.pdf

Kaufmann, J., & Steudler, D. (1998). Cadastre 2014; A VIision for a Future Cadastral System. FIG. Retrieved from http://www.eurocadastre.org/pdf/cadastre_2014 .pdf

LINZ. (2014). Cadastre 2034, A 10–20 Year Strategy for developing the cadastral system: Knowing the 'where' of land-related rights. Land Information New Zealand.

Matthias A Lee (2020). Python Tesseract . Mart 15, 2021 https://github.com/madmaze/ pytesseract

Mehmet Alkan, Hicret Gursoy Surmeneli, Zeynel Abidin Polat, fig congress 2018 - Design and determine cadastral and land management performance of Turkey with Cadastre 2034 Vision, FIG Congress 2018: Embracing our smart world where the continents connect - Enhancing geospatial maturity of societies

Neha Joshi (2019). Text Image Extraction and Summarization. ISSN NO: 2350-1146 I.F-5.11

Polat, Z. A., & Alkan, M. (2015). Türkiye'nin Kadastro 2014 Performansı ve Kadastro 2034'ten Beklentiler. The World Cadastre Summit 2015. Istanbul.

Sahil Thakare; Ajay Kamble; Vishal Thengne; U.R. Kamble (2018). Document Segmentation and Language Translation Using Tesseract-OCR. ISSN: 2164-7011 https://ieeexplore.ieee.org/document/8721372-DOI: 10.1109/ICIINFS.2018.8721372

Sergio Canu (2020). Text recognition (OCR) with Tesseract and Opencv. Mart 15, 2021 https://pysource.com/2020/04/23/text-recognition-ocr-with-tesseract-and-opencv/

Steudler, D. (2014). Cadastre 2014 and Beyond. Copenhagen: FIG.

Stoter, J. E. (2004). 3D Cadastre. Delft: Publication on Geodesy

Sonia Bhaskar, Nicholas Lavassar, Scott Green (2020). Implementing Optical Character Recognition on the Android Operating System for Business Cards. EE 368 Digital Image Processing.

The Tesseract open source OCR engine, http://code.google.com/p/tesseract-ocr.

Open Source Computer Vision (2021). OpenCV Mart 15, 2021- https://docs.opencv.org/master/d6/d00/tutorial_py_root.html

Zeynel Abidin POLAT, Mustafa USTUNER and Mehmet ALKAN, On the Way to Vision of Cadastre 2034: Cadastre 2014 Performance of Turkey, FIG Working Week 2015 From the Wisdom of the Ages to the Challenges of the Modern World Sofia, Bulgaria, 17-21 May 2015