Vision Based Localization on Smartphones
Arjun Ramesh & Art Zhu   Mentor: Jan Harkes

Abstract

Developed an end-to-end system to accurately localize a user on CMU campus using vision-based information from their environment. Our Android application enables users to scan their surroundings and receive their latest location update!

Problem Space

Accurate positioning is key to support many applications like Google Maps, Uber, and autonomous drones/vehicles. Current technologies commonly use GPS or wireless access points to obtain this information. Some drawbacks include:
- Blindspots, weak reception areas, e.g. indoors
- Ineffective for fine-grained localization (within a room, < 5m)

Vision-based techniques provide a complementary tool to GPS that can enhance localization in these challenging environments. With a combination of magnetic orientation and image recognition to identify landmarks, we can infer a user's location using triangulation algorithms.

System Architecture

Two Asynchronous flows:
- Gather flow for surrounding information;
- Retrieval flow for synthesized location.

Gather:
Client streams images to the MaaS Cloudlet offload machine using the Gabriel Framework. Images are delivered to the Location Synthesis Backend App for object detection. The backend application caches learned objects for use in the Retrieval flow.

Retrieval:
The Location Synthesis Backend App launches periodic tasks to analyze the location of the user based on the set of currently detected objects and stream the results back to the Android client.

Implementation

- Bidirectional Communication between smartphone and Backend Cloudelet Machine:
  - Loose integration with Gabriel Framework.
  - To server: [Image, Orientation (compass), View Angle];
  - From server: [Lat. & Long., Confidence Level, Error range].

- Object Detection: CVAT + OpenTPOD
  - 15 ML Models (combined + individual) recognizing: CFA, Hunts Library, The Fence, Walking to the Sky…
  - Retrieve bounding boxes for high-confidence objects.

- Location Analysis Service:
  - Triangulate using top 2 confident objects w/ orientation.

Key Results

Delivered a fully functional standalone app to perform vision-based localization w/ an Android App using an extensible Backend pipeline.

Challenges Resolved:
- Localization Inaccuracy
  - Report confidence level (transparency) and error range (radius of circle) for user decision making.
- Data denoise (e.g. altering azimuth readings)
  - Delayed streaming for confidence through consistent sensor readings. (Drop outliers)

Evaluation:
- Real user localization delay is caused by inability to detect. Application coverage is dependent on the ML models.
- This technique requires a ML training process, unlike GPS. Generalizing to a broader environment needs more effort.
- Quality of user input when using the application must be controlled. (User cooperation may be needed.)
  - Hard to model confidence of orientation from sensors.
- Full potential to serve as the cornerstone of other applications.

Usage:
1. Open App and scan surroundings;
2. Dialog prompt on Location found;
3. Accept and View Location (see Abstract);

Future Work

- Object Detection adaptability to environmental changes (Snow);
- Speed up user localization
  - More object detection models w/ improved accuracy;
  - Model prioritization; Reduced Fidelity if needed.
- Incorporate GPS as a reference/complementary service.
- Continuous location tracking.