

DEVELOPMENT OF AN OSINT TOOL FOR GEOLOCATION ANALYSIS BASED ON COORDINATE TRIANGULATION

Ivan Drankov, Yoana Konstantinova

University of Mining and Geology “St. Ivan Rilski”, 1700 Sofia; E-mail: ivan.drankov@mgu.bg; y.y.konstantinova@gmail.com

ABSTRACT. This paper presents the development of an OSINT tool for estimating a target's geographic location through coordinate triangulation from image metadata. The tool extracts GPS data from digital images, normalises and validates the collected coordinates, and applies spatial analysis methods, such as radius-based intersection, weighted centroid calculation, and Delaunay triangulation. The results are visualised on an interactive map, supporting automated geolocation analysis in OSINT investigations.

Keywords: OSINT, geolocation, triangulation, metadata, digital forensics.

РАЗРАБОТВАНЕ НА OSINT ИНСТРУМЕНТ ЗА ГЕОЛОКАЦИОНЕН АНАЛИЗ, БАЗИРАН НА КООРДИНАТНА ТРИАНГУЛАЦИЯ

Иван Дрънков, Йоана Константинова

Минно-геоложки университет „Св. Иван Рилски“, 1700 София

РЕЗЮМЕ. Настоящата публикация представя разработването на OSINT инструмент за определяне на географското местоположение на обект чрез координатна триангулация на база метаданни от изображения. Инструментът извлича GPS данни от цифрови изображения, нормализира и валидира събраните координати и прилага методи за пространствен анализ като анализ на пресичащи се радиуси, изчисляване на претеглен центроид и Delaunay триангулация. Получените резултати се визуализират върху интерактивна карта, подпомагаща автоматизирания геолокационен анализ в рамките на OSINT разследванията.

Ключови думи: OSINT, геолокация, триангулация, метаданни, дигитална криминалистика.

Introduction

The increasing volume of publicly available information has significantly expanded the possibilities for intelligence gathering through open-source intelligence techniques. Modern smartphones, cameras, and mobile devices generate large amounts of metadata that may contain valuable information regarding the location, time, and context of captured digital content. Geolocation through metadata analysis has become an important component of cyber investigations, law enforcement activities, digital forensics, and threat intelligence operations. In many cases, investigators are required to determine the approximate location of an individual, based on multiple photographs from different sources. Manual processing of such information is time-consuming and often requires specialised knowledge and software tools. The objective of this research is to develop an automated OSINT tool capable of extracting geolocation metadata from digital images and estimating the most probable location of a target through coordinate triangulation and spatial analysis.

Metadata Extraction

Location data can be extracted from multiple different sources, e.g. GPS data, metadata, IP address, Apps, etc. This research will focus on metadata solely from camera shots with attached metadata from a built-in GPS module on the device which took the picture. “Direct data sources, such as GPS, conventional geodetic measurements, and process monitoring, are described as data received from devices specifically designed for data collection” (Dimitrov et al., 2021).

This ensures the most accurate data for analysing near exact location. “Taking into account all data points, the maximum radial error for the Standard and Experimental phones was 11'533 m and 27'259 m” (Ryser et al., 2024).

The first step of the tool is to extract useful metadata from the input media files. In this implementation, the main focus is on GPS metadata stored in digital images. The tool scans a selected folder and processes supported file formats, such as JPG, JPEG, HEIC and other common image formats.

For GPS information like latitude, longitude and accuracy, date and time of creation, etc., the proposed system extracts this information if it is present. The feature is implemented with some help from Exif Tool due to the range of supported image formats and structured output of metadata. Python EXIF parser is used as a redundant feature for standard image files.

Once metadata is extracted, the coordinates are normalised into decimal degree format, as the GPS data might be stored in different formats, like in degree, minute, and second degrees. After normalising, the tool validates the information it extracts. Files which do not contain GPS Data, files which have nonsensical coordinates and files with duplicate coordinates in pairs are rejected from the core analysis.

Rejected files are not deleted from the workflow. They are stored in a separate rejected log together with the reason for rejection. This makes the analysis traceable and allows the analyst to check which files were accepted and which were excluded.

The accepted records are converted into evidence points. Each point contains the file name, coordinates, timestamp, extraction method, and uncertainty values. These evidence points are then used in the next stages of the tool for clustering, coordinate hypothesis generation and visualisation.

Implementation

In the case of Open-Source Intelligence, locations are mostly used to locate an object or a threat. The focus of this tool is to locate objects and predict movement, based on source points defined by the extracted data. The source points are used as a basis for the analysis. Choosing the right method depends on the data. At the moment, this is done by the operator. Humans have a good understanding of spatial and temporal data and can quickly determine the context. In further research, an AI with context knowledge will be useful for determining the best method.

Method 1: Triangulation

When an object's location is unknown, triangulation is the best method for estimating it. This is often seen in pictures taken near the object's home location. Those pictures are mostly clustered in a region with the same date of creation or very close to each other. The method used for finding locations is Delaunay triangulation. Here, the algorithm is weighted by the operator based on the context of the data. Each observed location point is weighted by a human representing the confidence, frequency, or reliability of the observation. This allows more reliable or more frequently observed points to have a stronger influence on the result.

Weighted Delaunay Triangulation

Each point has a human-selected weight that controls its influence on the triangulation.

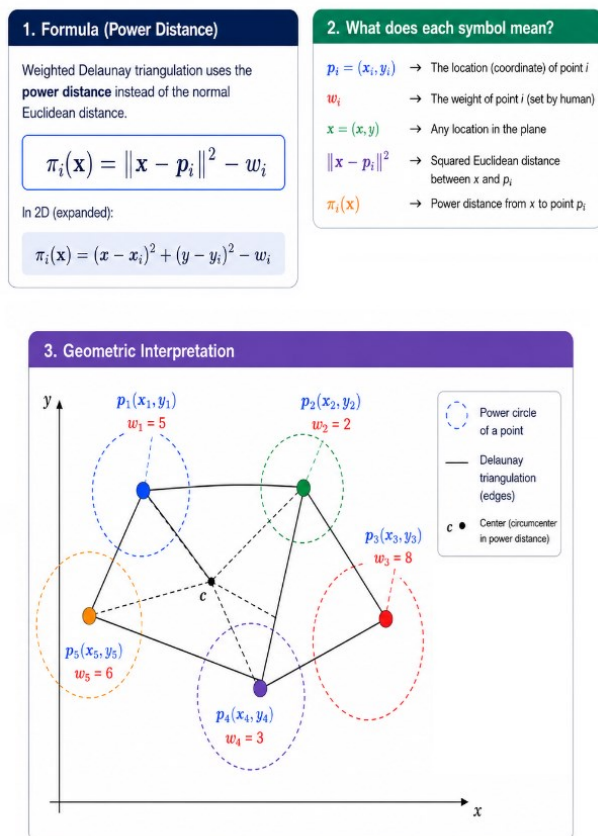


Fig 1. Delaunay triangulations

“Regular triangulations are generalised Delaunay triangulations. A regular triangulation is defined for a set of weighted points. Each weighted point can be considered as a

sphere whose square radius is equal to the weight. Then, the regular triangulation of a set of weighted points can be defined as the dual of the power diagram of the associated spheres.” (Boissonnat et al., 2002). The specific method used is shown in Fig 1.

Method 2: Last known location

The last known location is calculated by ordering all accepted GPS records by their creation time and selecting the most recent valid point. The result shows the latest available latitude, longitude, timestamp, and source file.

This point is useful as a reference in the analysis, but it does not prove the current location of the subject. It only represents the latest location recorded in the available metadata. For this reason, it is presented separately from the main coordinate hypothesis and from the path prediction layer.

Last Known Location Analysis

Observed locations are ordered by date and time, then connected to form a chronological chain of events.

1. Chronological Path

To analyse the last known location, observed points are sorted by timestamp and connected in chronological order. The sequence shows the movement path and the most recent point is treated as the last known location.

$$p_1 \rightarrow p_2 \rightarrow p_3 \rightarrow \dots \rightarrow p_n$$

$$t_1 < t_2 < t_3 < \dots < t_n$$

The point p_n is the most recent observation.

2. Geometric Interpretation

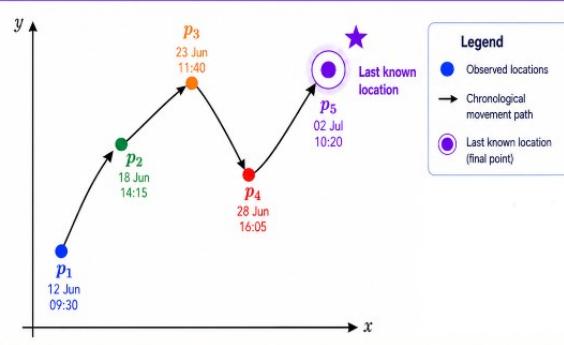


Fig 2. Last known location

Method 3: Path prediction

Path prediction in the scope of the proposed tool works as a movement hypothesis module which is based on chronological geolocation evidence. It is essential to notice that “prediction” does not mean the forecasting of movement at all, and refers to the generation of a probable movement pattern based on the GPS data and timestamps. Hence, the output of the tool should be seen as a contextual layer of analysis, not the proof of the exact route.

As the step following metadata extraction and validation, all the geolocation records with information on timestamps are sorted by date. Further, within each date group, the records are sorted in chronological order according to the extracted time of creation. Then, consecutive points are linked to create a movement segment. Each movement segment is described by the following four elements – starting coordinate, ending coordinate, time interval, and calculated distance.

Path Prediction Analysis

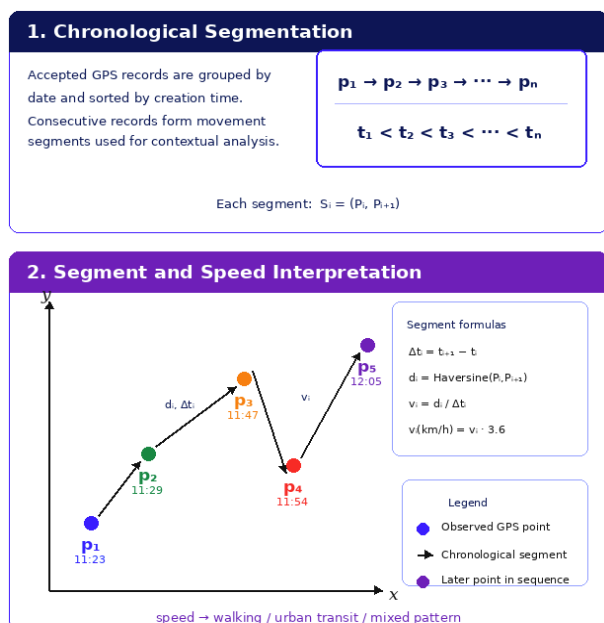


Fig 3. Path prediction

The distance between the two consecutive points is measured via the Haversine formula, which gives the estimated great-circle distance between two latitude/longitude pairs on the Earth's surface. "Haversine is a popular formula used to find the distance between two points. But no research proves the number of formulae to find which distance is the "best."" (Maria et al., 2020).

$$D = 2r \arcsin \sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)}$$

For the case where two points are presented as P1 (lat1, lon1) and P2 (lat2, lon2), the calculated distance is divided by the difference between the two timestamps in order to estimate the average speed for the particular segment. Speed is measured in kilometres per hour and reflects the possible way of movement.

The tool assigns each segment to a particular movement hypothesis depending on its speed. Thus, low speed is considered to be walking-plausible, medium speed may refer to mixed or uncertain movements, and high speed may represent urban transit or motorised transport. The classification categories are heuristic and cannot be used as proof of the actual way of movement. The idea is that they should be used as an assistance to the analyst in checking if the chronological order of points is spatially and temporally plausible.

The total travelled distance, total duration, speed values on the segment level, and a dominating pattern are calculated for each daily sequence. If there is more than one type of movement in the sequence, it is referred to as a mixed-day pattern. This is helpful in case the set of photographs includes both clustered points and distant points from the same day. With the help of the path-prediction module, it is explained why some coordinates can be seen as belonging to the primary geolocation cluster, and others are considered secondary clusters or contextual evidence.

The movement hypothesis is represented as an optional temporal layer on the interactive map. It is separated intentionally from the final coordinate hypothesis. The final coordinate hypothesis is produced via the spatial clustering and triangulation workflow, whereas the path-prediction layer shows the chronological context. It helps avoid the situation where the system treats every valid coordinate as evidence of a single location, helping the analyst identify a stable location cluster or movement between different locations.

During the experimental run, the tool found a daily sequence containing nine accepted geolocation records for one date. The total length of the sequence was about 7.64 km, and the dominating pattern was classified as mixed. This means that the records from the day cannot be seen as evidence of a single stationary location, but of the combination of movement and clustered evidence. Hence, the path-prediction layer assists in interpreting the coordinate results by showing the temporal transitions between spatial clusters.

There are several limitations related to this module. First, the metadata timestamps may be missing, modified, influenced by the time zone setting, or inconsistent across devices. In addition, the route calculation is based on the straight-line distances between the consecutive GPS points and not on the reconstruction of the real path through the streets. So, the path prediction should be regarded as the movement hypothesis and not the confirmation of the physical movement.

In future research, deep neural networks may be implemented to improve the path prediction module. Similar approaches have been applied for predictive analysis, where operational data are used to train different types of deep neural networks in order to predict a useful indicator. This suggests that, with enough chronological location data, a neural network could support the estimation of probable movement patterns and future locations (Arsova-Borisova and Hristov, 2021).

Results

The data in this research is collected with OSINT methods and is extracted from metadata. Publicly available sources are not subject to GDPR, but the results of the experiment may violate it. To ensure compliance with both ethical and moral laws, the results shown are only of the authors' data.

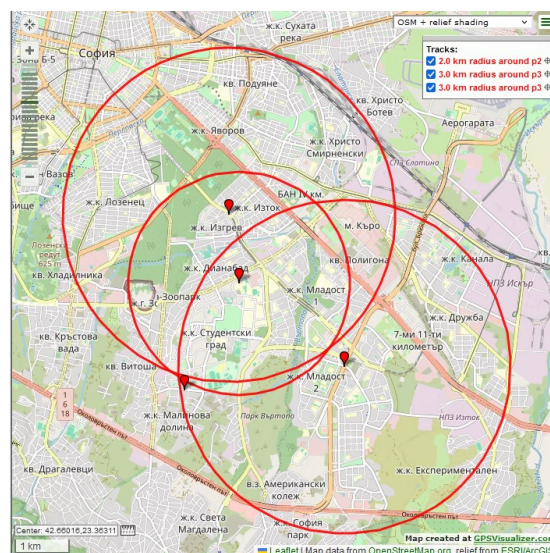


Fig 4. Method 1 triangulation

Triangulating results with real data shows the exact home address with about a 50-meter offset. This particular data belongs to one of the authors (I. Drankov). In this case, the experiment was made with 3 of the most occurring data points of the subject.

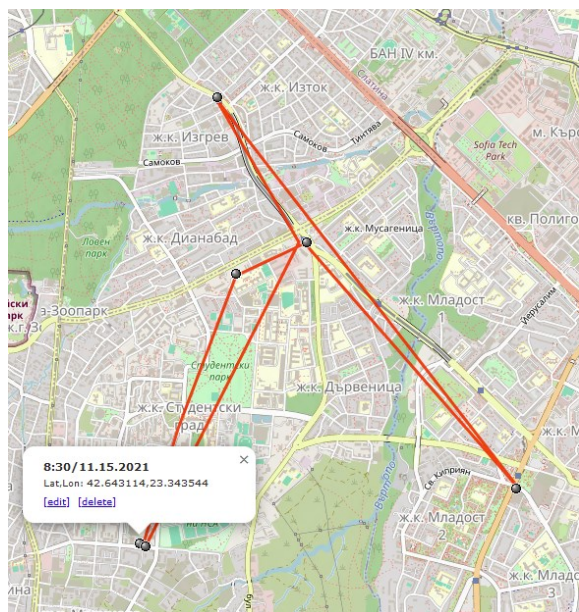


Fig 5. Last known location



Fig 6. Path prediction

The last known location method was tested by ordering the extracted location points chronologically. Each point represents a known position of the subject at a specific moment in time. By connecting these points in a sequence, the system creates a

chain of movement events. This allows the operator to understand the subject's movement history and identify the most recent known position. The method is useful when the goal is not to calculate a fixed home location, but to determine where the subject was last observed based on the available data.

The path prediction method uses the chronological chain of known locations as a basis for estimating possible movement. When two or more points are separated by a significant time interval, the system can analyse the direction of movement and the distance between the points. Based on this, a possible path between the known locations can be suggested. In cases where the last known location is available, the same logic can be used to estimate possible future locations. This does not provide an exact prediction, but it helps narrow down the area where the subject may be located or may move next.

The results show that location metadata can reveal highly sensitive information, even when the original data appears limited or fragmented. Repeated location points can expose important places such as a home address, while time-ordered points can show movement patterns and support path prediction. This confirms that OSINT-based location analysis can be effective for threat identification, but it also highlights the need for strict ethical limitations when processing personal geolocation data.

References

- Arsova-Borisova, K. & Hristov, V. (2021). Deep Neural Networks Applications in the Study of a Geological Indicator. *Sustainable Extraction and Processing of Raw Materials*, 2, pp. 25–27, doi: 10.5281/zenodo.5594817.
- Boissonnat, J. D., Devillers, O., Pion, S., Teillaud, M., & Yvinec, M. (2002). Triangulations in CGAL. *Computational Geometry: Theory and Applications*, 22, (1–3), pp. 5–19. doi: 10.1016/S0925-7721(01)00054-2.
- Dimitrov, A., Ivanova, N., Nesheva, R., & Yanev, N. (2021). Preparation of industrial data for implementation in Big Data. *Annual of the University of Mining and Geology “St. Ivan Rilski”*, 64, 151–154.
- Maria, E., Budiman, E., Haviluddin & Taruk, M. (2020). Measure distance locating nearest public facilities using Haversine and Euclidean Methods. *Journal of Physics: Conference Series*, 1450, p. 012080, doi: 10.1088/1742-6596/1450/1/012080.
- Ryser, E., Spichiger, H., & Jaquet-Chiffelle, D. O. (2024). Geotagging accuracy in smartphone photography. *Forensic Science International: Digital Investigation*, 50, Supplement, doi:10.1016/j.fsidi.2024.301813.