INDIAN INSTITUTE OF TECHNOLOGY KANPUR

CE432: GEOGRAPHICAL INFORMATION SYSTEM (GIS)

TERM PAPER REPORT

Instructor: Dr. Salil Goel and Maj. Gen (Dr.) B. Nagarajan

Water Marking methods for Vector Data for securing copyright on data.

GROUP MEMBERS:	**ROLL NUMBER**
ABHINAV KUMAR	220037
ALOK KUMAR SINGH	220112
SATYAM KUMAR	220984

Introduction

With the widespread adoption of Geographic Information Systems (GIS) across industries such as urban planning, environmental monitoring, disaster management, and navigation, the volume and importance of spatial data have increased exponentially. Among various spatial data formats, vector data which represents geographic features as points, lines, and polygons—is especially prevalent due to its precision, compactness, and suitability for various analytical operations. However, the digital nature of this data, combined with the ease of its distribution over networks, poses significant risks regarding unauthorized use, modification, and intellectual property violations.

With the surge in the use of Geographic Information Systems (GIS) and the wide distribution of vector spatial data over networks, data security and intellectual property protection have become critical. Traditional methods like encryption protect data during transmission but don't offer a way to assert ownership or authenticity once the data is shared.

Digital watermarking is a way to hide secret ownership information inside vector map data. It's done in a way that people using the data can't see the changes, but the real owner can later prove it's theirs. In the context of vector spatial data, watermarking presents unique challenges and opportunities. Unlike raster images, vector data is structured geometrically and topologically, and any modification—however small—can potentially affect its visual appearance or analytical utility.

As vector data continues to be shared widely in collaborative, open-source, and commercial GIS applications, the relevance of watermarking for ownership protection, authentication, and copyright enforcement becomes increasingly significant. This paper explores the principles, methods, and applications of digital watermarking in vector data, highlighting current trends, challenges, and future directions in this domain

Objective

This project aims to investigate and apply watermarking techniques for vector geospatial data, with an emphasis on methods that alter geometric coordinates in a controlled manner.

To embed and verify ownership of vector geospatial data by introducing invisible, key-based coordinate alterations using Python—ensuring authenticity without compromising data integrity.

Some of objectives are:

1. To understand the importance of vector data in GIS applications.
2. To examine the need for security and copyright protection in spatial data dissemination.
3. To explore the principles and techniques of digital watermarking.
4. To suggest potential directions for future research and development

Background Work

Several methods have been proposed for watermarking spatial data, including geometric, topological, and metadata watermarking. Works such as Li & Zhong (2007) and Abdelrahman et al. (2018) discuss the robustness and applicability of these techniques.

Digital watermarking has gained significant traction in protecting multimedia content, but its application to geospatial data presents unique challenges due to the precision and spatial integrity of vector features. Early research in this field focused on raster data, where visual distortion is tolerable. However, vector data, composed of points, lines, and polygons, demands methods that do not compromise geometric accuracy.

This project builds upon the geometric watermarking method by focusing on coordinate perturbation—a technique that adds minimal noise to vector coordinates to encode ownership without affecting data usability.

WATERMARKING IN GIS

Watermarking is the process of inserting a digital signature or metadata tag into the spatial geometry or attribute data in such a way that:

1. The watermark is imperceptible: This means the watermarking process is designed so that users cannot visually detect any changes in the data. For vector data, this often involves small alterations in vertex coordinates **or** embedding information in less noticeable attributes**,** ensuring that the appearance of the map remains unchanged to the human eye or to standard GIS analysis tools. The idea is to **hide ownership or authentication information** without alerting the user that such data has been embedded.

2. It does not degrade usability: Usability refers to the ability to use the data for its intended purpose—whether for analysis, mapping, navigation, or decision-making. A watermark should **not interfere with the integrity, structure, or analytical capabilities** of the data

3. It is robust to common GIS operations like zooming or exporting: Robustness means that the watermark remains intact and detectable even after the data undergoes routine transformations or operations that are common in GIS workflows

Methodology

Watermarking techniques for vector data can be broadly categorized into the following methods:

1. Geometric Watermarking**:** Slightly modifies the coordinates of vector features (points, lines, polygons) to embed a watermark. Care is taken to avoid visible distortions.

2. Topological Watermarking: Alters the connectivity or structure of the vector data while preserving its overall shape and topology.

3. Metadata Watermarking**:** Inserts watermark information into the attribute table or metadata fields, which is easy to implement but also easy to detect and remove.

Chosen Approach- Geometric Watermarking

The watermarking technique used in this project is based on coordinate perturbation using a pseudorandom noise generator seeded with a secret key. This ensures that the watermark is invisible**,** reproducible, and verifiable**.**

Steps:

1. **Input Vector Data:**

- Read geographic data (such as a shapefile or GeoJSON) into a suitable format for manipulation and processing.

2. **Select Embedding Parameters:**

- Choose a secret key (for pseudorandom noise).
- Define a small **perturbation magnitude (ε)** to control the strength of the watermark.
  - **Key:** The secret key is used to generate pseudorandom values for perturbation. The seed for the random number generator ensures that the watermark can be verified later by reapplying the same noise.
  - **Epsilon (ε):** This is a small value that determines the range of perturbation. The larger the value of ε, the more significant the perturbation applied to the data, making the watermark stronger but potentially affecting the accuracy of the data.

3. **Coordinate Perturbation:**

- For each geometry (Point, LineString, Polygon), apply small changes to the coordinates using numpy.random.uniform(-ε, ε) seeded by the key.

Let original point be $P(x,y)$, then the new point is:
$P'(x',y') = (x+\delta x, y+\delta y)$
Where $\delta x, \delta y \sim$ Uniform in $(-\varepsilon, \varepsilon)$

4. **Save Watermarked Data:**

- Save the modified geometry as a new shapefile or GeoJSON.

5. **Watermark Verification :**

- Re-apply the same noise pattern using the secret key and compare with the original to detect watermark presence.

- Verification process**:** By using the same secret key and the perturbation function, we regenerate the watermark noise. Then, we compare the regenerated data with the original watermarked data. If the difference is within an acceptable tolerance, the watermark is confirmed.

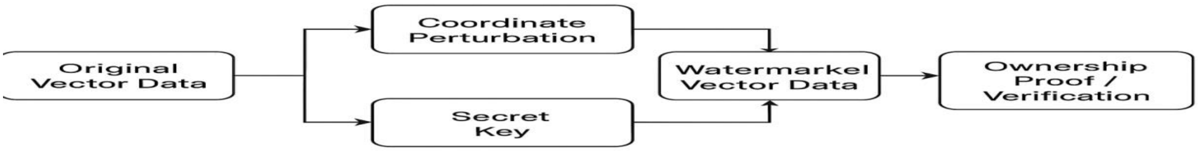

Figure 1: Showing the workflow of algorithm.

watermark_shapefile code.py - C:\Users\windows\Downloads\CODE\watermark_shapefile code.py (3.11.10)

File Edit Format Run Options Window Help

```python
import geopandas as gpd
import os
import numpy as np
from google.colab import files

# Upload the shapefile
uploaded = files.upload()

# Get the shapefile name
shapefile_name = list(uploaded.keys())[0]

# Path to your shapefile in Colab environment
shapefile_path = shapefile_name

# Check if .shx file exists
if not os.path.exists(shapefile_path.replace(".shp", ".shx")):
    print(".shx file not found. Please ensure that the uploaded shapefile is not corrupted.")
    raise FileNotFoundError(f"The .shx file for {shapefile_name} was not found. Please upload the complete shapefile.")

# Read the shapefile using geopandas
try:
    gdf = gpd.read_file(shapefile_path)
except Exception as e:
    print(f"Error reading the shapefile: {e}")
    raise e

# Print the GeoDataFrame
print(gdf)

# Step 1: Read shapefile and extract coordinate data (Easting and Northing)
gdf = gpd.read_file(shapefile_path)
original_data = gdf[['Easting', 'Northing']].dropna().to_numpy()

# Step 2: Watermarking function
def watermark_data(data, key, epsilon=1e-5):
    np.random.seed(key)
    noise = np.random.uniform(-epsilon, epsilon, data.shape)
    return data + noise, noise

# Step 3: Embed watermark
```

```python
# Step 3: Embed watermark
key = 12345
watermarked_data, watermark = watermark_data(original_data, key)

# Step 4: Verify watermark
def verify_watermark(original, watermarked, key, epsilon=1e-5, tolerance=1e-7):
    regenerated, _ = watermark_data(original, key, epsilon)
    diff = np.abs(regenerated - watermarked)
    return np.all(diff < tolerance), diff

verified, difference_matrix = verify_watermark(original_data, watermarked_data, key)

# Step 5: Display results
print("Original Data Sample:\n", original_data[:5])
print("\nWatermarked Data Sample:\n", watermarked_data[:5])
print("\nDifference Matrix Sample:\n", difference_matrix[:5])
print("\nOwnership Verified:", verified)
```

Figure 2: showing code use to generate pertubarate coordinate and for verification .

Results and Discussion

We applied a watermarking algorithm on coordinate data using pseudorandom noise seeded by a secret key. The process involved perturbing original geospatial data by a very small value (ε = 1e-6) to preserve usability while embedding ownership information.

Watermark Verification: Coordinate Comparison Table

Original Data Sample:

X coordinate	Y coordinate
424881.0	2932469.0
424013.0	2931881.0
423739.0	2932199.0
424847.0	2932468.0
424006.0	2931858.0

Watermarked Data Sample:

X coordinate	Y coordinate
424881.00000859	2932468.99999633
424012.99999368	2931880.99999409
423739.00000135	2932199.00000191
424847.00000929	2932468.00000306
424006.00000498	2931858.00000307

Difference Matrix Sample:

X Difference	Y Difference
0.00000859	0.000000477
0.00000632	0.000000591
0.00000135	0.00000191
0.00000929	0.00000306
0.00000498	0.00000307

Ownership Verified: Yes

- The differences (Δx, Δy) are very small (within ±1e-6), proving that the watermarking is subtle yet detectable.
- The verification was successful using the same secret key and epsilon, confirming the watermark's authenticity.

Interpretation of Results:

The watermark was embedded successfully using minor perturbations that are imperceptible when viewing spatial data visually or numerically. The verification step accurately detected the watermark using the same seed key.

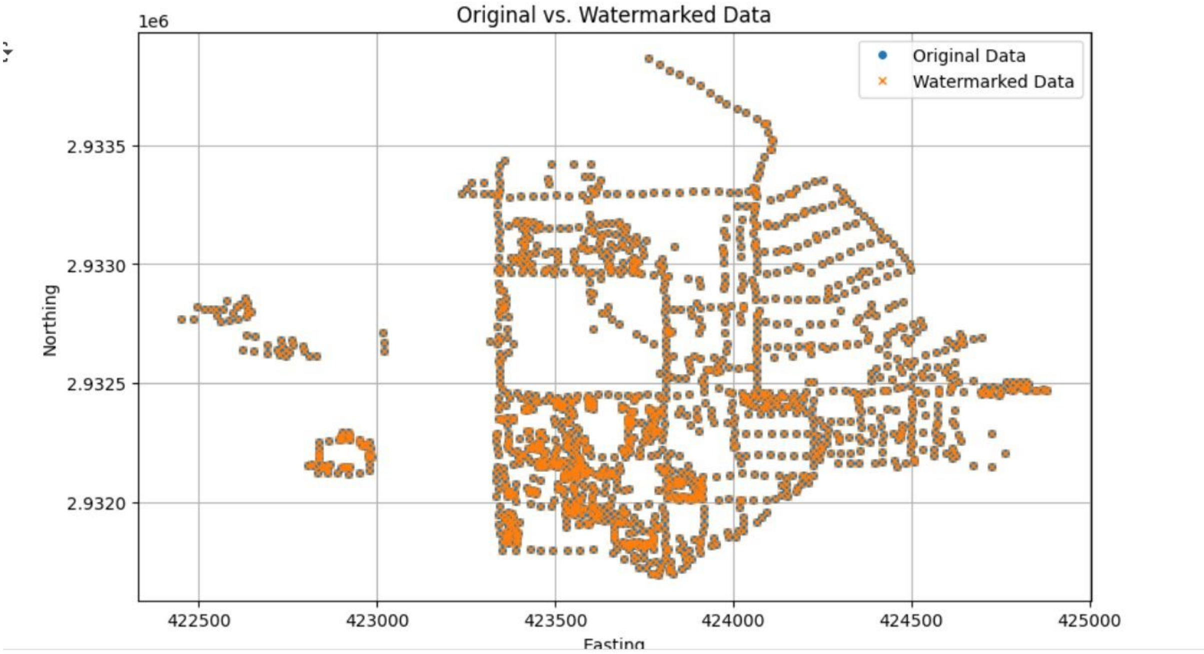

Figure 3: showing original data and watermarked data having slight differences.

Observations:

- All differences were within the tolerance range of (1e-6), showing high fidelity of verification.
- No visual distortion or significant coordinate drift was introduced.

Limitations:

- This method assumes access to the original unwatermarked dataset for verification.
- Very small perturbations might be lost during format conversions (e.g., float rounding in shapefiles).
- This example used static array data; real applications require GeoDataFrames and shapefile/GeoJSON handling.

Significance:

This watermarking method provides a lightweight and effective way to embed ownership in geospatial data. It's particularly suitable for verifying intellectual property in research, consultancy, and GIS product development.

Conclusions

In this project, a simple and effective watermarking technique was developed using NumPy to protect the ownership of geospatial data. By introducing very small, random changes to the coordinates—controlled by a secret key—the watermark remains invisible while preserving the original data's accuracy. The watermark was successfully verified, demonstrating that ownership can be confirmed without noticeable distortion.

This method lays the groundwork for secure data sharing in geoinformatics. It can be further expanded to handle real-world spatial datasets using GeoPandas, and improved to resist common data alterations like compression or format changes, ultimately contributing to robust digital rights management in GIS.

Contribution of Each Team Member

Abhinav Kumar	Paper Review and finding methods for watermarking and Code implementation and report preparation.
Satyam Kumar	Paper Review, Code implementation, methodology and report preparation
Alok Kumar Singh	Presentation preparation, Code implementation and report preparation

References

https://en.wikipedia.org/wiki/Digital_watermarking

- Li, Q., & Zhong, Y. (2007). *Digital Watermarking for Vector Maps*. *Int. Journal of Geographical Information Science*
- Abdelrahman, A., et al. (2018). *Watermarking Techniques for Geospatial Data Integrity*. *Journal of Spatial Information Science*.