Automating Measurement Extraction from Deeds

Amir Bar-Maor, Netherlands, Christine Leslie, United States of America and Tim Hodson, United States of America

Key words: deeds, dimensions, automation, optical character recognition, OCR

SUMMARY

Many legal descriptions in the USA use the metes-and-bounds system to describe land boundaries. Metes-and-bounds descriptions define parcel boundaries using physical features, references to other parcel corners, and directions and distance measurements.
Parcel boundaries are typically described in a sequence starting from a point-of-beginning and returning to the same point. Straight lines are described using a direction and a distance, while curved lines are described with a direction, radius, and another curve parameter such as an arc length, chord length or a delta angle.
Map technicians manually type these measurements, also known as COGO (Coordinate Geometry) dimensions, into software to construct and map parcel boundaries.
While data entry of COGO dimensions has been made efficient and user-friendly over the years, it is still a labour-intensive process that is prone to data entry typos and mistakes.
This paper presents a novel approach utilizing Optical Character Recognition (OCR) technology to extract measurement data from legal deeds and convert it into numerical values. The extracted measurements undergo a review process before being transformed into Coordinate Geometry (COGO) values compatible with the software. This paper details the implementation of this innovative capability and highlights its potential to significantly enhance the productivity of thousands of map technicians across the United States.

Automating Measurement Extraction from Deeds

Amir Bar-Maor, Netherlands, Christine Leslie, United States of America and Tim Hodson, United States of America

METES-AND-BOUNDS LEGAL DESCRIPTIONS

The metes-and-bounds system is a method used to describe land boundaries in the United States. Metes are the measurements of the boundary lines, typically described in terms of direction and distance. For example, a boundary might be described as running "North 54°32’10” East a distance of 100.23 feet."
A typical deed contains multiple courses that depict the parcel boundaries. Courses can be straight lines as well as circular arcs. A deed may represent one or more parcels that each start from a well-established point of beginning and loop back to the same point of beginning.

COMMENCING at the West Right-of-Way line of Reddleshire Lane 50' wide, said point being the most Southerly cutback point of Lot 1, Block 5 as originally platted in Westwick Section 1 as recorded in Volume 258, Page 143 of the Harris County Map Records;

THENCE S 87° 30' 01" W, a distance of 14.60 feet to the POINT OF BEGINNING of the herein described tract;

THENCE, S 03° 25' 14" E, a distance of 3.17 feet to a point for corner;

THENCE, N 86° 35' 46" E, a distance of 4.00 feet to a point for corner;

THENCE, S 03° 25' 14" E, a distance of 6.00 feet to a point for corner;

THENCE, S 86° 34' 46" W, a distance of 4.00 feet to a point for corner;

THENCE, S 03° 25' 14" E, a distance of 18.42 feet to a point for corner;

THENCE, S 86° 34' 46" W, a distance of 57.73 feet to a point for corner;

THENCE, N 01° 21' 30" W, a distance of 27.60 feet to a point for corner;

THENCE, N 86° 34' 46" E, a distance of 56.74 feet to the POINT OF BEGINNING and containing 1602 square feet of land, more or less.

Figure 1: Deed with courses depicting parcel boundaries

Coordinate geometry (COGO) is used to map feature locations and boundaries based on ground measurements recorded in survey plans, deed descriptions, and other physical or electronic land records.

COGO measurements typically describe features relative to each other. The following example survey plan shows COGO measurements for a road centerline and the dimensions relative to its adjoining parcel boundaries. (Esri Documentation Reference 1)

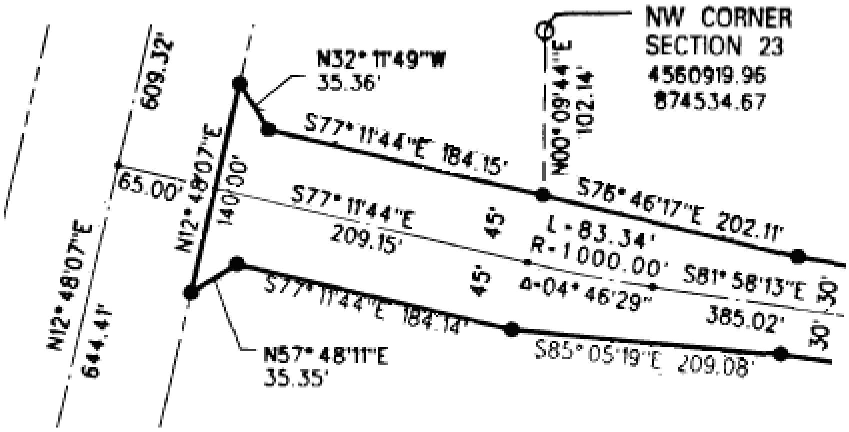

Figure 2: Survey plan showing COGO measurements

Survey plans can include references to points with known coordinates within a coordinate system. These points are known as control points, monuments, or cadastral reference points.

SYSTEM OF RECORD

A system of record is an information storage system that serves as the authoritative data source for a particular data element or piece of information. It ensures the integrity and accuracy of the data it holds, making it the primary source for that data within an organization. (Wikipedia, Reference 1, 2)

Cadastral systems qualify as systems of record because they manage crucial data that supports the conveyance of land, land tenure, taxation and more. The ArcGIS parcel fabric technology supports workflows that are record-driven, meaning that the legal land documents drive the process, and are referred to as record-driven-workflows.

REPRESENTING THE LEGAL AND PHYSICAL WORLD

The parcel fabric represents the physical world by mapping the parcel features and their boundaries, while also representing the legal, recorded boundary line measurements. The physical representation and mapped position of parcel boundaries may change over time, but the legal recorded measurements are never changed in this system of record. Each COGO measurement is associated with the legal source document from which it came.

In the parcel fabric, parcel boundaries are stored in a lines feature class. The line feature classes in the fabric are automatically COGO-enabled for all parcel types. This means that they have additional attribute fields to store values from cadastral documents, such as the directions, distances for straight-line boundaries, or radius and arclengths for circular arc boundaries. It is also possible for these fields to be empty if one or all the values are not on the original document.

The parcel fabric also keeps track of the source from which each dimension was derived: it distinguishes between dimensions that were entered from a legal document, dimensions that were computed from the geometry, and dimensions that were calculated. The ability to compare the legal and the physical dimensions can help to identify blunders as part of quality-driven workflows.
COGO-lines are expected to be two-point, single-segment lines though there are exceptions such as natural boundaries like rivers and shorelines. Since parcels are usually defined by a closed loop of two-point lines that store dimension information, a mathematical misclosure can be computed and stored on the polygon feature and provides quality check on the line dimensions.

GROUND TO GRID CORRECTION

It is typical for surveyors in the USA to produce digital representations of parcel (and other) data in a ground coordinate system. The intent is for the computed distance between any two points in a CAD drawing to represent the “true” horizontal ground distance as would be measured between the same two physical locations in the field. In GIS datasets the map projection is referenced to the ellipsoid and projected to the “grid.”

Ground to grid correction converts the bearing and distance values into the coordinate system of the map displaying the feature data. When configured in ArcGIS Pro, it is a drawing mode that automatically corrects bearing and distance values entered when using editing tools. (Esri, Documentation Reference 2) The entered values are always stored in their original, uncorrected form, as these represent the true ground values, and represent the values from the official source as recorded at the county.

CREATING A PARCEL FROM A DEED

Even though the original data prepared surveyors is in digital vector formats, like CAD, the deed descriptions are required to be recorded at the county in the textual form shown in Figure 1. The original CAD is not included, and very few counties have a digital submission process.

Deeds ready to be recorded are on physical paper with official stamps and signatures. These documents are then scanned and converted into a digital image format or PDF. Bringing these documents into an information system database requires a manual data entry process by county mapping technicians, using the official document as the source. The technician interprets the language of the deed and transcribes the sequence of directions and distances by manually typing them using a traverse tool.

OPTICAL CHARACTER RECOGNITION (OCR)

Optical Character Recognition (OCR) is a technology that converts different types of documents such as scanned paper documents, PDFs, or images captured by a digital camera, into editable and searchable data. OCR transforms images of text into machine-readable text.

OCR technology has been available for decades, but the specialized language of deed descriptions has made the automated extraction of their COGO dimensions into a traverse format uniquely challenging. In a forthcoming release of ArcGIS Pro, the parcel fabric technology will include a tool called *COGO Reader*, the primary subject of this paper, and described later herein.

COGO TRAVERSE TOOL

In ArcGIS Pro, parcel dimensions are entered using a traverse tool. The traverse tool is designed to make the process of entering dimensions as efficient as possible.

Single-handed data entry

The traverse tool supports single-handed data entry, where map technicians can use the numeric keypad, with one hand, to enter dimensions from a deed. Every course entered creates a feature on the map. Smart labeling is used to show the entered COGO values on the parcel boundaries as they are created.

Keyboard shortcuts

To improve data entry efficiency, a set of keyboard shortcuts is available, which allows exclusive use of the numeric 10-key pad. Use of the mouse can be avoided.
For example, deeds in the US typically use the quadrant bearing format with the “N” “W” “S” and “E” letters paired to specify the quadrant of directions. The traverse tool will accept the directly entered string including the alpha characters, but experienced technicians can use the quadrant short-cuts instead. If a direction on a deed reads N44°45'30"W technicians can use the “–4” suffix to indicate the northwest quadrant, and can enter the following:
44-45-30-4 (All characters from the 10-key pad)

This is an example of one shortcut available for the traverse tool. An extensive list of documented keyboard shortcuts is available for rapid and efficient data entry of dimensions.

Circular arcs

The user can easily switch from entering straight lines to curves. Curves can be entered using a variety of parameters such as radius and arc length or chord length or a delta angle.

Misclose used for quality checks

A closed traverse begins at a known point of beginning and either returns to that same point or ends at another point with a known location. Comparing the calculated end-point coordinates with the known, expected location for the point provides a mathematical consistency check. This ensures that there are no data entry mistakes, and that there are no incorrect values on the source document.

For most traverses a relatively small misclose is to be expected, and the cause depends on the source of the data. Causes can include a loss numerical precision due to the rounding of numbers, or direct field-observed measurements that contain measurement error.
The misclose ratio is l ratio of the misclose distance divided by the overall length of the traverse, and a typical requirement is for the misclose to be smaller than 1:5000.
If the misclose is too large, this indicates a mistake in the data, or in the data entry. The use of misclosure is therefore an important quality metric for the validation of data consistency.

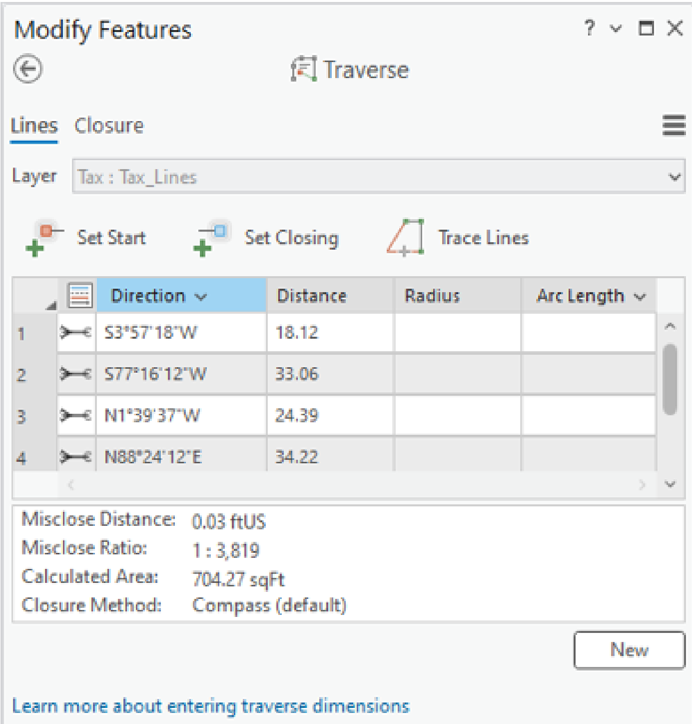

Figure 3: Traverse showing misclose in ArcGIS Pro

A misclose value is also a simple way for a machine-based data consistency check after COGO has been extracted by an OCR process.

Common data entry mistakes

Manual data entry is prone to user error. Common mistakes include entering the wrong digit, skipping a course, and entering the wrong quadrant for the direction. Any such mistake is likely to cause a large misclose.

Labour intensive

Manual data entry is tedious and labour intensive. Some deeds might contain dozens of courses, and the process can take hours. Finding a mistake and correcting it also takes a long time, and technicians often use rulers and highlighters on physical paper copies to try to minimize their data entry mistakes.

COGO READER

OCR technology can be applied to the scanned images of legal deeds to convert them into text. COGO dimensions can be extracted from the text and used to create line features that describe a parcel's boundaries. This new capability in ArcGIS Pro will be called COGO Reader and will be part of the parcel fabric technology.

It comprises four components: a display view of the source document, a display view of the extracted text, a display view of the extracted COGO dimensions, and a map preview of the resulting lines.

Motivations for creating COGO Reader

COGO Reader reduces the time it takes to enter a parcel traverse from a deed description. A process that might take an hour using the conventional methods can be reduced to minutes. The productivity gains allow more time for map technicians to focus on quality checks and other elements of their record-driven workflows. Even with a 50% success rate in the OCR, there are benefits in the reduced amount of manual data entry.

Implementation

The COGO Reader user interface allows for an interactive, *person-in-the-loop* workflow where the quality of the data extraction can be evaluated and fixed before adding lines to the map.

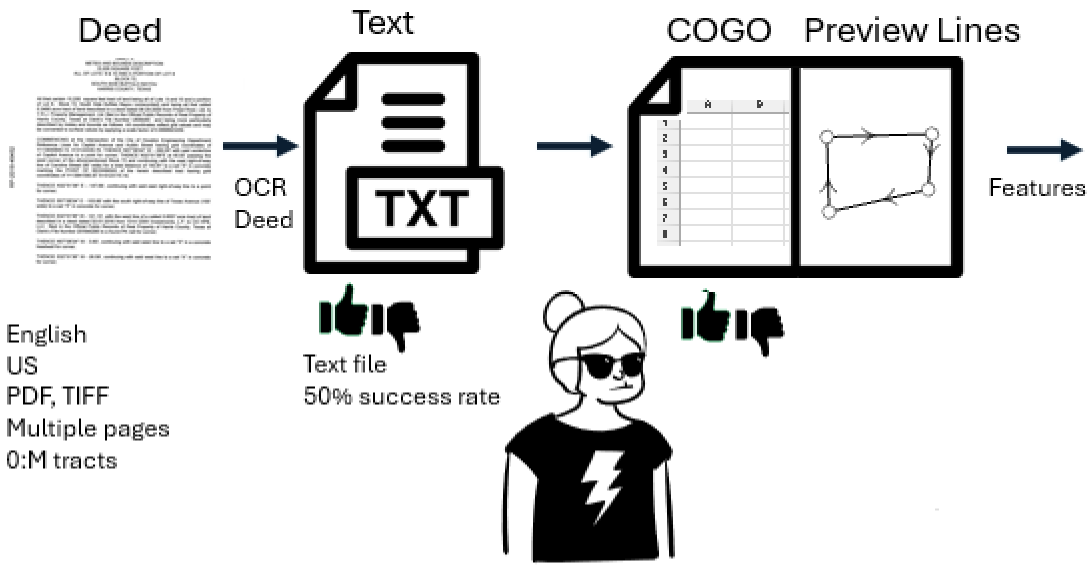

Figure 4: Extracting dimensions from deeds using OCR

The types of problems that may occur include poor OCR or incorrect COGO extraction. Since some documents are poorly scanned it is common for the OCR process to fail at extracting the text properly. For example, a degree symbol ° may have been interpreted as a '0', or a watermark can obscure the original image. The tool can calculate the misclose and notify the user when further review is necessary before lines are added to the map.

In these cases, the technician compares the source document with the extracted text. When the operator selects a COGO measurement, the corresponding extracted text is highlighted in the user interface.

The map preview helps the user identify any blunders by viewing the lines created by the extracted dimensions. Selecting the erroneous dimensions highlights the extracted text in the source document for a quick and easy review. For complicated deeds that produce partial results, the user can also send the results to the traverse tool in ArcGIS Pro and use them to finish the data entry of the traverse.

Can AI chatbots be used instead?

Artificial Intelligence (AI) and Large Language Models (Large Language Models) do use OCR to extract COGO dimensions from scanned deeds. However, the current AI models are not capable of reliably interpreting the specialized language used in textual deed descriptions.

There are also variety of ways to record the same description. Differences can occur based on the date of the document, with older documents using different language.

Differences can also occur in language conventions used in deeds from different regions and jurisdictions. These variations often cause commercial AI assistants to hallucinate. AI hallucination refers to a phenomenon where an artificial intelligence model, such as a LLM or a computer vision system, generates outputs that are incorrect or nonsensical but presented as if they are accurate. (Wikipedia Reference 3)

Another disadvantage of using AI chatbots is that they do not let you review and fix the intermediate extracted text before they attempt to extract the COGO dimensions. There is no *person-in-loop* to review and correct mistakes.

Chatbots also upload and process the documents in the cloud, and this can take significant processing time since documents are often large. Some organizations are also reluctant to process this data outside of their organization. The effort required to understand where the process has failed can often take longer than entering the data manually.

In the future, it might be possible to use reliable resources from COGO Reader to train a model dedicated to fixing OCR mistakes and to accurately interpret a wide variety of deed descriptions.

ROAD AHEAD

The COGO Reader technology will continue to evolve through the feedback received by users on the Esri community ideas board. (Esri Community Reference). Future work may include the extraction of additional information such as the record document name, recording date, point descriptions and parcel name.

Esri anticipates the development of training models that are evolving in sophistication and able to interpret ever more complex and varied description types.

The use of AI can help to enhance the image before OCR, such as the removal of watermarks and noise in the image that obscures readable text.

AI can also assist in extracting COGO values from survey plans, and other diagrammatic representations of parcels. The challenges in accomplishing this include interpreting directions and distances placed at angles on the page to match the orientation of the lines that they annotate.

CONCLUSIONS

Emerging technologies, particularly advancements in artificial intelligence (AI), are accelerating the demand for highly efficient and adaptable cadastral systems. Esri is actively working on AI assistants in ArcGIS Pro to perform common operations and reduce the learning curve for new and inexperienced users.

Automated quality assurance and efficient cadastral data ingestion can significantly reduce processing time. Optical Character Recognition (OCR) technology can assist in extracting dimensions from deeds, greatly reducing hours spent on tedious data entry.

With OCR technology, scanned documents can be automatically processed in the parcel fabric. This paper presents the parcel fabric technology called COGO Reader that will be available in ArcGIS Pro. This tool automatically creates a parcel traverse from a scanned deed instead of requiring the manual data entry of dimensions. This greatly improves the rate at which new parcel data can be added to the parcel fabric.

REFERENCES

Esri Community Reference:
https://community.esri.com/t5/arcgis-parcel-fabric-ideas/idb-p/arcgis-parcel-fabric-ideas

Esri Documentation Reference 1:
https://pro.arcgis.com/en/pro-app/latest/help/editing/introduction-to-cogo.htm

Esri Documentation Reference 2:
https://pro.arcgis.com/en/pro-app/latest/help/editing/introduction-to-ground-to-grid.htm

Wikipedia Reference 1:
https://en.wikipedia.org/wiki/System_of_record)

Wikipedia Reference 2:
https://www.indeed.com/career-advice/career-development/system-of-record

Wikipedia Reference 3:
https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)

CONTACTS

Amir Bar-Maor
Environmental Systems Research Institute
Wholesale building, Weena 695 (B2 - 036),
3013 AM ROT TEMP
Rotterdam
NETHERLANDS
Tel. +31 (0)10 217 07 00
Email:

Christine Leslie
Environmental Systems Research Institute
380 New York Street
Redlands
USA
Tel. +1 909 793 2853
Email: cleslie@esri.com
Web site: www.esri.com

Tim Hodson
Environmental Systems Research Institute
111 Market Street NW, Suite 250
Olympia
USA
Tel. +1 360 754 4727
Email: thodson@esri.com
Web site: www.esri.com