Functional Scene Element Analysis of Open-Source Data
Geospatial clustering of regions based on similar social media and kinematic behavior

Kitware, Inc

Principal Investigator: Matt Turek

Area of Interest: a7. Social Media for Situational Awareness.

Capability Description: Kitware’s Functional Scene Element (FSE) capability provides a geospatial centric framework to represent, cluster, visualize and analyze data from multiple disparate data sources, including social media data, such as Twitter, and kinematic data, such as that provided by the Caltrans Performance Management System (PeMS). These datasets were chosen because they can be publicly obtained and a dataset with spatial and temporal overlap across modalities can be assembled. The framework generalizes methods developed for representing moving objects (e.g. people, vehicles) in wide area video (such as WAMI) to a variety of non-image data types. This provides a common representation for different types and quantities of data that can be used to cluster, visualize, compare, and cross reference analysis results, and thereby provide a more complete understanding of behavior within particular geographic areas. In this experiment we will explore FSE capabilities with two datasets:

1. Twitter text messaging data obtained from Twitter’s public API which provides a real-time sample of tweets over a geographic region.

2. Aggregated vehicle statistics provided by the State of California’s Performance Measurement System (PeMS). This service provides real-time and archived measurements from over 39,000 detectors that span the Californian freeway system.
across all major metropolitan areas. These statistics are an open source surrogate for the motion statistics that could be available from typical ISR assets.

We are interested in exploring the relative contributions of each dataset to enhancing situational awareness.

**Experiment Objective/Hypothesis:** There are multiple hypotheses underlying our work. The first hypothesis proposes that automated machine learning tools can provide useful situational awareness capabilities based on big data sources, such as social media or kinematic data (typically derived from ISR sensors, e.g. tracks). An associated hypothesis proposes that the geospatial information in social media, such as geo-tagged Tweets, can provide useful situational awareness summaries. Finally, we hypothesize that the fusion of social media and kinematic data can provide more utility than each sensor in isolation. Our objectives include assessing the utility of the FSE approach applied to this problem domain, exploring the balance between real-time and forensic analysis tools, identifying the highest value features and use cases, and quantifying the data volume/throughput needs and challenges in a more dynamic user environment.

**Experiment Plan / Data Collection Plan:** We will assess the ability of the user to understand and summarize the behavior in a large geo-spatial region with and without the prototype FSE tools. The particular challenge from a user perspective involves the ability to productively ingest and understand a large quantity of social media and kinematic data.

Regarding data collection, we will collect all the data required for the experiment prior to the event. This will include both Twitter social media data and PEMS traffic data. A baseline dataset incorporating both modalities over a shared time frame and geospatial range has already been collected.

**Measures of Performance & Effectiveness:** We will assess the computational requirements for the algorithms automatically assessing the data. We will compare the ability of a user to develop a situational awareness summary with and without the prototype tools. If we have developed sufficient ground truth by the time of the experiment, we may be able to quantitatively measure (probability of detection, false alarm rate) the performance of the algorithms to discover interesting regions or behaviors.

**What new capability does this represent?** A vast amount of social media data is openly available. Most approaches focus on providing new methods for searching, filtering, and visualizing the data. We take a different approach, and instead attempt to characterize geospatial regions based on the behavior that takes place in those areas. Our approach is general, and can fuse data from multiple sources to characterize and summarize the behavior in the geospatial regions. We will investigate the fusion of social media and traffic monitoring data to characterize the behavior in an urban region.

**What capability gap does this address?** Our technique is able to integrate information across multiple sources, including fusing data such as social media data and kinematic information to characterize geospatial regions based on joint behavior. Our technique uses advanced machine
learning techniques to summarize potentially vast amounts of data into a geospacial behavior map. Such a representation is valuable to analysts as it provides a baseline situational awareness that can leveraged to find other regions with similar behavior and to detect changes in behavior patterns.