ENS 491 – Graduation Project (Design)

Proposal

Project Title: Indoor Localization Using Camera Images

Group Number: 128

Group Members: Kağan Kağanoğlu, Yarkın Alpmen Akyosun

Supervisor(s): Mustafa Ünel

Date: 10/12/2023

Sabancı Üniversitesi

1. **ABSTRACT**

Indoor localization is critical for precisely navigating in large, enclosed spaces, especially in areas like shopping malls, airports, and for individuals with visual impairments. Traditional methods like GPS and Wi-Fi face limitations indoors due to signal obstruction. Unlike radio frequency based methods, it doesn’t require equipment setup, it only requires a single monocular camera which keeps the cost low. It also doesn’t rely on artificial marks. In contrast, image-based indoor localization, leveraging computer vision and deep learning, presents a promising solution. This proposal aims to advance existing approaches by designing and implementing an image-based indoor localization system. The complexity arises from the absence of a straightforward solution, combining issues of accuracy, hardware requirements, and response time. The project involves a comprehensive literature review, dataset search, and algorithm selection. This project addresses the complexity of indoor localization, emphasizing innovation in computer vision and deep learning. The proposal outlines a comprehensive plan, considering challenges, risks, and ethical considerations to ensure the successful development of an advanced indoor localization system.

2. **INTRODUCTION**

Indoor localization aims to provide precise localization within buildings, and it is particularly important in large places like shopping malls, airports, and for people with visual impairments. Unlike outdoor localization, traditional methods such as GPS, satellite imaging, Wi-Fi, and Bluetooth receive poor results within enclosed spaces where the radio signals encounter difficulty passing through walls. Meanwhile, image-based indoor localization presents a promising solution. This approach provides stable and reliable results without requiring high maintenance costs.

Image-based methods have piqued interest with advancements in computer vision designs, as the development of deep learning models enhanced accuracy and efficiency. The existing state-of-the-art methods mostly utilize feature extraction methods in deep learning models. They maintain a database consisting of indoor images and utilize closeness metrics to match these images with captured ones.

For example, Giovanni Fusco and James Coughlan proposed a sign recognition based solution with visual-inertial odometry [1].

Qun Niu and Mingkuan Li suggest highly automated image-based localization algorithm (HAIL), a powerful method that focuses on recognizing visual features in images [2]. HAIL operates in a distributed manner on mobile devices, eliminating the dependency on server-side processing while optimizing resource utilization, in terms of processing and storage.

Figure [1]. 3 sample test cases HAIL is used for localization of images. Distance between actual position and predicted position are also given.

Orhan Akal, Tathagata Mukherjee and others developed a position estimator system that uses 4 monocular cameras with disjoint fields of view as input Their system is based on Posenet CNN architecture with minor changes which achieves 12cm indoor accuracy [3]. Pisco has created a two-step architecture with an initial expressive descriptor stage, followed by a second finer localizing stage [4].

Another two-stage architecture is used by Chen, where in the first stage a CNN is used to find images with similar features from images recorded in a database. Second stage then uses the original and similar images to estimate the pose of the new image [5].

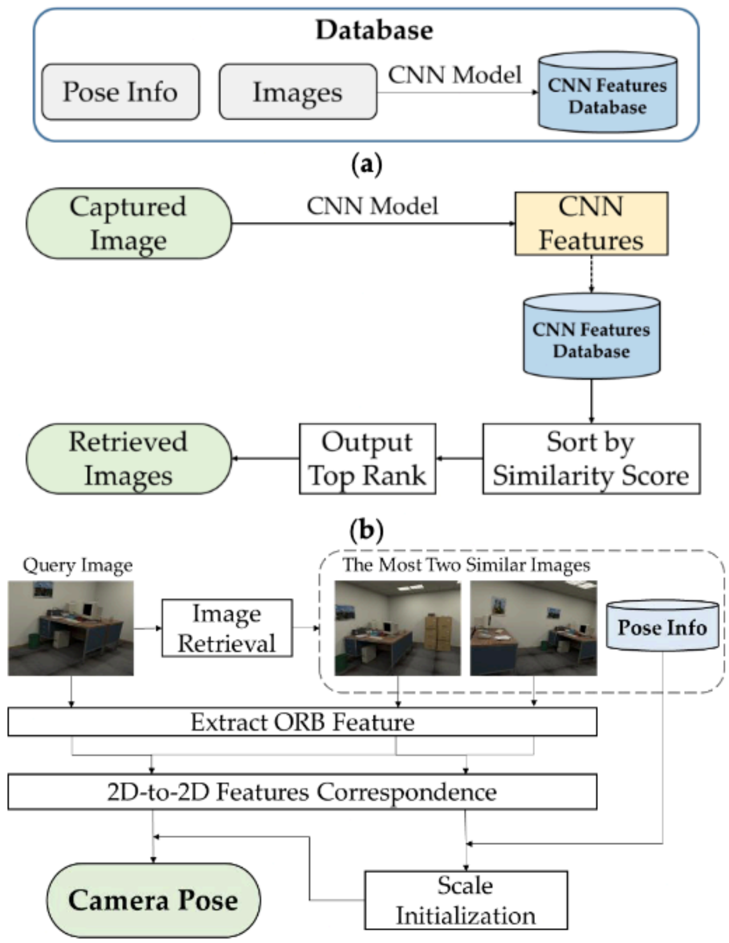

Figure [2]. Complete flow of the system including initial database generation. Then stages a and b show feature based retrieval and pose extraction respectively.

A web-server assisted system developed by Xiao, where images captured by a mobile phone are processed in a server. Yet another two stage algorithm is used where a given image is matched with specific measurement points by matching features. Then geometric transformation is extracted from point pairs to yield the position of the new image [6].

Shuang Li, Baoguo Yu and others present a novel indoor positioning system that utilizes deep belief networks [7]. By making use of image data and depth information, and additionally utilizing extracted features, this approach achieves positioning accuracy with an error of less than a meter.

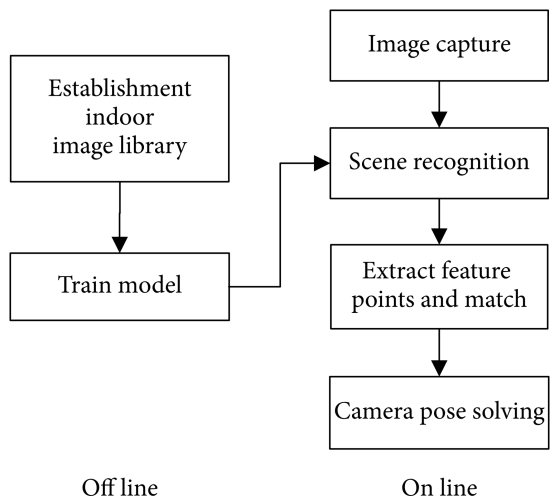

Figure [3]. The framework of the visual localization system.

In their 2023 article, Qing Li, Rui Cao and others introduce an end-to-end convolutional neural network structure [8]. They achieve high localization performance by presenting depth maps, which represent the distance of a pixel from the camera, and edge information which helps capture key features.

While these methods are powerful, the proposed project seeks to elevate the current state of image-based indoor localization by introducing innovative methodologies and addressing specific challenges encountered by existing techniques. Leveraging the advancements in computer vision and deep learning, the project aims to contribute to novel approaches that go beyond traditional feature extraction methods.

Computer vision based localization has certain important advantages over other methodologies. It doesn't require any environmental tags or expensive equipment setup (unlike wifi based methodologies) since the method only requires monocular images from a single camera, the equipment cost per localized client is also very low, only a camera and microprocessor with sufficient resources is enough. There are certain important real world applications where computer vision based localization is more preferable compared to other methodologies because of the given advantages. First one is industrial robotics. Mobile robots used in industrial applications such as factories or warehouses require reliable localization systems in order to be able to execute their tasks. Computer vision based localization is a good option for such robots due to its stated advantages and also that they already contain embedded microcontrollers which can handle our lightweight localization system. Another field of application is smartphone based locating assistance. A smartphone app that utilizes a positioning system can navigate individuals in indoor locations, allowing them to reach destinations they specify to the app. Such an application would be especially beneficial for people with visual disabilities.

3. PROPOSED SOLUTION AND METHODS

The project's aim is to design and implement a computer-based indoor localization solution that advances the current state-of-the-art approaches.

It encompasses the whole area of image-based indoor localization, including large enclosed spaces and deep learning techniques such as vision based models and feature extraction methods. However, the project does not generalize to outdoor localization, or non-vision based localization methods like GPS, sensors and WiFi.

The project can be considered complex for several reasons:

<u>No Straight-Forward Solution:</u>

Image-based Indoor Localization is a developing field with no perfect solutions. It combines several issues that can be utilized such as accuracy, hardware requirements, response time and there is no simple solution that can solve all those issues completely.

Wide-Ranging Issues:

There are several realistic constraints including technical and engineering constraints. To name a few: Limited data, hardware limitations, budget constraints and data storage capacity.

Research-Based Requirements:

Solutions are based on computer vision based deep learning algorithms. Incorporating an up to date solution requires a deeper understanding of these specialized fields.

3.1. **Objectives/Tasks**

- Literature review for better understanding of problem definition and finding state of the art techniques / methodologies.
- Search for public datasets which can be used in this project for training / testing purposes (may also include methodology development for creating our own database in implementation phase)
- Determination of candidate techniques / algorithms for implementation phase.

3.2. **Realistic Constraints**

Economic constraints of this project can be evaluated in two dimensions. Project development phase budget constraints and implementation phase cost / client budget.
Our project does not have any active budget / grant unfortunately, therefore we are planning to mainly rely on free to use resources and our universities resources available to undergraduates. Main development phase constraints caused by lack of budget are, limited custom grand truth set creation ability and limited real world testing capability and constrained computation power for training which limits model complexity and training set size.

In addition to our development phase constraints, it is also necessary to account for potential economic constraints which interested third parties may experience in a possible utilization of our navigation system. Our navigation system is most likely to be deployed on a mobile robot for robotic navigation purposes or to a smartphone for assisting human navigation. In both of these cases our algorithm will be deployed on embedded computing platforms. While such platforms with high computational power / GPU acceleration does exist, oftentimes embedded systems with such capabilities are quite expensive, therefore our navigation system should be designed to account for budget constraints of potential clients and should be able to work properly under inexpensive hardware with limited capabilities.

3.3. **Engineering/Scientific Standards**

In the development of image-based indoor localization systems, adapting the industry standards is crucial. Following IEEE standards will guide the implementation of the project:

1. **IEEE Recommended Practice for Framework and Process for Deep Learning Evaluation [9]**

   This standard is a valuable resource for the project to set guidelines that improve the reliability and performance of deep learning models.

   Abstract: This standard includes recommendations for improving and evaluating deep learning algorithms. Specifying an assessment index system and corresponding assessment process.

   Scope: Designates requirements for the field of deep learning, covering the aspects of assessment index system, assessment process and assessment of stages encapsulating demand, design, operation and implementation.

   Date of Publication: 10 April 2023

   DOI: 10.1109/IEEESTD.2023.10097701

   ICS Code: 35.080 35.240.01 - Software Application of information technology in general

   Publisher: IEEE

2. **IEEE Standard for Camera Phone Image Quality [10]**

   This standard is employed in the project as it provides a comprehensive framework for evaluating and ensuring the quality of images captured by camera-equipped mobile devices, which the model will be built upon.

   Abstract: This standard quantifies the performance of mobile devices with cameras. It establishes metrics and procedures for sensors, lenses, and signal processing routines. Metrics include spatial frequency response, color uniformity, chroma level.

   Scope: Image and video quality are addressed with this standard. It also defines standardized test methods for measuring camera phone image quality and attributes.

   Relevance: Many of the papers use generic mobile devices to test the models they developed. Some publications use mobile devices for building datasets as well. In both cases, mobile cameras become the main source of monocular RGB photographs / camera streams and IEEE phone camera quality allows standardized evaluation of data obtained in such a way.

   Date of Publication: 10 April 2023

DOI: 10.1109/IEEESTD.2023.10097701

ICS Code: 35.080 35.240.01 - Software Application of information technology in general

Publisher: IEEE

4. **RISK MANAGEMENT**

Fundamental aim of our literature survey is to observe state of the art methodologies, algorithms and system architectures utilized by recent studies for our problem. An important step of the design phase of our graduation project will be architecture design / methodology selection for solving the problem. While the ending implementation will only have a single architecture, we are planning to create a "pool" of potential techniques in the design phase since some of the candidate techniques may turn out to be unusable in the implementation phase due to either insufficient accuracy of the technique or our inadequate level of expertise in the field to implement the given technique.

Another potential risk is unavailability of computational resources. While cloud computing services may provide us some computation time for free, freely available quota may run out. In such a case, our alternative will be using our own local machines but that may potentially limit our models' quality.

5. **PROJECT SCHEDULE**

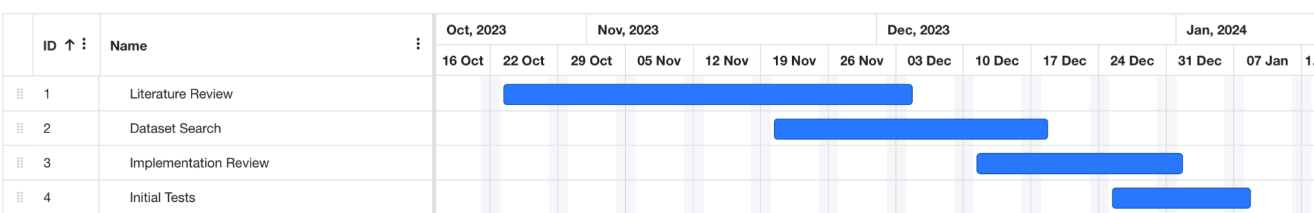

Figure [4] : Gantt chart of design phase tasks

The project plan in the Gantt chart outlines our timeline, with each team member contributing equally to every objective. Delays in one task might impact others. For instance, if implementation reviews run late, it could affect initial tests since they rely on those reviews. Similarly, if literature search is delayed, it pushes back the implementation review.

To prevent the negative impacts of delays, we started the literature search early and will handle it alongside other tasks. During the implementation review phase, we conduct searches simultaneously, and if there's a delay, we test the first found implementations without waiting for the entire review process to finish.

6. ETHICAL ISSUES

The only potential ethical problem related to the project may be privacy. Since the project involves usage of visual data of public places, protecting privacy of individuals is an important matter. Since it is practically impossible to ask consent of every individual who will appear in recorded training & test material for the project, we are ethically required to remove any potential individual identification feature from our datasets such as faces. For this purpose, we plan to implement a special face removal or blurring pre-processing stage in our data collection which will work before any other type of operation on the data, and the filtered version of the dataset will be placed in our final development output, effectively alleviating any ethical dilemmas regarding privacy.

7. **REFERENCES**

[1] Fusco, G., Coughlan, J.M. (2018). Indoor Localization Using Computer Vision and Visual-Inertial Odometry. In: Miesenberger, K., Kouroupetroglou, G. (eds) Computers Helping People with Special Needs. ICCHP 2018. Lecture Notes in Computer Science(), vol 10897. Springer, Cham. https://doi.org/10.1007/978-3-319-94274-2_13

[2] Niu, Q., Li, M., He, S., Gao, C., Gary Chan, S.-H., & Luo, X. (2019). Resource-efficient and automated image-based indoor localization. *ACM Transactions on Sensor Networks*, *15*(2), 1–31. https://doi.org/10.1145/3284555

[3] Akal, O., Mukherjee, T., Barbu, A., Paquet, J., George, K., & Pasiliao, E. (2018). A Distributed Sensing Approach for Single Platform Image-Based Localization. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. https://doi.org/10.1109/icmla.2018.00103

[4] Piasco, N. (2019). *Vision-based localization with discriminative features from heterogeneous visual data* (Doctoral dissertation, Université Bourgogne Franche-Comté).

[5] Chen, Y., Chen, R., Liu, M., Xiao, A., Wu, D., & Zhao, S. (2018). Indoor visual positioning aided by CNN-based Image retrieval: Training-free, 3D modeling-free. *Sensors*, *18*(8), 2692. https://doi.org/10.3390/s18082692

[6] Xiao, A., Chen, R., Li, D., Chen, Y., & Wu, D. (2018). An Indoor Positioning System Based on Static Objects in Large Indoor Scenes by Using Smartphone Cameras. *Sensors*, *18*(7), 2229. https://doi.org/10.3390/s18072229

[7] Li, S., Yu, B., Jin, Y., Huang, L., Zhang, H., & Liang, X. (2021). Image-Based Indoor Localization Using Smartphone Camera. *Wireless Communications and Mobile Computing*, *2021*, 1–9. https://doi.org/10.1155/2021/3279059

[8] Li, Q., Cao, R., Liu, K., Li, Z., Zhu, J., Bao, Z., Fang, X., Li, Q., Huang, X., & Qiu, G. (2023). Structure-guided camera localization for indoor environments. ISPRS Journal of Photogrammetry and Remote Sensing, 202, 219–229. https://doi.org/10.1016/j.isprsjprs.2023.05.034

[9] "IEEE Recommended Practice for Framework and Process for Deep Learning Evaluation," in *IEEE Std 2841-2022* , vol., no., pp.1-32, 10 April 2023, doi: 10.1109/IEEESTD.2023.10097701.

[10] "IEEE Standard for Camera Phone Image Quality (CPIQ)," in *IEEE Std 1858-2023 (Revision of IEEE Std 1858-2016)* , vol., no., pp.1-170, 4 Aug. 2023, doi: 10.1109/IEEESTD.2023.10205967.