WHITE PAPER

INTELLIGENT SPATIAL DATA TRANSFORMATION

USING DIGITAL TO AUTOMATE GIS FEATURE EXTRACTIONS

Abstract

Recent advances in spatial data technologies have greatly increased the quantum of data available for geographic information systems (GIS), creating new opportunities as well as data management challenges for asset-centric industries. This POV examines the issues involved as well as the benefits of using digital technologies to automatically extract feature data from GIS data sources.

Infosys®
Navigate your next

The growing popularity of GIS

Recent analyst research reports indicate that most CXO/ CIOs in utilities, energy, mining, natural resources, and other asset-centric industries are increasingly leveraging spatial data to manage their business operations better – right from planning and fulfillment to assurance.

However, with the faster deployment of operational data becoming essential for maximizing opportunities, these industry leaders are especially keen to optimize and modernize their use of technology to better leverage their decision support systems.

Thus, geographic information systems (GIS) are steadily transforming from mere standalone tools to real-time operation support systems, with most organizations keen to deploy them for their many benefits including faster go-to-market and reduced cost of operations.

Why traditional processing is a challenge

Asset-centric industries use GIS systems to meet their needs for spatial analysis, network planning and design, geotechnical/ geological operations, and constructions. These systems deal with vectors and raster map data of various formats, at scale, and need the utmost importance given to quality during pre- and post-transformation activities.

When it comes to topographic data, traditional methods for mapping or extracting the features from the source datasets — either single or disparate — is considered as monotonous, having a long turnaround time, and being of a primitive in nature in the GIS field. Also, though traditional data conversion methods are still the best for addressing the various challenges in interpreting complex geometry diagrams, reducing rework, and ensuring quality outcomes, they involve exclusive operational set ups involving large production workforce deployments and substantial infrastructure.

Because recent times have seen revolutions in gathering spatial data – such as through point cloud data, drone images, Internet of Things (IOT) data, and earth observatory systems – there has been massive growth in data volume and scale. Thus, effective data management is now a serious challenge for many organizations operating with traditional methods, and they are now recognizing their need for automated interventions in acquiring, manipulating, analyzing, and presenting topographic data.

The shining promise of digital in spatial data management

Industry experts have been opining that the smarter option for leveraging all the spatial big data available is to incorporate innovative digital technologies, such as analytics augmented by artificial intelligence (AI) and/ or machine learning (ML). These technologies are already helping enterprises to better understand the potential of their data for large-scale planning and operations.

Resolving data management challenges by using reusable digital components and innovative architecture is also helping to bring about service differentiation through newer offerings such as location-based services centered around geospatial insights. Thus manual, rudimentary, and traditional approaches in GIS data are being overtaken by the new value proposition of augmentation of traditional methods by complementary new digital technologies.

The digitization or data conversion of the traditional map update is usually either by vectorization or photogrammetry/ image processing. Recently however, the newer AI/ML solutions have made automated feature extraction possible. These solutions enable the identification of geographic features and their outlines in remote-sensing imageries, scanned raster, etc., through post-processing technology that enhance feature definition, often by increasing feature-to-background contrast or using pattern recognition software [1]. Several of these solutions use off-the-shelf products and platforms and can involve either full or partial automation to reclassify the image data using feature extraction, segmentation, matching, classification, or simulation.

Moreover, advanced optical character recognition (OCR) and raster to vector (R2V) conversion technologies to digitize large-scale maps and data obtained from terrestrial surveys, reinforce the agility of digital and its ability to deliver continuous incremental value in managing large volumes of data.

[1] http://wiki.gis.com/wiki/index.php/GIS_Glossary

Building workflows for automation

Advance imaging, LiDAR mapping, and drone surveys have rapidly emerged as quick and flexible acquisition and mapping systems for most real-time applications including in mining, cadastral, roof measures, and coastal areas. Though still at a nascent stage, it has been exploited to its fullest potential in the mining and utilities space.

Because spatial data today comes from these different source systems, there are specific limitations to completely automate feature extraction, especially from drone or LiDAR imagery. Even partial automation will contain some errors, data loss, and discrepancies, and thus manual corrections cannot be fully eliminated.

For an OCR platform, accuracy is measured in term of transforming the geometry objects as-is to the target format, however, feature classes are not maintained when GIS data is leveraged. As the OCR figure below shows, the line object is moved to a line object but not as a polyline, and not continuous, snapped, or topologically connected. Therefore, accuracy is always debatable on how GIS data is converted to the feature class against the raw data.

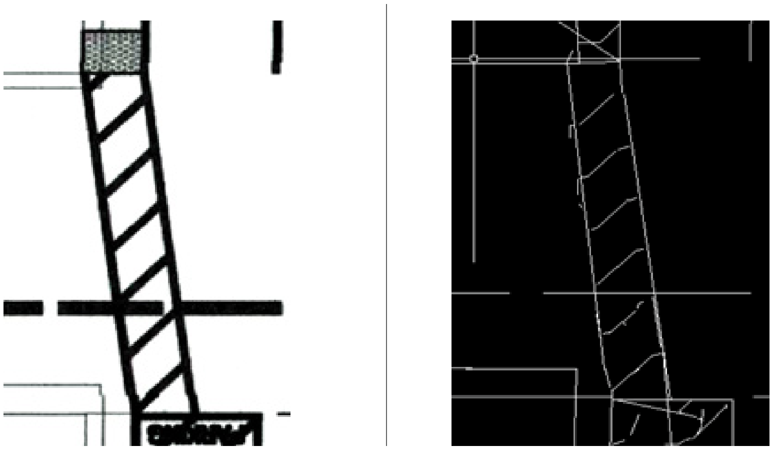


Fig 1: Pre-Post conversion challenges faced in Data Operations environment

Thus, while leveraging AI/ML and robotic process automation brings in efficiencies in data management, considerable manual interventions are still required for post-production aspects. These include pre/post processing, image segmentation, line extraction, contour generation, all of which are necessary to realize the advantages of becoming data-driven. To eliminate redundant legacy methods used for these areas and drive automation further, organizations need to adopt additional applications including object detection and data extraction (OCR like) and AI/ML based annotations.

This can be done through evaluating various approaches for feature extraction based on various application fields and synthesizing these into a hypothetical or real-time workflow.

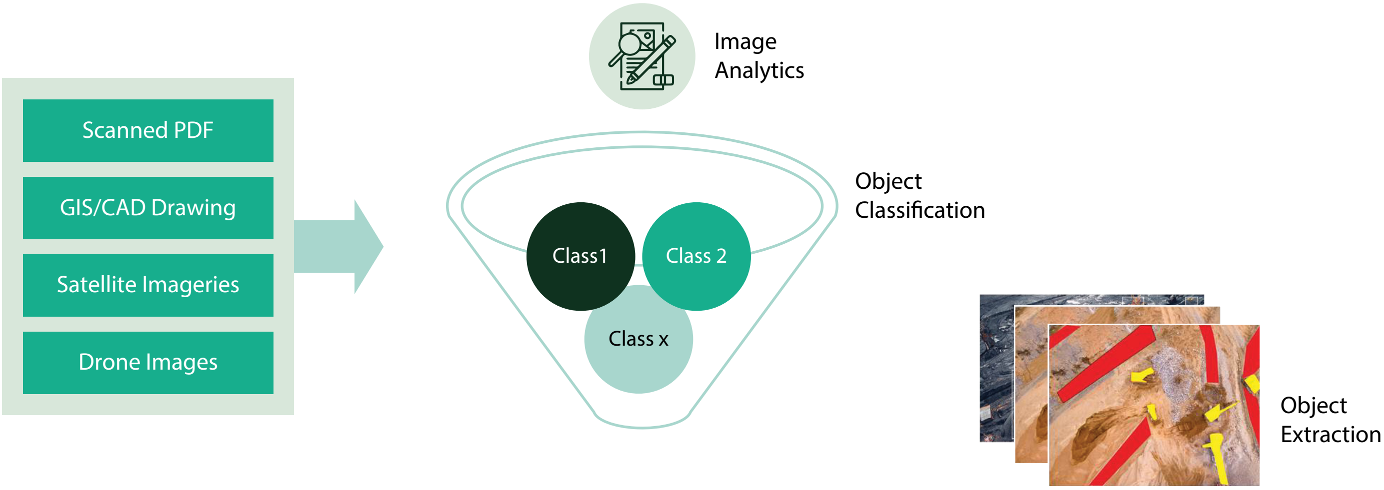


Fig 2: GIS based Object classification & Image extraction: Illustrative workflow for a data extraction process

Training the bots for accuracy

With respect to location data, trained bots would adopt similar interpretation methods as those used by humans to identify and locate target assets in the drawings, which could include anomaly features, printed and handwritten features, as well as symbols and icons. One of the methods involves building bounding boxes around the targeted assets and then allowing bots for focused extraction so that accuracy is not impacted by the adjacent objects and assets.

In a typical application scenario, as shown in the below figures, any software technique includes building training data sets involving thousands of annotated images for ML platforms that performs the human like interference for many utility applications such as vegetation management, autonomous fleets, railroads cracks and tracks, or cargo movements.

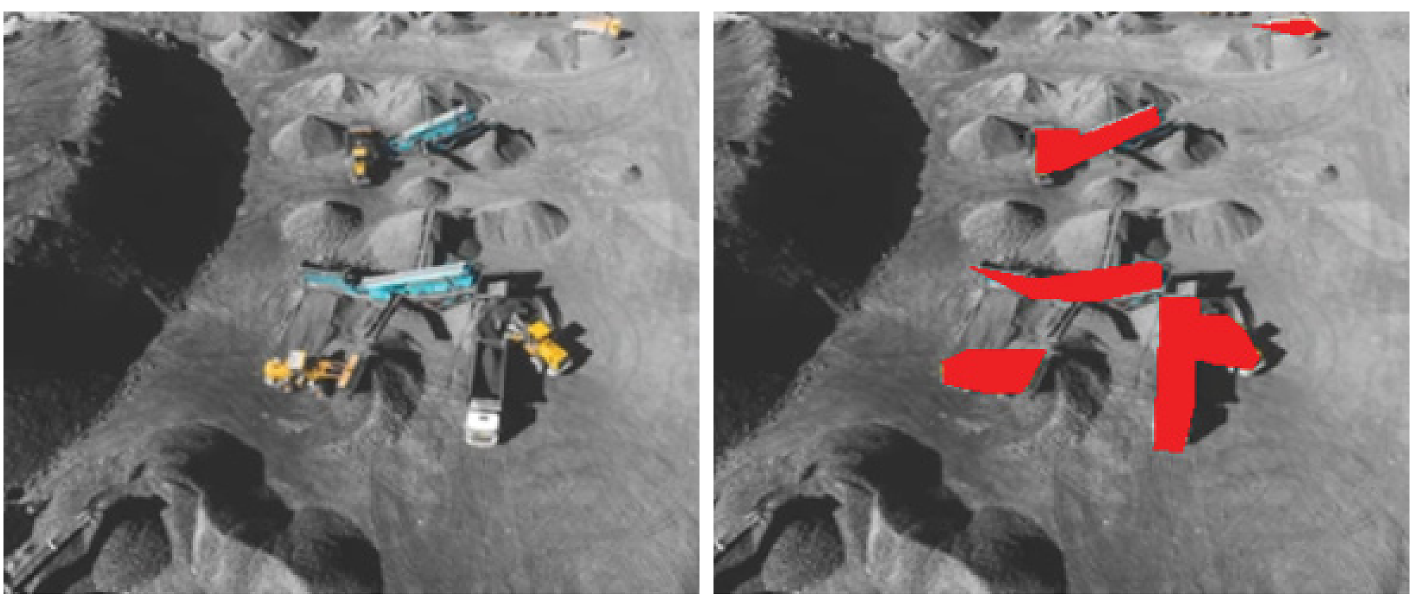


Fig 3: Building training datasets (annotated Images) for ML solution (Imagery source - https://www.shutterstock.com/)

Like above, there are other aspects involved in managing spatial data conversion with AI/ML methods, very similar to OCR technologies, but a robust AI centric engine assessing more volumes of sample data can be used to build hyper-automated solutions.

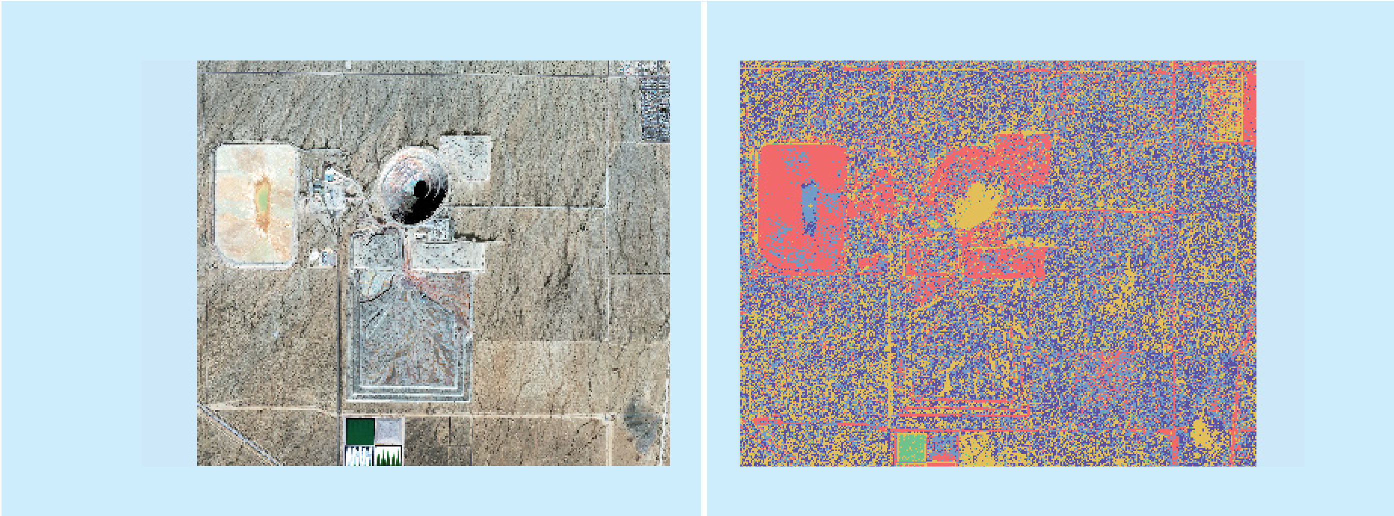


Fig 4: Image processing and classification post applying analytics logic
(Image source: https://effigis.com/en/solutions/satellite-images/satellite-image-samples/)

Calculating success outcomes

While training bots to attain data accuracy, the key challenges currently are getting outputs compatible with GIS and domain centric formats. As shown in the below diagram (fig), a bot understands the source objects as geometry objects and converts to the polylines, however, the lines are discreet and topologically incorrect data and auto-extraction was able to fulfill only 10% to 20% of the accuracy required. Even more challenging is extracting hand-drawn or user drawn info from the source data.

However, based on prevailing experience, involving both automated extraction and manual involvement of engineers to interpret the data ingested and processed by the BOT algorithms can yield successful outcomes.

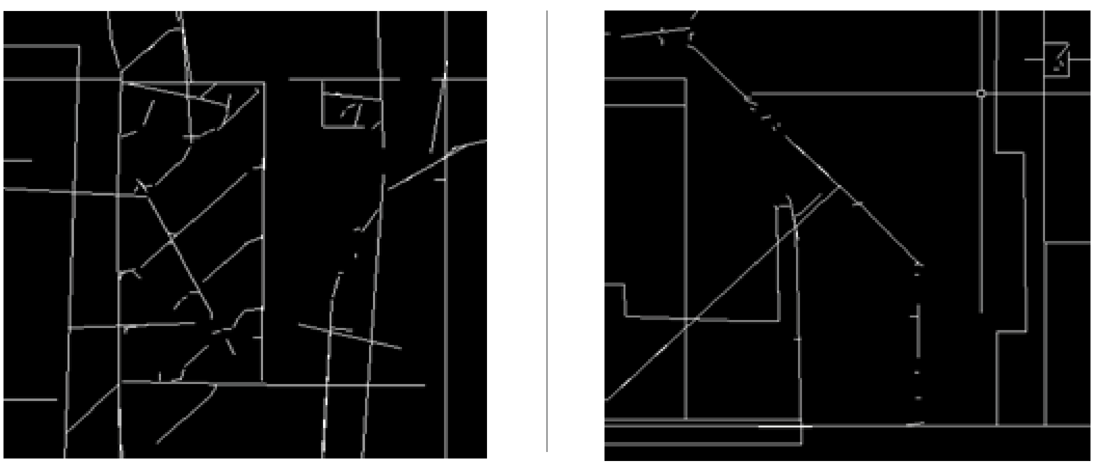
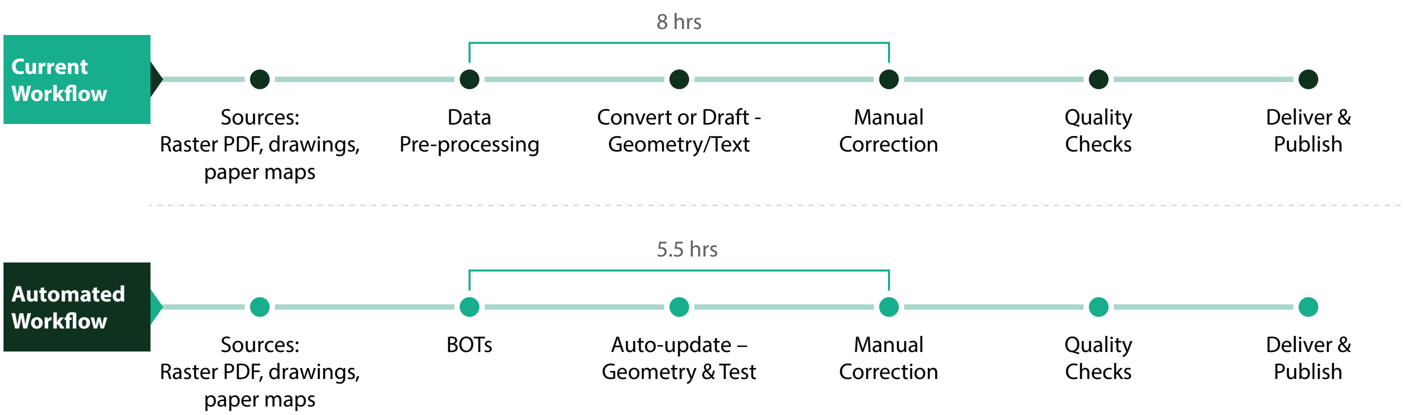


Fig 5: How auto-extraction wasn't able to fulfill geometrical accuracy

Thus, recently, a utility company with traditional data operations needed to migrate and update data on overhead and underground inventory/assets from legacy formats into the GIS format using custom AutoCAD/ESRI or other tools. For this, pre- and post-asset inventory changes were manually identified. Each point, line, text, and coordinates needed to be manually compared and changes identified, which was error prone and time-consuming.

To gain efficiency, the company adopted simple bot tools, with the ability to run multiple transaction at a single point and converted the legacy assets to GIS format with a certain level of accuracy. It then used manual intervention to correct remaining errors and the method worked successfully as the company's data source types were uniform.

Process Overview



Leading towards GIS Maturity

Today, telecommunications, energy, and utility organizations are considered primary users of GIS to manage mission critical applications, including planning, field optimization, order placement, fulfillment, and assurance. Spatial data is imbibed into the value chain of above industries and is part of their IT ecosystem, enabling the achievement of their business and organization goals.

Since the quantum of data has been increasing by leaps and bounds, these organization cannot now resolve their problems in data management and operations, unless they increasingly deploy the latest digital technologies discussed above. While the approaches may be different, adapting to the changing data needs of the organization is of prime importance. Managing assets better by progressing through the phases of GIS maturity, will deliver manifold benefits, such as improved efficiency and productivity, and reduced costs, that cut across the value chain of planning, design, supply, and delivery.

Further, organizations gaining experience with leveraging AI-centric object extractions tools for their spatial needs will adapt and respond more easily to future opportunities and lead the pack in driving further innovation.

References

1. https://www.thedrum.com/news/2020/02/25/data-the-lifeblood-your-business-s-why-you-need-data-strategy-its-heart
2. https://link.springer.com/article/10.1186/s12940-018-0386-x
3. https://www.techerati.com/features-hub/opinions/artificial-intelligence-is-set-to-transform-data-management/
4. https://www.precisely.com/blog/data-quality/machine-learning-improve-data-quality
5. https://www.infosys.com/insights/ai-automation/data-driven-enterprise.html
6. https://link.springer.com/article/10.1007/s41064-020-00102-3
7. https://www.geospatialworld.net/blogs/why-ai-for-gis-is-still-under-the-radar/
8. https://ehjournal.biomedcentral.com/articles/10.1186/s12940-018-0386-x
9. https://www.europeanbusinessreview.com/three-trends-driving-the-geospatial-ai-revolution/
10. http://news.aag.org/2019/03/ai-and-gis-finally-delivering-on-the-promise/
11. https://geobuiz.com/geobuiz-report-2019/
12. https://www.isprs.org/proceedings/XXXIII/congress/part2/241_XXXIII-part2.pdf
13. https://support.esri.com/en/other-resources/gis-dictionary/term/02fc5f0b-23e1-41e6-8cd2-d3ced57e71f1
14. file:///C:/Users/pradeep_kishore/Downloads/remotesensing-08-00689%20(1).pdf
15. https://proceedings.esri.com/library/userconf/fed18/papers/fed-060.pdf
16. https://link.springer.com/chapter/10.1007/978-3-540-77058-9_8
17. https://support.esri.com/en/other-resources/gis-dictionary/term/02fc5f0b-23e1-41e6-8cd2-d3ced57e71f1
18. https://www.mdpi.com/2072-4292/8/8/689/htm

About Author

Dr. V. Pradeep Kishore

Solution Design Head, Infosys BPM

Dr. Pradeep leads the GIS data practice in Infosys BPM. He has close to 22 years of experience in the field of GIS, telecom OSS/BSS, mobility and telematics domain, and is involved in the consulting, solution architecture, program/project management, portfolio management for Indian and overseas clients (US/Middle East). In Infosys BPM, he is responsible for leading the GTM strategy, presales, consulting, COE accountabilities - creating/owning business modeling/framework and contribution to all lines of GIS business/new wins.

Pradeep holds a Doctorate Degree in Disaster Management (landslides prediction) using GIS/remote sensing and has done his post-graduation in Applied Geology.

Karthigaiselvan R

Practice lead, Digital Interactive Services, Infosys BPM

Karthi has over 16 years of experience across the geospatial industry. He is involved in electric utilities, water utilities, disaster management and smart city project management, solution architecture, and geospatial products support. At Infosys BPM, he is responsible for delivery support, consulting, and managing geospatial projects.

Karthi has done his post-graduation in Remote Sensing and under-graduation in Electrical and Electronics Engineering.

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next