

Introduction

Historical maps offer valuable insights into long-term trends in urban development and environmental change—but many of these maps exist solely in physical form. Cartographers, archivists, and government employees tasked with map digitization must strike a balance between maintaining accuracy and improving processing speed via automation.

While automation has improved map digitization—through georeferencing and feature extraction—historical maps often contain distortions and scale inaccuracies. As a result, experts must manually validate each step to prevent error propagation, making verification labor-intensive.

We ask:

How can we systematically verify that a historical map has been accurately digitized?

Background

Time Series Shapelets

Time series shapelets are short subsections extracted from a time series representation of a shape that help distinguish between different classes [1]. By focusing on local structural features instead of the entire sequence, shapelets provide a more efficient and interpretable way to classify shapes. Calculating the minimum distance between the points of the shapelet and a given time series can give a measurement that can identify classification of the shape.

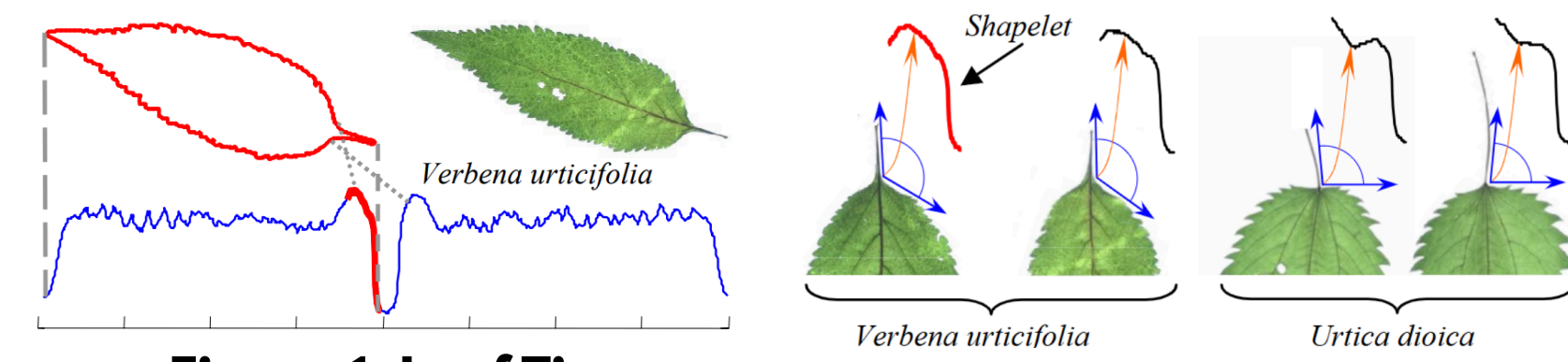


Figure 1. Leaf Time Series

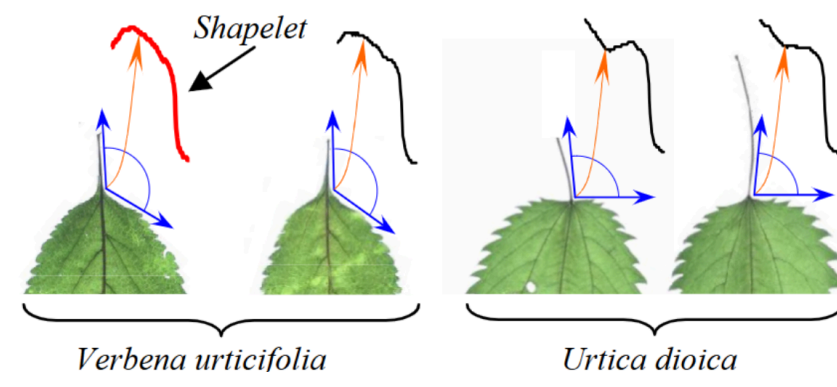
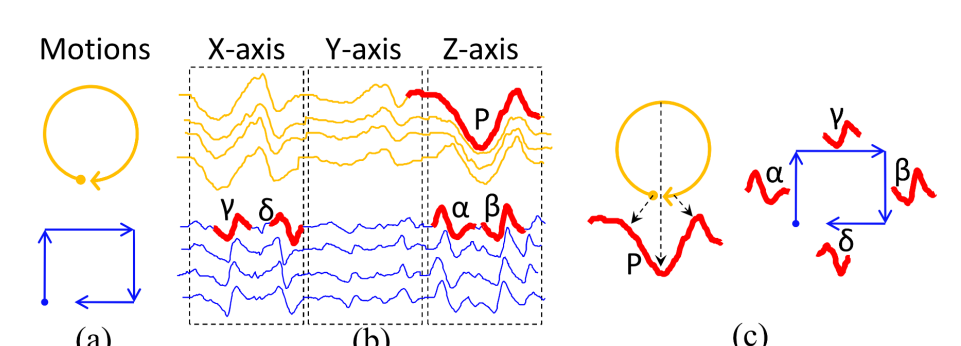


Figure 2. Leaf Shapelets

Logical Shapelets

Logical-Shapelets extend this idea by incorporating speedup algorithms such as Efficient Distance Computation, Candidate Pruning, and Fast Shapelet Discovery, which significantly reduce the total time complexity of the original time series classification algorithm, allowing quicker computations for large data sets. Logical shapeletes are also significantly more accurate than classic shapeletes or other state-of-the-art techniques [2].

Figure 3. Motion-Based Shapelet Patterns



Proposed Methodology

Raster Format

Physical historical maps are captured in pixel form using high-resolution cameras or scanners

Georeferencing

The raster is bound to a coordinate system by matching select control points to their known coordinates and applying a transformation.

Feature Digitization

Features of interest are manually or automatically converted into a polygon form

Centroid and Hub lines

The center point and boundary points are calculated for each polygon, and hub lines are drawn to connect each boundary point to the centroid

Radial Distance data

Each hub line is converted into an angle and distance datapoint to create a numerical representation for each hub line (angle, distance)

Time Series Creation

The extracted angle and distance data points are then used to convert two-dimensional features of the map into a one-dimensional time series signal

Subsequences and Shapelets

A Smaller subsequence of the timeseries graph is then used to compare with the reference map

Verification

Each map's distance to the learned shapelet is plotted showing how closely the digitized maps align with the reference signal - the smaller the distance, the more accurate the digitization.



Figure 4. Feature

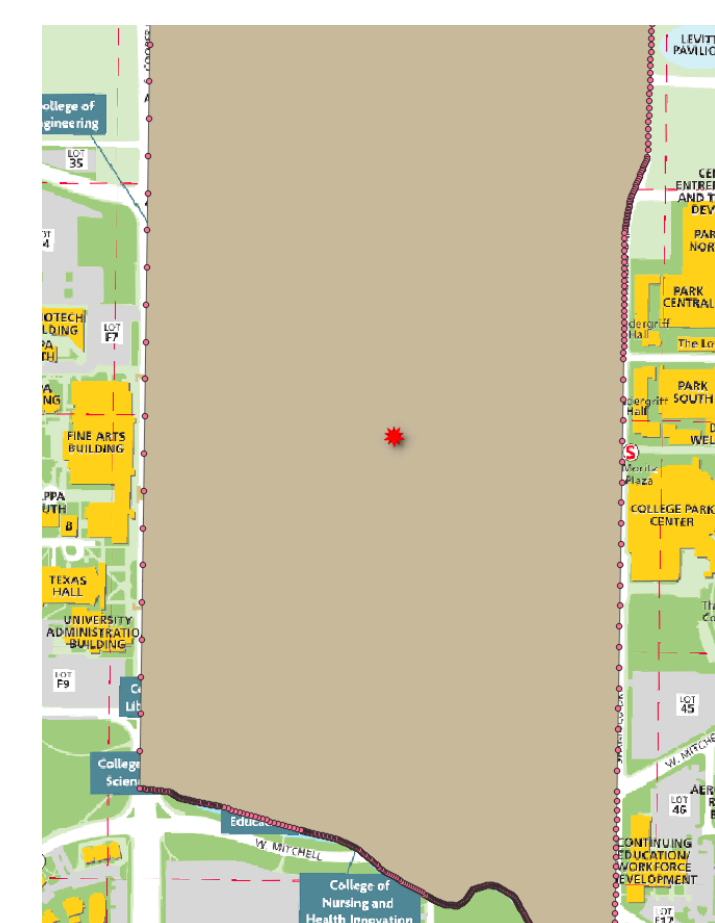


Figure 5. Polygon and Centroid

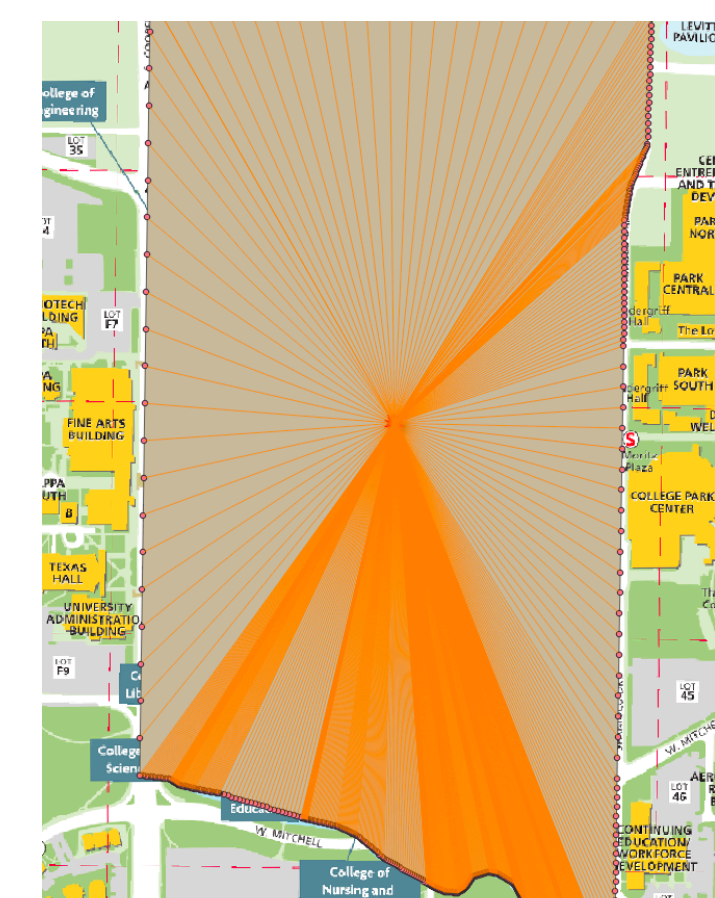


Figure 6. Hub lines

Looking Ahead

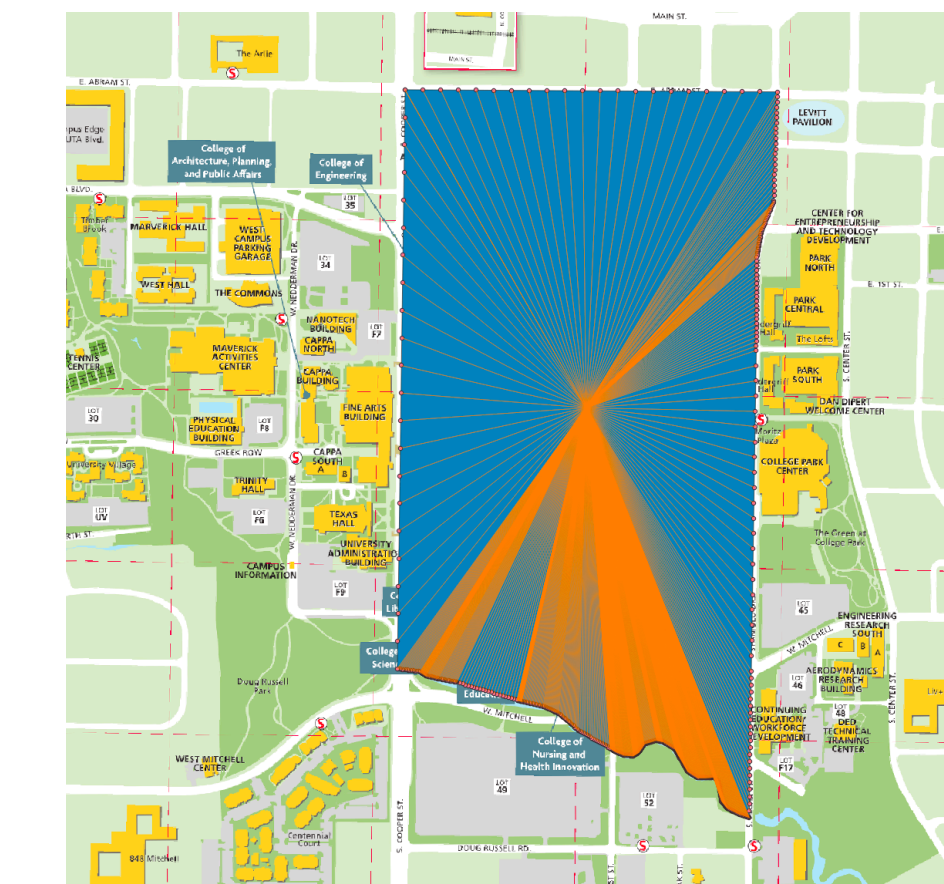
We are looking to automate the geo-referencing and feature digitization steps using computer vision, which would make the front end of the pipeline fully hands free. We also are working to test our methodology on maps from other cities and explore other machine learning shapelet approaches to push verification accuracy even further.

References

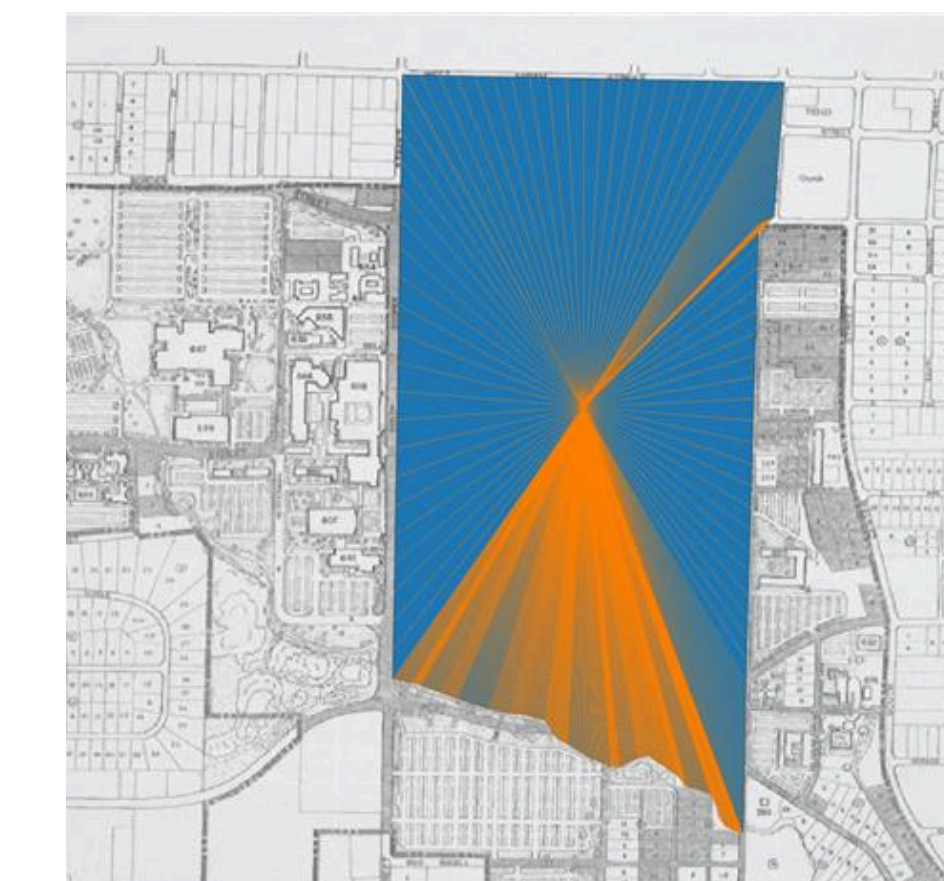
- [1] L. Ye and E. Keogh, "Time Series Shapelets: A New Primitive for Data Mining," Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 947–956, Jun. 2009. doi:10.1145/1557019.1557122
- [2] A. Mueen, E. Keogh, and N. Young, "Logical-Shapelets: An Expressive Primitive for TimeSeries Classification," Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1154–1162, Aug. 2011. doi:10.1145/2020408.2020587

Verification of UTA Campus Maps

The signals evolve across eras and the distance measurements reflect the structural distances between them



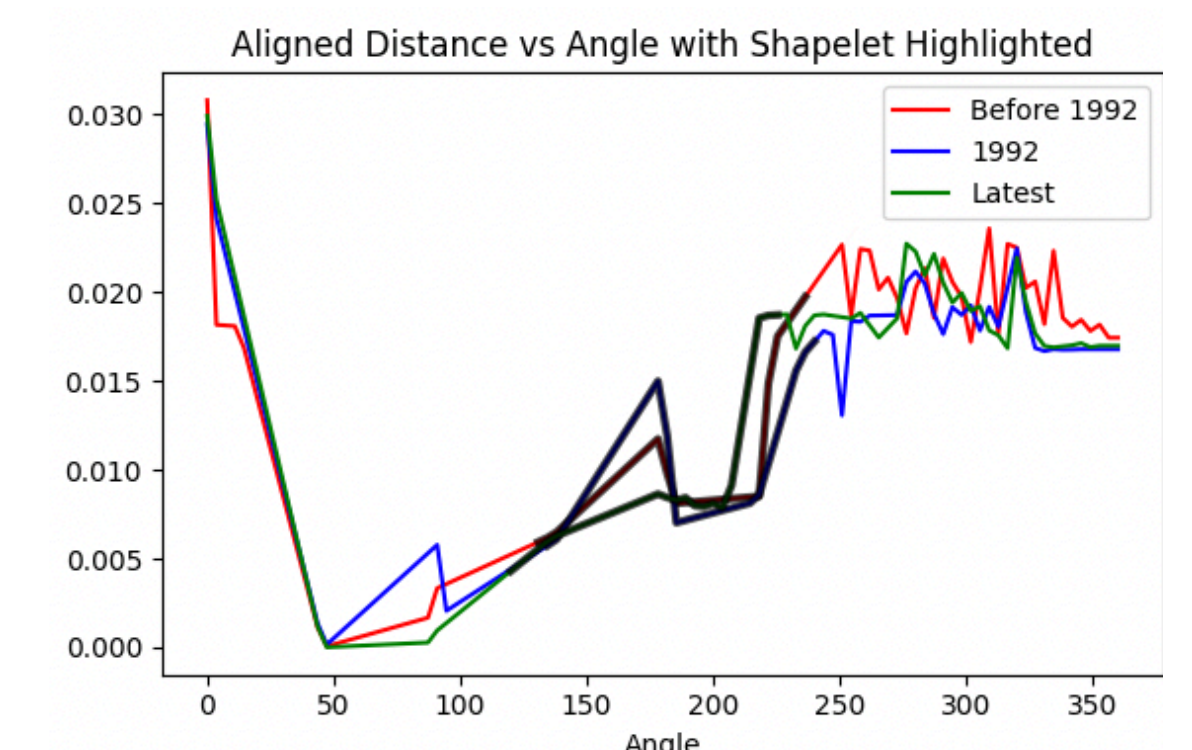
2025 UTA Campus Map



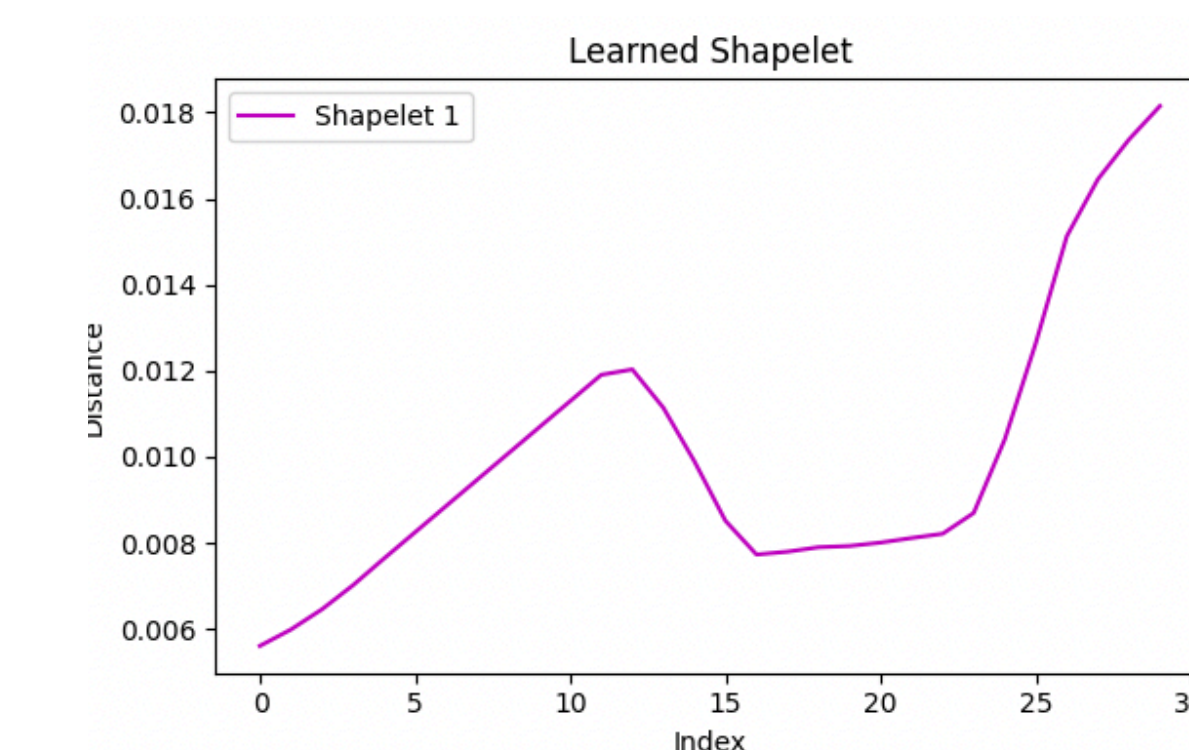
1992 Arlington City Map



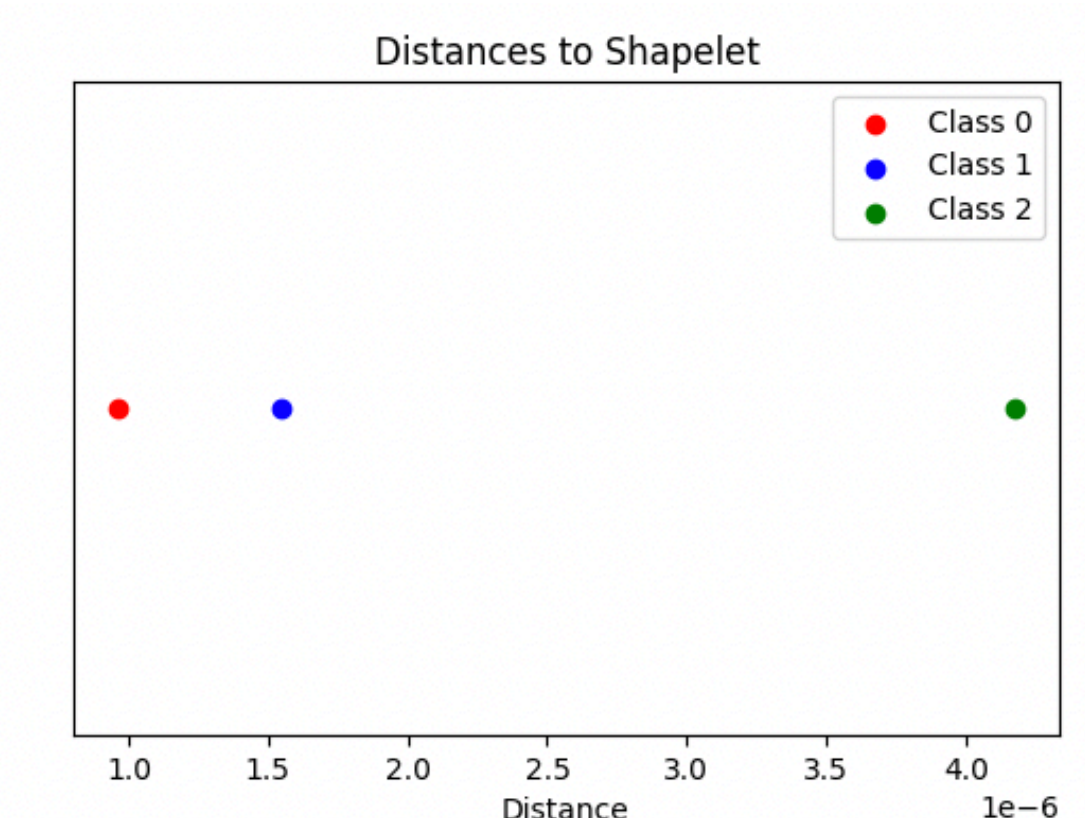
1926 Arlington City Map



Time Series Signal from pipeline



Extracted Shapelet



Plotted distance of each map

Conclusion

We propose a signal-based framework for digitizing historical maps using time series shapelets. By converting two-dimensional spatial features into one-dimensional time-series signals using radial and distance datasets. It allows for the extraction of distinctive subsequences/shapelets for verification. This approach offers a scalable, computationally efficient pathway for automating verification in historical map digitization.



More References