A mobile application for navigating HKU visitors with computer vision

Project Plan

Supervisor:
Dr. Luo, Ping

Team Members:
Liu Kan Man, 3035790733
Wong Riley Hoi-Kiu, 3035829039
Ng Enoch, 3035781536

Table of Contents

1. Background

In this day and age, Google Maps has been a powerful tool in navigation and benefited users by providing various details such as satellite images, geographical features, and street views. However, as the services provided greatly relied on satellite and aerial imagery to create detailed maps, indoor images may not be mapped detailly. Buildings certainly block satellite signals to our devices. Hence, GPS location information is insufficient and inaccurate for indoor navigation [3].

Being one of the well-known universities, the University of Hong Kong is located on the Bonham Road and Pok Fu Lam area of Hong Kong Island, consuming about 14 hectares of land [1]. According to the statistics provided by the University of Hong Kong [2], there are more than 13,000 student admissions in the year 2022. For newcomers, many may have been confused by the lecture hall code mentioned in their timetable, not knowing their way to their destination, on their first visit to the university. It is not a unique situation for first-year students but also visitors. There are numbers of heritage buildings at the university [4], making the university campus an attraction. Also, vast amounts of events, such as talks, ceremonies, and workshops, holding frequently at the university attract visitors to the campus. Although the university offers text-based guiding information, an easy-to-use and detailed indoor navigation system may come in handy for everyone.

Current Approach: Global Positioning System (GPS)

Currently, the university does not provide visitors a navigation application for campus uses but only an undetailed map and some text-based guidelines. Visitors need to rely on other common navigation apps, such as Google Maps, which are usually GPS-based.
Google Maps evaluates users' location by calculating their attitude from the results obtained from the Global Position System (GPS), which is the most used positioning method. The accuracy of GPS will be affected by obstacles and extreme atmospheric conditions [3]. Therefore, GPS is incapable of determining what floor that users are on. It led to misleading location whenever users attempt to access indoor information from Google Maps with GPS.

Proposed Approach: Vision-based Navigation

This project aims to provide users with an alternative way of obtaining their current location. Shifting the core technology for location from GPS to computer vision reduces the reliance on satellite signals and allows the computer to learn about the surroundings and interpret users' positions. This method will alleviate the error caused by GPS.

Motivation of using Computer Vision

Due to GPS's limitations in indoor usage, wireless technology based positioning systems, using such as ultra-wideband (UWB), radio frequency identification (RFID), Bluetooth, and Wi-Fi, are proposed. However, they are usually erratic in accuracy and require extra external infrastructure, which result unsatisfying performance yet high cost in initialization and maintenance [5],[6],[7]. Moreover, these indoor positioning technologies are not popular comparing to GPS.

Compared to other indoor positioning technologies, such as Wi-Fi-based, Bluetooth-based, or RFID-based systems, vision-based indoor positioning offers high accuracy, infrastructure flexibility, robustness, and additional functionalities.

Vision-based indoor positioning can leverage features like object recognition, depth sensing, and motion tracking to provide precise location estimates, which make it less susceptible to interference from environmental factors like signal attenuation, multi-path propagation, or radio frequency interference [5]. Furthermore, vision-based systems do not require the installation of additional infrastructure or hardware like beacons or RFID tags [6]. Also, Vision-based systems can provide additional functionalities beyond just positioning, like object tracking, gesture recognition, activity monitoring, and augmented reality overlays, enhancing the overall user experience [8].

2. Project Objectives

- Evaluate and compare the effectiveness of using vision-based positioning with GPS positioning
- Investigate various well-known supervised machine learning algorithms in object detection / image classification
- Research and optimize the selected machine learning algorithm to obtain high accuracy in real-time object detection / image classification
- Map users' location with results obtained from the machine learning model
- Visualize detection results and navigational guides on mobile application

3. Methodology

a. Scope

The Main Building of the University of Hong Kong is chosen as our first testing site due to its popularity and complexity. In other words, the goal for this project should be able to navigate thousands of visitors through this complex building effectively.

In view of limited time and human resources for developing this project, we will mainly focus on the first testing site. Further development in more areas will be considered if possible.

b. System Structure

The system will consist of mobile app frontend, backend, a server, and a database.

1. Mobile App Frontend

The frontend will provide a user interface to visualise a virtual map, navigational cues on the camera feed, and user controls. The app should also be responsible for the path finding features using downloaded app data.

React Native will be chosen for front-end development because it is a cross-platform mobile development framework with a large community.

2. Mobile App Backend

The backend will handle communication between the mobile app and the server via secure channels. The app will also pre-process images before sending them to the server, ensuring optimal image amount, quality, and size.

For data transmission, RESTful APIs and WebSocket are planned to use since they provide easy and secure channels.

3. Server

The server will host a trained machine learning model capable of performing object detection. Received images will be processed and apply to the machine learning algorithm. Leveraging the trained model, the server will generate geomatical results.

4. Database

The database will be responsible for image storage. The database will also store process history, resource usage, and user reports for future reference, performance analysis, or retrieval.

Data are in a fixed structure, so MySQL, which is a relational DBMS, is suitable for the project.

The following are the flowcharts for client and server sides.

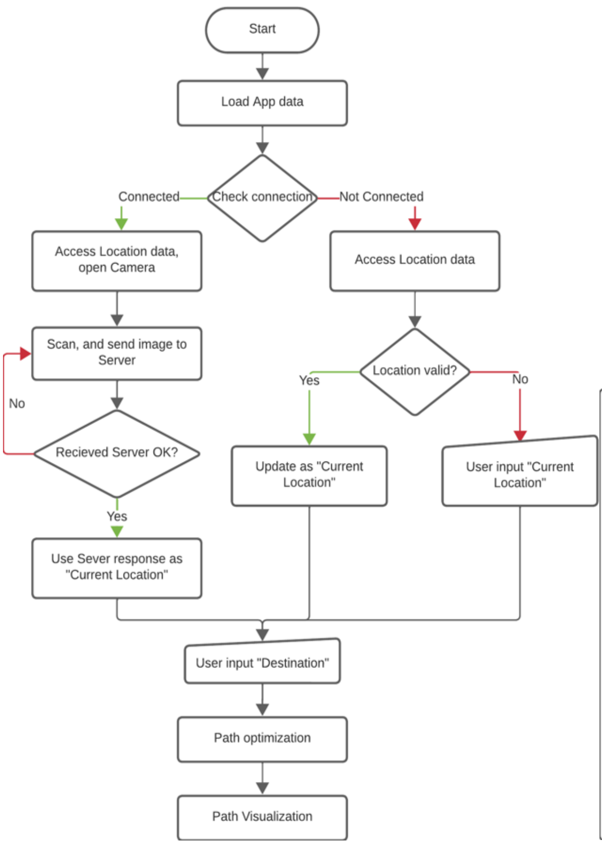

Figure 1 Client-side Flowchart

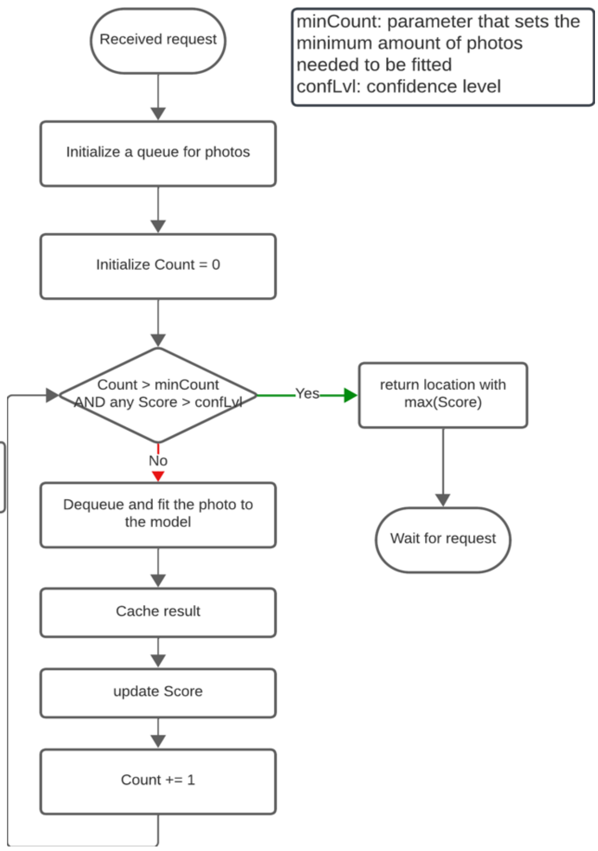

Figure 2 Server-side Flowchart

c. Application Features

1. Vision-based Positioning

The app utilizes the camera on the user's phone to sample images of their surroundings at a designated frame rate. These images will then be transmitted to the server through secure channels and fitted into the object detection model. Once a sufficient number of images have been processed and the model's confidence level has been met, the resulting location will be returned by showing on the map. The location can be used for pathfinding as well.

2. Path Finding

Getting the user's location and destination within the HKU (the University of Hong Kong) campus, the application will calculate the optimal route and display it on the map. Navigational cues will be provided to assist the user throughout the journey. This feature will be available in both online and offline mode. In online mode, the user's location can be accessed through vision-based positioning, phone GPS location, or manual input. In offline mode, the user's location is identified via phone GPS or manual input.

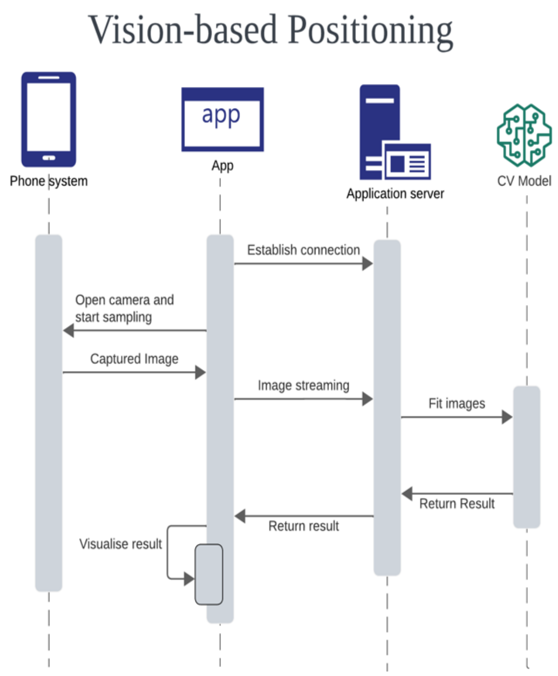

Figure 3 Sequence Diagram: Vision-based Positioning.

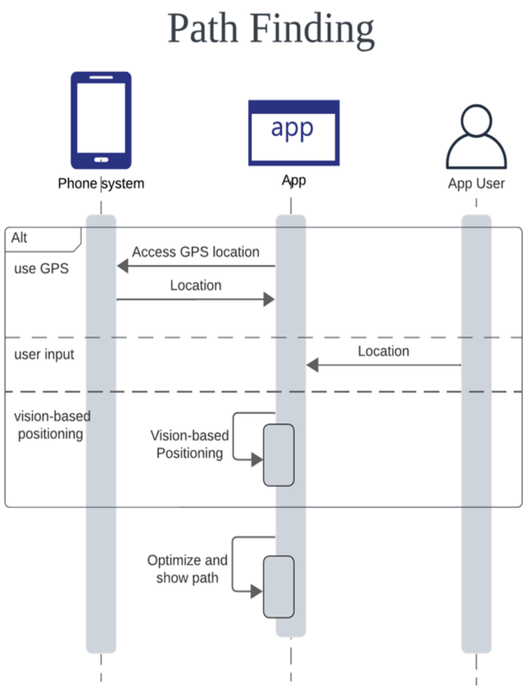

Figure 4 Sequence Diagram: Path Finding

d. Computer Vision Model

The accuracy of the positioning system is dependent to the model design and the quality of training data. Thus, it is important to choose a suitable model and perform strict data management.

Model Selection

In the selection of a deep learning model, two popular models have been identified for consideration: YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector).

YOLO is a real-time object detection model that will be one of the shortlisted candidates for our implementation. It divides input images into grids to predict bounding boxes and class probabilities. It uses a single neural network to directly predict object locations and class labels. Being able to recognize multiple objects within an image, fast-paced real-time navigation can be supported.

SSD is also designed for real-time object detection. SSD utilizes a series of convolutional layers with different scales to predict object classes and bounding box coordinates. It provides better accuracy than YOLO but may have slightly higher computational requirements. Considering the need for accurate object detection in the vision-based navigation app, SSD can be a strong candidate.

To determine the most suitable model, a comprehensive evaluation will be conducted, considering factors such as performance, accuracy, computational requirements, and the specific needs of the vision-based navigation app. The evaluation will also account for the availability of labelled training data, hardware resources, and the project timeline.

Data Management

Training data should go through the following steps:

1. Data Collection,
2. Data Pre-processing,
3. Data Labelling,
4. Data Augmentation

1. Data Collection

Manual photographing will be used to collect relevant images. Operators will walk through the testing area and record the surroundings with a mobile camera. Meanwhile, location information will be marked. Images will be sampled at a designated frame rate to optimize the size of training data. Static images will be supplemented afterwards.

2. Data Pre-processing

Some images may still be repeated, unfruitful, corrupted, or with noises. Unwanted and bugged images will be filtered out at this stage.

3. Data Labelling

Every image will be labelled manually with a geomatical location. The location data should be marked during data collection. Moreover, object labelling, and landmarks annotation are necessary for the machine learning process.

4. Data Augmentation

Data augmentation techniques will be implemented to increase the sample size of training data. By performing augmentation such as image cropping, mirroring, and adjusting brightness, an adequate amount of data can be generated with varying shooting angles, lighting conditions and sizing for model training.

e. Map and Navigation

To provide indoor positioning and navigation, it is necessary to construct indoor maps of the university campus.

Thus, the first step is to gather floor plans of the testing area (the Main Building). We will first obtain floor plans from various sources, including websites, and the university. After multi-source verification, we will perform map rendering, map visualisation and geocoding with corresponding applications such as Mapbox. The maps will then be used for model training, integrating into the mobile application through provided APIs (e.g., Mapbox APIs).

For navigation, Mapbox offers indoor navigation SDKs that provide tools and components for creating turn-by-turn navigation and route guidance, building users' indoor navigation experiences.

4. Schedule and Milestones

Period	Work Description
Sep - Oct	- Analyse different proposed ideas' feasibility and effectiveness - Research and literature review of computer vision-based positioning - Design interface for Mobile application
Oct - Nov	- Collect data in Main Building (1st floor) - Research in different image classification/ object detection machine learning model - Construct mobile application - Data processing (image labelling, mapping) - Indoor map construction
Nov - Dec	- Build and train shortlisted machine model - Perform hyperparameter tuning for each model - Evaluate and select model with best performance
Dec - Jan	- Prepare interim report and presentation - Prepare rough demo for interim presentation
Jan - Feb	- Compare and select appropriate path searching algorithm - Integrate machine learning model and path finding algorithm to mobile application
Feb - Mar	- Deploy mobile application for alpha testing - Debug mobile application - Fine tuning the model
Mar - Apr	- Prepare final report - Source code cleanup - Prepare PowerPoint and Demo/Demo Video for final presentation - Prepare for the project exhibition and project competition

5. Bibliography

[1] "HKU Estate Office - Our Campuses". https://www.estates.hku.hk/campus-planning/campuses#:~:text=The%20Main%20Campus%20of%20the,well%20as%20the%20Main%20Library (accessed Sep 19, 2023)

[2] Communications and Public Affairs Office, "HKU Quick Stats 2022". https://www.cpao.hku.hk/qstats/files/Archive/2022.pdf (accessed Sep 19, 2023)

[3] H. Motte, J. Wyffels, L. De Strycker, and J. P. Goemaere, "Evaluating GPS data in indoor environments,". https://www.researchgate.net/publication/269978666_Evaluating_GPS_Data_in_Indoor_Environments. Advances in Electrical and Computer Engineering, vol. 11, no. 3, pp. 25–28, Jan. 2011, doi: 10.4316/aece.2011.03004. (accessed Sep 21, 2023)

[4] "HKU Estate Office - Heritage and Conservation". https://www.estates.hku.hk/campus-information/Heritage-features/heritage-and-conservation (accessed Sep 21, 2023)

[5] [1] T. Wu, L.-K. Chen, and Y. Hong, A Vision-Based Indoor Positioning Method with High Accuracy and Efficiency Based on Self-Optimized Ordered Visual Vocabulary, https://lightweb.ie.cuhk.edu.hk/api/publication/1658736041110-07479682.pdf (accessed Sep. 21, 2023).

[6] D. Khan, Z. Cheng, H. Uchiyama, S. Ali, M. Asshad, and K. Kiyokawa, "Recent advances in vision-based indoor navigation: A systematic literature review,". https://www.sciencedirect.com/science/article/abs/pii/S0097849322000371. Computers & Graphics, vol. 104, pp. 24–45, May 2022, DOI: 10.1016/j.cag.2022.03.005. (accessed Sep 22, 2023)

[7] C. Wiegand, "Achieving Blue Dot: Best types of indoor positioning systems," Mar. 22, 2023. https://www.inpixon.com/blog/what-is-the-best-system-for-achieving-blue-dot-indoors (accessed Sep 22, 2023)

[8] S. Ahmad, "How is Mobile Computer Vision Changing the World?," Mobisoft Infotech, Dec. 01, 2020. https://mobisoftinfotech.com/resources/blog/how-is-mobile-computer-vision-changing-the-world/ (accessed Sep 23, 2023)