Deep Data Dive

Extracting personal telemetry and metadata maps

>> **Workshop 8**

>> Design Tools and Skills 1

>> *Semester 1 - August-December 2024*

>> *Instructor:*
Joris Putteneers

>> *Coordinator:*
Tom Jenkins

>> *Faculty Team:*
Bahnfun (Dream) Chittmittrapap, Deniz Guvendi, Hayden Minick, Hseng Tai Lintner, Joris Putteneers, Stefan Svedberg, Warisara (Nice) Sudswong.

1. Pedagogical Activities

1.1 what is telemetry data

Have you ever thought about Google's revenue model? Its primary stream of income comes from personalized advertisements. But how does Google obtain this personalized data, and what does it consist of? This data is called telemetry data and refers to automatically and remotely transmitted information.

Google and Facebook, for example, are quite infamous for collecting a very large volume and scope of user data

Among many other types, this data is likely to include:

- Network connectivity (to check who is near you)
- GPS coordinates (to check where you are)
- Vehicle type, speed, direction, etc. (where you are going)
- Temperature, humidity, air quality, light levels (environmental conditions)
- Audio and video calls (what you are talking about)
- Spending patterns (what are your spending habits)
- Click patterns (what you like to do on the internet)
- Ambient audio recording
- Eye tracking patterns (what you like to look at)
- Heart rate, blood pressure, body temperature, emotional state, sleep patterns (some smartwatches only)

You can practically create a full digital fingerprint of a person by looking at the telemetry data, which sounds very scary (which it is), and it can be used for both good and bad purposes.

Luckily, there are governing bodies like the General Data Protection Regulation (GDPR) and Personal Data Protection Act (PDPA) that are instantiated to protect the users. Many others around the world exist.

Most software includes some nuanced version of this telemetry system. This includes Adobe, Rhino, and even your operating system like Windows or macOS. This data is mostly used by software companies to receive feedback on their programs, to pick up on sales and trends, and to manage your licenses, etc.

An extension of this is a field called IoT (Internet of Things), where devices are connected to the internet all the time. This includes devices such as your smartwatch, Google Home, and even your cell phone can be classified as such.

However, if used correctly and ethically, with a common consensus, where the code is open-sourced, and with a clear agenda, it can be an extremely useful tool for data analysis, visualization, and predictions in many professions.

For us as architects, this can mean space utilization analysis and optimization for urban planning proposals, as well as performance monitoring to check if your building is functioning as well as our predictions.

In this exercise, the service we will utilize to capture this telemetry data is 'google takeout'.

1.2 what is metadata

Globally, in 2023, about 400 million terabytes are uploaded, which includes 5.3 billion images being uploaded every day. That is as many images in a day as the whole year of 2011. We are living in a scrolling culture that is oversaturated with imagery. When you click on an image on your phone, you most likely can see the resolution or where the image was taken. This is called metadata, and it is embedded in most images. It can look similar to this:

Image: example_photo.jpg (JPEG, 3024x4032, sRGB) EXIF: Apple iPhone 12 Pro, iOS 15.1 2023:09:18 14:30:15, 1/120 sec, f/1.6, ISO 100 Focal Length: 4.2mm, Flash: Off GPS: 40°41'21.91"N, 74°2'40.20"W Altitude: 10.5m, Speed: 0, Direction: 45.0° IPTC: © John Doe Photography Title: Sunset in Central Park Caption: Beautiful sunset view from Central Park, New York City Keywords: sunset, park, nature, city, landscape XMP: Created: 2023-09-18T16:45:30-04:00 (Adobe Photoshop 2023) Modified: 2023-09-19T09:15:22-04:00, Rating: 4 Color Profile: sRGB IEC61966-2.1 (Display Device Profile) Photoshop: Document Ancestors: xmp.did:9A8B7C6D5E4F3A2B1C0D9E8F7A6B5C4D Thumbnail: JPEG, 9012 bytes, 160x120 pixels

In This exercise, the most interesting data is the gps data and timestamp.

This GPS data, combined with data from a telemetry service like google takeout, can provide interesting insights into not just where or when a photo was taken, but also the circumstances surrounding its capture. For instance:

The time and date can reveal patterns in user behavior or popular times for certain locations.

Camera settings can indicate whether it was a casual snapshot or a more deliberate, professional shot.

2. Objectives

With this metadata, and the semantic telemetry data from google takeouts, we can start displaying the two on top of each other, and hopefully finding interesting insights that can benefit your project.

Project Preparation:

Choose a site in or around Bangkok for your 1st year project within the image:

- Site has to be beneficial to your project (e.g., bridge, rooftop, house, forest, boat, temple).

Ensure your Google Maps Timeline is active:

- View your timeline in Google Maps. If you can see a route, that means it is successful.
- If it's not working, watch this tutorial:
- If issues persist, extract a Google Takeout, deselect everything, and only select timeline, download and verify the timeline sematic history data has a .json files like "2024_AUGUST.json" with data inside of it. https://takeout.google.com/
- Report any additional problems in the chat

Verify that your phone can store images with metadata:

- Check if you can see the map when scrolling up on your photos

Site Visit Instructions: Turn on your phone's location services and cellular data BEFORE leaving for your site. Your location and cellular data must stay on during your whole trip. Take a minimum of:

- 25 photos of your site (front views, side views, interior, exterior, etc.)
- 15 photos of the site surroundings (neighboring architecture, waterways, railroads, people, etc.)
- 10 photos during transportation (multiple modes encouraged and will giove more interesting results: bike, boat, foot, on foot, car, etc.)

Videos can be found here:

Make sure to read the information posted on the discord channel regarding capturing your trajectory and images.

WS8_presentation

WS8_1_google_takeout

WS8_2/3_reconstruction_fase_A

WS8_4_reconstructing_fase_B

WS8_5_Bangkok_route

WS8_6_uploading_ZIP_and_tasks

All deliverables are explained in video 6.

Quantity over Quality

Given that we have approximately 90 students participating, we can expect a significant amount of data to be produced. In this exercise, we should be able to extract more information and correlation the more data we have to analyze. More data means:

1. Increased Granularity: With more data points, we can increase the resolution and detail of our map.
2. Enhanced Resolution: A larger dataset allows us to create a more precise and accurate representation of the information we're mapping.
3. Reduced Impact of Anomalies: As the volume of data increases, individual anomalies or outliers will have less influence on the overall results,
4. Extraction of Meaningful Information: A larger dataset provides more opportunities to identify patterns and extract meaningful information that might not be apparent with less data.

While I will provide the final map in the following week, you are highly encouraged to show interest and attempt to work with the data yourself

All deliverables are explained in video 6..

3.   Procedures

3.1.   Location
online

3.2.   Agenda
13:00 - 13:15 - Attendance Check *(Aj.Joris)*
13:15 - 13:45 - DTS Introduction *(Aj.Joris)*
13:45 - 14:15 - Workshop Introduction *(Aj. Joris)*
14:15 - 14:30 - Set up
14:30 - 17:30 - Task A-1
17:30 - 18:00 - Attendance Check + Tutorial *(All)*

*3.3.   In-Class Task | **semantic mapping** (in Class)*
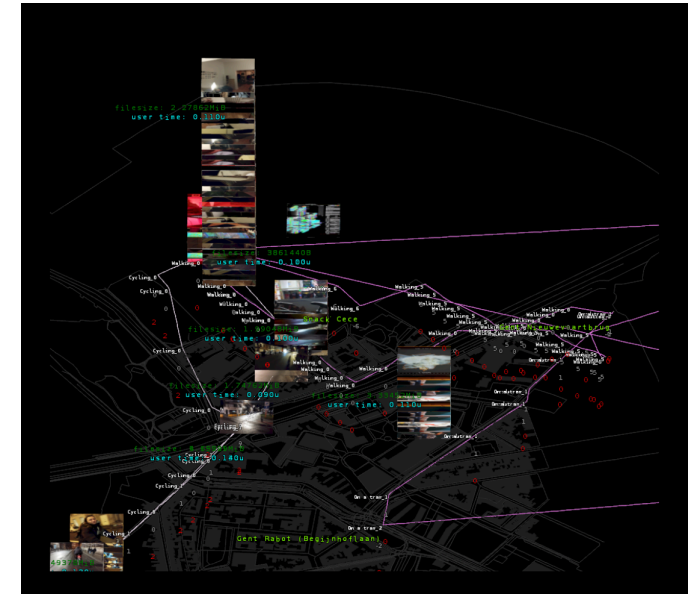
4.   Material required

A google account with location history turned on

Houdini apprentice 20.5

Houdini labs tools

Images with gps metadata

5.   Submissions

All deliverables are described in video6

Material to be submitted should be complete and in accordance with the guidelines presented in class.

File Naming:
[[Your DTS Code]_WK08_fase_A.png
[Your DTS Code]_WK08_fase_B.png

Submission data dump Form Link:
Submission fases Form Link:

Submit the digital copy of your work via this google form by Tuesday 1st October, 8pm.

6.   Grading Criteria

All submissions are present and performed according to the instructions defined by the brief procedure.

Completion
Timely submission in good condition ………....……….….. 40%

Quality
Conceptual clarity and craftsmanship …..………………….. 60%

7.   Keywords

This is a list of words that will be used in class and you should familiarise with:

#telemetry

#metadata

#privacy

#python

#houdini

#mapping

#analysis

#dataviz