Title

ALPR DataDive: The Visualization and Analysis of Automated License Plate Recognition (ALPR) Databases – from Data Science to Data Ethics

Organizers

Bryce C. Newell (University of Kentucky), Michael Katell (University of Washington), Jeff Hemsley (Syracuse University)

Key Participants

This is not an instructional workshop/session with planned keynotes or presentations beyond those offered by the organizers (see below for details). As such, key participants include the organizers and all of the attendees. We will make personal invitations to certain other individuals prior to the conference.

Abstract

The growing ability of law enforcement to easily and affordably draw information-rich insights from the surveillance data they collect implicates important social, political, and ethical concerns. For example, many police departments have incorporated automatic license plate recognition (ALPR) into their surveillance portfolios by equipping squad cars with devices that capture the license plate number of every vehicle that comes into sensing range. The fact that some police agencies have been legally required to disclose ALPR databases to members of the public creates an opportunity for researchers seeking to better understand the impacts that the creation, analysis, and disclosure of these surveillance databases can have in society. The purpose of these sessions is to provide scholars access to ALPR data and to provide an interactive and collaborative opportunity for scholars from multiple disciplines to investigate the relevant social, technical, and ethical ramifications through data visualization, data analysis, and ethical/policy-based analysis.

Purpose and Intended Audience

The purpose of the proposed SIE is to provide an interactive and collaborative opportunity for scholars from multiple disciplines to investigate the social, technical, and ethical ramifications arising from the collection and public disclosure of large police surveillance databases. The intended audience includes scholars from a variety of disciplines within information science/studies, including data visualization, data science, social informatics, computational social science, geographic information systems (GIS), data curation, data and information ethics, and information policy. During the opening session, we plan to make space for scholars across these areas to join together to dive into the available data and make plans for visualizing and/or analyzing the data, and for unpacking the social and ethical implications that such analyses of ALPR data engender.

Proposed Activities

Pre-conference. Prior to the conference, we will make the datasets available to potential participants who request it (we will promote the SIE on a dedicated webpage with a “request the data” link), both to allow additional time for those who would like to begin exploring the data

early as well as to promote the session and generate interest among potential participants. Importantly, the license plate numbers in the scan data we provide will be altered to protect the privacy of those drivers whose plates have been captured in the datasets.

During the conference. At the conference, we propose to hold an opening session near the beginning of the conference in which we will make multiple ALPR datasets available to participants (who haven't already downloaded them), with the goal of spurring multiple, parallel efforts to dive into the data over the course of the conference, culminating in a closing session where participants come back together and present their analyses (empirical, technical, or ethical) and/or data visualizations and discuss the social and ethical implications that have arisen during their analysis. The data shared with participants will include multiple databases generated by the use of ALPR capture devices (cameras) by police departments in Washington, California, and Massachusetts, and that generally include precise latitudinal and longitudinal data from the scans of a vehicle's license plate, actual license plate numbers, as well as date and timestamp information. The data was sourced through public records requests under each state's access to information law and, as such, is considered part of the public record. (The use of ostensibly "public" but non-consensually sourced data for research provides a secondary opportunity for scholarly discourse in these sessions.)

Post-conference. After the conference, we plan to post the results of the SIE (e.g., visualizations, short summaries of analytic findings, etc.) to the website, and to organize these results into a collaboratively-written report.

Relevance to the Conference/Significance to the Field

This SIE is relevant to a variety of areas of inquiry within the information field and to those who generally attend the iConference, including data visualization, data science, computational social science, social informatics, GIS, data curation, data/information ethics, and information policy (among others). In particular, the existence of the sort of data we will unpack throughout this SIE and the possibilities that this makes available to researchers (and others), potentially poses significant social, ethical, and legal risks to individual privacy, and raises important questions about the appropriate boundaries of information access, information control, the transparency of government information, information privacy, data ethics, and research ethics. Additionally, we have chosen an interactive format to more fully engage participants in the event, and to provide space for presentation and the development of ideas over a (short) period of time.

Duration

We expect to hold one introductory session at the beginning of the conference and one closing session on the last day of the conference (for a total of two 90-minute sessions).

Special Requirements

None required, beyond the ability to project visual projects to a screen (a projector and a screen).