Conveying Uncertainty in Archived War Diaries with GeoBlobs

Johannes Liem[1]* Eirini Goudarouli[2]† Steven Hirschorn[2]‡ Jo Wood[1]§ Charles Perin[3,1]¶

[1]City, University of London
[2]The National Archives UK
[3]University of Victoria, Canada

ABSTRACT

We introduce GeoBlob, an abstract representation of spatio-temporal data dedicated to conveying uncertain positions and uncertain temporal information of entities that move over time. We describe the preliminary design space of GeoBlobs by exploring variations of the technique to visualize data extracted from hand-written World War One diaries, which is our driving context and motivation for this project.

Index Terms: Human-centered computing—Visualization—Visualization techniques; Human-centered computing—Visualization—Visualization application domains—Geographic visualization

1 INTRODUCTION AND BACKGROUND

Since 2014, The National Archives (UK) have been digitizing analog, hand-written War Diaries from the First World War (WWI), documenting the story of the British Army and its units on the Western Front. Using the crowdsourcing platform Operation War Diary [2] (OWD), "citizen historians" tag 1.5 million scanned and archived pages. They classified and annotated essential information on each diary page, generating extensive time-series data about military units, including labels for casualties, unit strength, weather, everyday army life, military activities, soldier names and ranks, location names and dates. As part of a project between City, University of London and The National Archives that aims to communicate the "life behind the trenches" [6], this poster focusses on the visual representation of the spatio-temporal data derived from the War Diaries.

The digitizing and tagging process introduces uncertainty on many levels due to missing records, misspellings (in original diaries and while digitizing), unreadable parts, lost diaries, the reliability of workers, and also through post-processing of the gathered data. As a result, misspellings and ambiguities during geo-referencing of mentioned place names can result in noisy and cluttered map displays when using common visualizations of movement such as line symbols (Figure 1). Furthermore, locations at which the unit did not reside at the time are also mentioned in the diaries but not identifiable as such (e.g., When a soldier was reported to go on furlough to his hometown only the place name but not the context was digitally collected). Additionally, the data typically do not capture the exact temporal order of locations on a single day.

Instead of giving the illusion that the data is complete and clean, our approach is to leverage this uncertainty to produce a more "organic" view of a unit's movement over time.

*e-mail: johannes.liem@city.ac.uk
†e-mail: Eirini.Goudarouli@nationalarchives.gov.uk
‡e-mail: Steve.Hirschorn@nationalarchives.gov.uk
§e-mail: j.d.wood@city.ac.uk
¶e-mail: cperin@uvic.ca

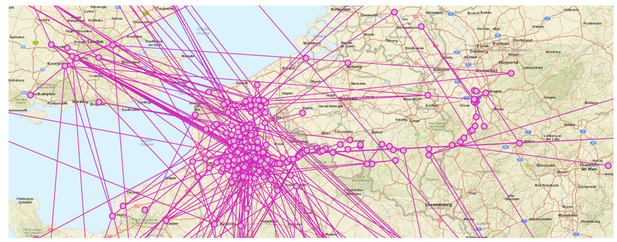

Figure 1: Location data (longitude, latitude) after geo-referencing place names for a single military unit connected in chronological order with lines results in a cluttered map display.

2 GEOBLOB DESIGN SPACE

Our aim was to design visual representations of military units that incorporate and acknowledge uncertainty or noise in the data. We started out by exploring possible geometries to visualize a unit's movement during the war.

We define GeoBlobs as an abstract representation of spatio-temporal data dedicated to conveying uncertain positions and uncertain temporal information of entities that move over time. Instead of showing an entity at a given point in time, GeoBlobs convey an unordered estimation of the possible locations over a temporal window using enclosed shapes. Many different parameters can influence the design of GeoBlobs. We defined and explored the following design variations in the context of the OWD data:

Temporal Window. A start and end date define a window for temporal aggregation of the included locations (Figure 2). Sliding the window on the temporal axis animates the GeoBlob over time.

Spatial Window. We apply a distance heuristics to weight each location within the temporal window and adjust the set of locations that form the GeoBlob (Figure 2). This helps to filter outliers (e.g., unlikely locations that have an incorrect geo-location).

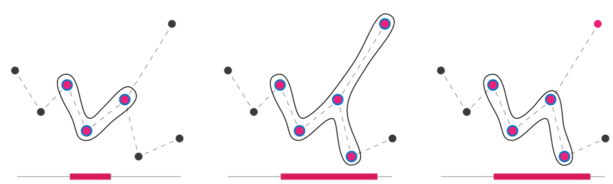

Figure 2: Variations of the temporal (red) and spatial (blue) window defining the inclusion of locations for the GeoBlob generation.

The temporal and the spatial window define the size and shape of the GeoBlob. Keeping both windows constant over time, the size and shape of the geometry convey information about the speed of a unit. While narrow and long shapes imply a greater distance covered (e.g., Figure 3 middle), small circular forms indicate local or no movements (e.g., Figure 3 right).

Shape Parameters. We distinguish convex-hull-like, skeleton-like, and graph-like shapes (Figure 4 top). While skeleton-like geometries connect locations along the shortest distance, graph-

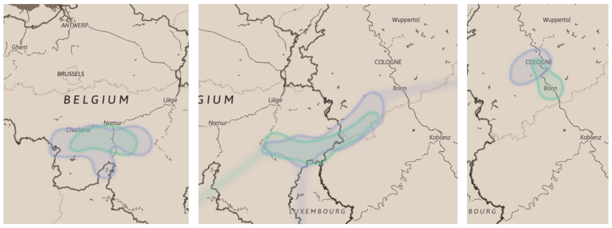

Figure 3: Snapshots of browser-based prototype of two units moving from Belgium into Germany after the war in early 1919 with a temporal window of 10 days.

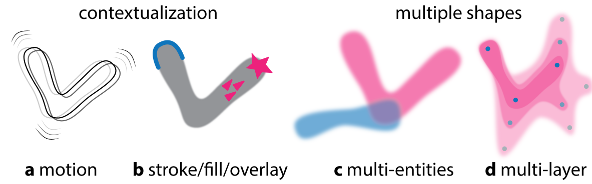

Figure 6: Contextualization and multiple GeoBlob options.

like GeoBlobs consider the temporal order of locations. Buffers for nodes and edges can influence the appearance of a GeoBlob; we distinguish between wide, narrow and mixed buffers (Figure 4 bottom).

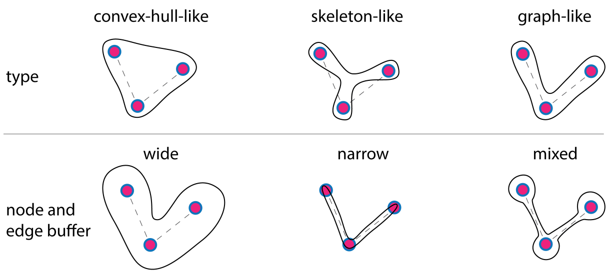

Figure 4: Design variations of the shape/form of GeoBlobs.

Style Parameters. Different visual variables, including color, transparency, focus (blur effect), pattern, or gradient fill can be used to vary the visual style of a GeoBlob's stroke or fill (Figure 5). Especially using a blurry or sketchy [8] style can help to visually convey the uncertainty of the data.

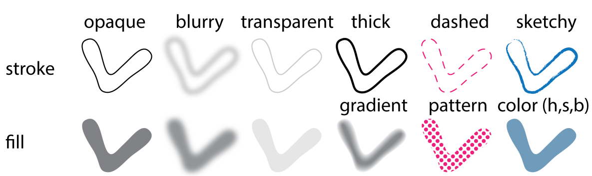

Figure 5: Design variations of stroke and fill style.

Motion. When animating a GeoBlob, its motion encodes the overall direction of the displayed army unit. Here the temporal and spatial window might introduce a certain generalization level (e.g. back and forth movements within a larger temporal window are not obvious), which on the other hand reflects the uncertainty in the data. We also investigate how motion can be applied to convey additional information like a shaking GeoBlob on fighting days (Figure 6a).

Contextualization. Besides motion, outline and fill overlays can provide context. For example by visually integrating unit activities like fighting, re-supplying the front, and resting behind the lines or events like famous battles (Figure 6b).

Multiple Shapes. Mapping multiple entities (i.e., army units) concurrently allows to compare their individual movements (e.g., two units were located at the same front, but split after a battle) (Figure 6c). Generating multiple layers for a single GeoBlob taking the different location probabilities into account and using transparency allows to convey the uncertainty of the locations (Figures 6d and 3).

3 Prototype

We implemented a browser-based prototype using the mapping framework leaflet [1] with custom D3 SVG overlays [3] to visualize the OWD data. To calculate the geometry for the GeoBlobs, we make use of the Bubble Sets algorithm [5].

A set of sliders allows to explore and vary the visualization design space. It includes interactive adjustment of the temporal window (how many days are visualized by a unit GeoBlob), spatial window (the distance from a blobs centroid to remove outlier locations forming the GeoBlob), blob properties (e.g.,size or style), playback speed, or duration and fade out time of overlaid events.

4 Discussion and Conclusion

GeoBlobs are an abstract representation of moving entities on a map over time with uncertain positions. In contrast to the visual certainty conveyed by crisp line or dot visualizations, GeoBlobs leverage the uncertainty in the data to produce an "organic" view of such entities, in this case of UK Army units during WWI. GeoBlobs also suffer less from scalability issues, while line or dot/glyph visualizations become cluttered when the number of data points is large.

GeoBlobs are well suited to any dataset with uncertain locations (e.g., hurricane prediction data). Moreover, the technique can be transfered to domains where the movement data is not uncertain but has a high density or the area of coverage is of interest. For example, GeoBlobs are a promising way of visualizing the movement, the spatial coverage and the pressure in team sports, as well as of visually aggregating spatio-temporal trajectories of players. All are important topics in sports visualization [7]. Due to its configurable visual appearance we think GeoBlobs can be used in Data Comics [4].

Future work will include the incorporation of additional information from the OWD data and additional data sources about WWI. Which will lead towards a narrative visualization for communicating "the life behind the trenches", one aim of the collaboration between City, University of London and The National Archives. We also plan to run empirical evaluations to assess the strengths and weaknesses of GeoBlobs and their variations.

References

[1] Leaflet. https://leafletjs.com/.
[2] Operation War Diary. https://www.operationwardiary.org/.
[3] Leaflet.D3SvgOverlay: Leaflet Plugin: D3 SVG Overlay, June 2018.
[4] B. Bach, N. H. Riche, S. Carpendale, and H. Pfister. The Emerging Genre of Data Comics. *IEEE Computer Graphics and Applications*, 38(3):6–13, 2017.
[5] C. Collins, G. Penn, and S. Carpendale. Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1009–1016, Nov. 2009.
[6] R. Grayson. A Life in the Trenches? The Use of Operation War Diary and Crowdsourcing Methods to Provide an Understanding of the British Army's Day-to-Day Life on the Western Front. *British Journal for Military History*, 2(2), Feb. 2016.
[7] C. Perin, R. Vuillemot, C. D. Stolper, J. Stasko, J. Wood, and S. Carpendale. State of the Art of Sports Data Visualization. *Comput. Graph. Forum*, June 2018.
[8] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. Sketchy Rendering for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2749–2758, Dec. 2012.