ENS 491 – Graduation Project (Design)

Progress Report I

Project Title: Indoor Localization Using Camera Images

Group Number: 128

Group Members: Kağan Kağanoğlu, Yarkın Alpmen Akyosun

Supervisor(s): Mustafa Ünel

Date: 07/01/2024

Sabancı Üniversitesi

1. **PROJECT SUMMARY**

1.1 Description:

This project aims to develop an innovative solution for precise indoor localization within large enclosed spaces, such as shopping malls and airports, utilizing computer vision and deep learning techniques. The motivation behind this project lies in the limitations of traditional localization methods like GPS and Wi-Fi indoors, where signal obstruction hinders their effectiveness.

1.2 Gap in the Literature:

While several notable studies, such as Fusco and Coughlan's work on indoor localization using computer vision and visual-inertial odometry [1], Niu and Li's automated image-based localization algorithm (HAIL) [2], and Akal et al.'s distributed sensing approach for single-platform image-based localization [3], have contributed significantly to the field, there remains a distinct gap that our project seeks to address.

The existing literature showcases various approaches, from discriminative feature-based localization [4] to two-stage architectures involving image retrieval and pose estimation [5]. However, a comprehensive solution that combines the strengths of computer vision and deep learning for accurate indoor localization is yet to be fully realized.

Moreover, while Li, Yu, and others have explored image-based indoor localization using deep belief networks [6], and Li, Cao, and others introduced an end-to-end convolutional neural network structure [7], there is room for further innovation and advancement in methodologies that enhance accuracy and efficiency.

1.3 Motivation:

With recognizing the limitations of current solutions, especially in large enclosed spaces like shopping malls and airports, the project aims to contribute to the existing body of knowledge by introducing new methodologies and addressing specific challenges encountered by current techniques. Through this project, we intend to bridge the gap in the literature by pushing the boundaries of accuracy, efficiency, and accessibility in indoor localization using camera images.

1.4 Objectives and Intended Results:

The primary objectives of the project include conducting a comprehensive literature review to understand the current state of the field, searching for relevant datasets, and selecting suitable algorithms to come up with a design for implementation. We acknowledge the complexity of the project, considering factors like accuracy, hardware requirements, and response time. Realistic constraints, such as economic limitations and research-based requirements, guide our approach.

1.5 Fundamental Elements:

The project will adhere to fundamental elements, including thorough literature synthesis, analysis of existing solutions, construction of an advanced indoor localization system, rigorous testing, and comprehensive evaluation. By incorporating IEEE standards for deep learning evaluation and camera phone image quality, we aim to ensure the reliability and performance of our proposed solution.

In summary, our project aims to push the boundaries of image-based indoor localization by introducing innovative methodologies, addressing gaps in the existing literature, and adhering to realistic constraints. The intended results include the design and implementation of an advanced indoor localization system that can be practically applied in various scenarios, ranging from industrial robotics to smartphone-based location assistance.

2. **SCIENTIFIC/TECHNICAL DEVELOPMENTS**

2.1 Design

In response to the challenges posed within indoor localization using camera images, the project proposes designs and ideas that will be implemented and tested during the flow of the project. As test and experiment procedure takes place, these designs will be evaluated and compared to discern their strengths, limitations, and overall performance. Through assessment and comparison, the project aims to find optimal solutions, shedding light on the most effective strategies for accurate indoor localization in diverse and dynamic environments.

2.1.1 Design A

This design combines ideas from the papers "Structure-guided camera localization for indoor environments" by Li, Cao, et al., and "Image-Based Indoor Localization Using Smartphone Camera" by Li, Yu, et al. The architecture incorporates different branches to leverage powerful feature extraction methods and Deep Belief Networks (DBN), promoting effective fusion of diverse features.

Methodology

Input Processing and Feature Extraction:

This part aims to Utilize Depth-Weighted Input Residual Fusion to incorporate multi-modal information, giving emphasis to different levels based on depth. Then, with the combination of Edge Attention-Based Feature Residual Fusion with Local Binary Pattern (LBP) Feature Extraction Techniques, capture both high-level semantic features and texture-based information.

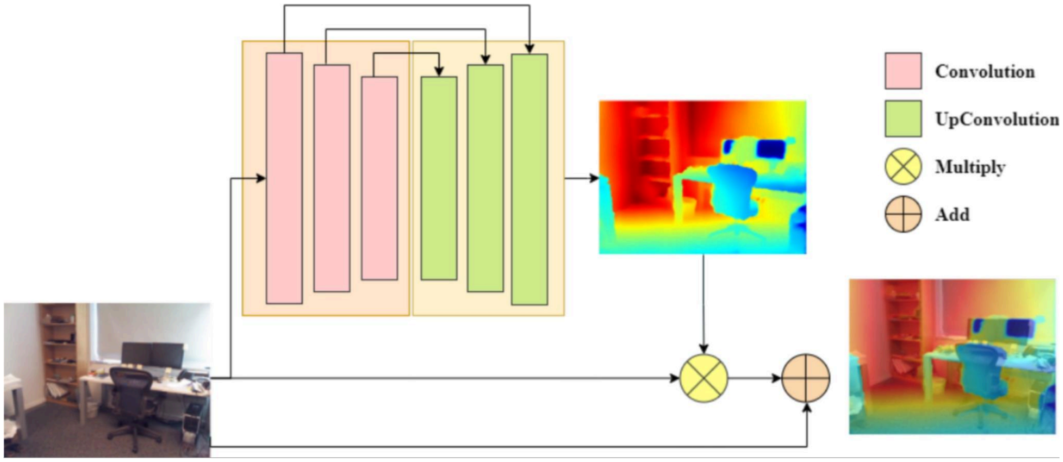


Figure [1]: Visualization of Depth Weighted Residual Fusion

Neural Network Architecture:

A Deep Belief Neural Network (DBN) is implemented to harness its unsupervised learning capabilities for feature representation. The model undergoes unsupervised layer-by-layer pre-training, effectively initializing weights. After pre-training, transfer learning is applied using a network pre-trained on a related dataset to boost performance. Self-attention is further incorporated into the neural network architecture to enable the model to weigh different features dynamically.

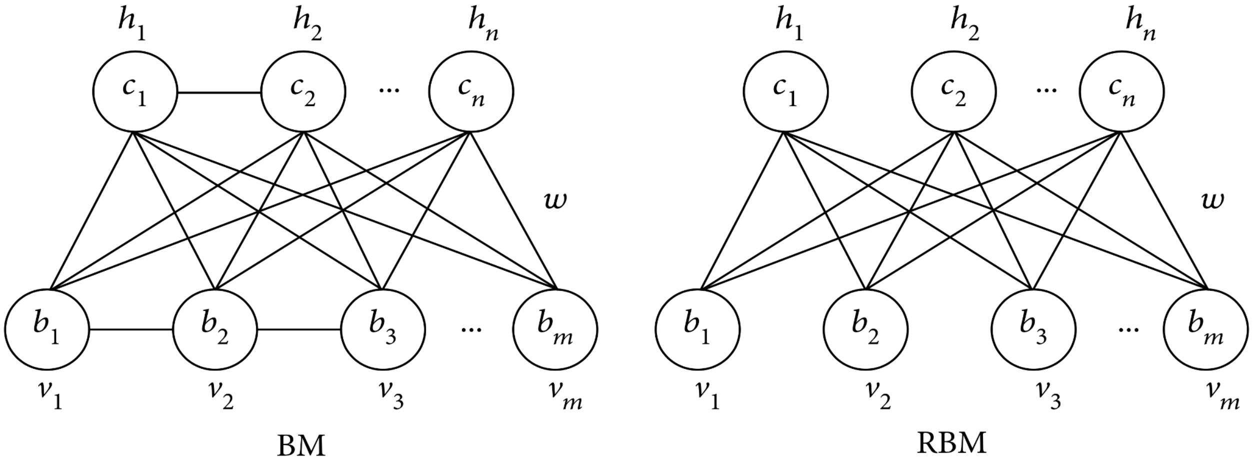


Figure [2]: Deep belief network consisting of a multi restricted Boltzmann machine (RBM) and a backpropagation (BP) neural network.

Ensemble:

Utilizing ensemble learning, the design aims to improve robustness by combining outputs from different branches. A specific ensemble technique, such as a weighted average or stacking, will be employed.

Flow Chart:

The features extracted during input processing, especially through Depth-Weighted Residual Fusion and Feature Residual Fusion, directly contribute to the subsequent DBN architecture, enhancing the network's ability to learn diverse features.

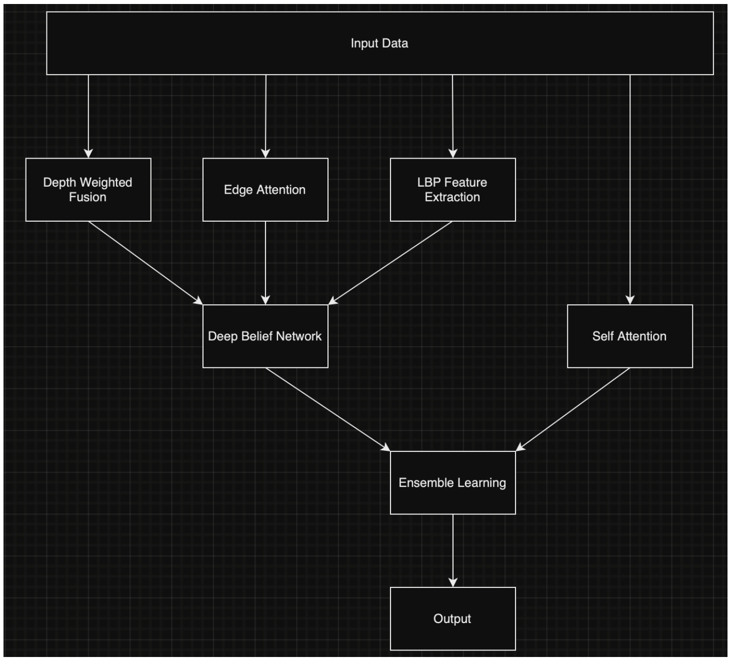


Figure [3]: Flow Chart of Design A

2.1.2 Design B

Our second design uses a different approach. Image retrieval based methodology works by extracting the most similar image from the training dataset to the given input image. Then the positional difference between two images is computed in order to find the final position. One paper employing this methodology is [5]. We also employ high level feature extraction with CNN [9]. We then utilize depth weighted fusion we have previously discussed in order to compute actual coordinates.

Methodology

Retrieval:

By using a CNN model trained for matching based on similarity as in [5], we retrieve a closely related image to the given input image. This allows us to focus on a potentially close location to the location of the target image.

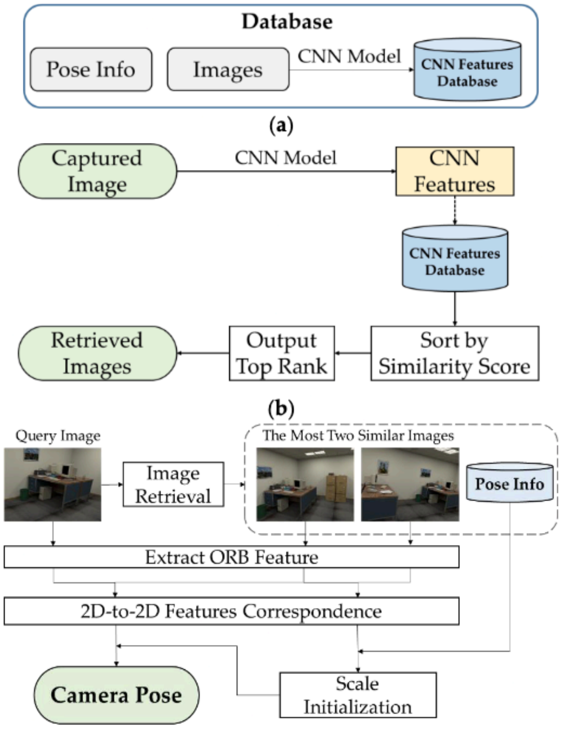


Figure [4]: Image Retrieval

High Level Feature Extraction:

Second step is high level feature extraction. This step uses another CNN model similar to one in [5] to extract high level object determination. This calculation is only applied on input images in the test phase as dataset images will be pre-extracted to avoid redundant computation.

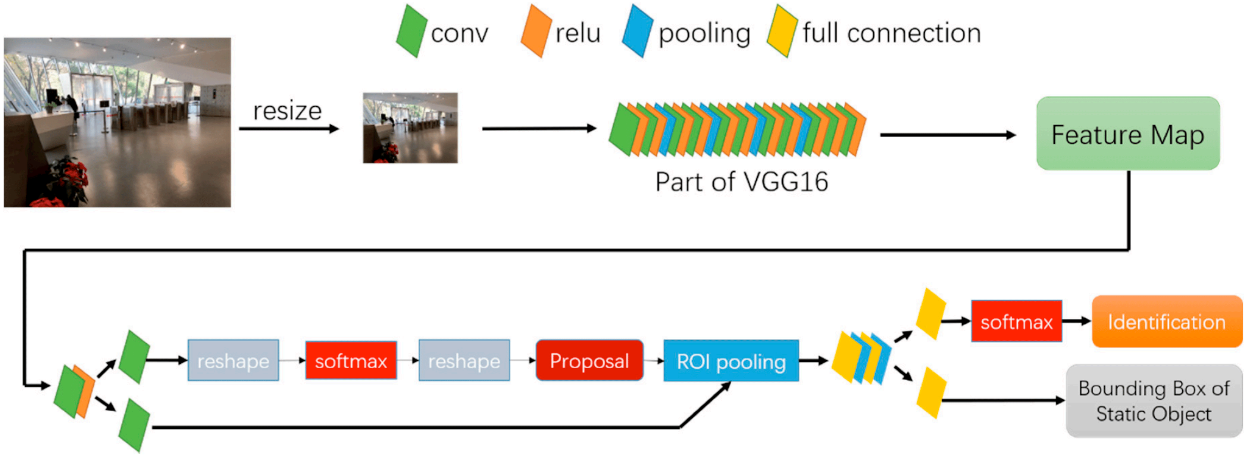


Figure [5]: CNN model pipeline employed in paper [5], our model uses CNN for high level object extraction only, rest of the position determination is carried out by DWF based calculations.

Position Determination Using Depth Weighted Fusion:

Final step is to utilize DWF to find the position difference between the queried and given image to calculate the absolute position of the input image. DWF and high level boundaries are used together to determine depth differences of objects between query and input image. By selecting an average distance for every extracted object from DFW, we reach a system of equations, where a redundant number of features allows elimination of some portion of the uncertainty, and the algorithm converges into a position difference. Then the given difference is simply combined with the position of the queried image to calculate the final position.

2.2 Test & Experiment

During our literature review, we have observed several performance metrics employed in assessment of monocular indoor localization models. Since the fundamental aim of indoor localization models is to retrieve indoor position, translation (position) and orientation errors are the most important evaluation metrics. Median translation and orientation errors are presented very frequently. Metrics regarding execution times are also included on some papers. Time required for position prediction over a sample is stated in FeatLoc: Absolute pose regressor for indoor 2d sparse features with simplistic view synthesizing [8]. Chen et al. also includes memory requirements of the database (on deployment) of the model they developed [5].

Models developed are benchmarked against train samples of datasets. While benchmark of developed model yields its metrics, raw metrics are oftentimes not sufficient to assess model. Comparing developed models with other successful models in literature allows relative assessment of the model and therefore elaboration of potential contribution of given work to literature. Usage of publicly available datasets commonly benchmarked in literature allows compatible evaluation of different models. Two such datasets are 7-scenes and 12-scenes. These datasets consist of a series of data of (per) multiple indoor locations. Each series per location consists of a monocular 24 bit RGB image, 16 bit depth image (where distance unit is millimeters) and matrix of a position descriptor (6DOF, position and orientation in 3D). Datasets are also pre-splitted into train-test samples.

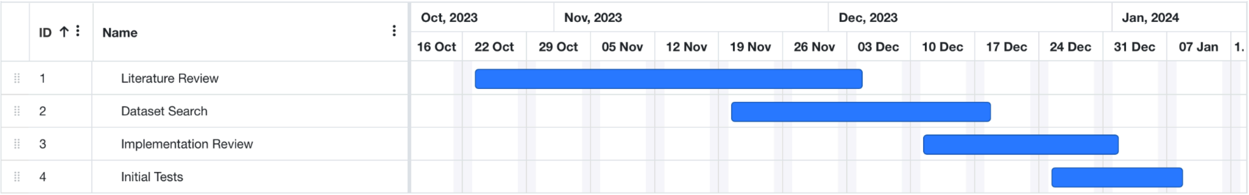
There are many papers which devise new models for indoor localization with monocular images. As in any other branch, some papers contain better novel ideas with highly desirable metrics. As every research aims to introduce incremental improvement in the field, oftentimes authors compare their models with such state-of-art ones. In our literature review we have observed such papers that are frequently benchmarked against new models. Common notions of given papers are that they employ 7-scene and/or 12-scene datasets for their benchmarks. Therefore they are compared with proposed models for assessing their performance.

We have determined an experimentation / evaluation pipeline for our own model which will allow us to relativistically evaluate our model to existing literature and optimize our design by experimenting with different topologies/components. 7-scene and 12-scene datasets will be used in development/benchmarking because of advantages stated above. We will experiment our model in two phases, optimisation and final benchmark. Optimisation involves testing different designs/components only with a training set (by sub-splitting the train set) and benchmarking different potential candidate designs against each other. Then the final step will be to benchmark our model using a test portion of datasets to compare final performance to existing literature. In both phases, translation and orientation errors will be the fundamental metrics. Time for position estimation of a single frame will also be recorded and reported.

3. **ENCOUNTERED PROBLEMS**

3.1 Adaptations and Challenges in Project Execution

The project has encountered some challenges that prompted adjustments to the initial plan and goals. One significant factor contributing to these changes is the complexity of the image-based indoor localization domain. As we delved deeper into the literature of computer vision and deep learning, we identified the need for a more nuanced and adaptive approach than initially anticipated. Although there were some setbacks, the original goals remain relevant, with necessary modifications.



ID	Name
1	Literature Review
2	Dataset Search
3	Implementation Review
4	Initial Tests

Considering the project timetable, there were some delays, primarily during the initial tests phase. The delays in the initial tests phase include unexpected technical difficulties, which consumed more time than anticipated. Additionally, the group members, although dedicated to the project, had to contend with high workloads, from ongoing classes, homework assignments and exams. Balancing academic responsibilities with the demands of the project introduced an additional layer of complexity.

3.2 Corrections Planned

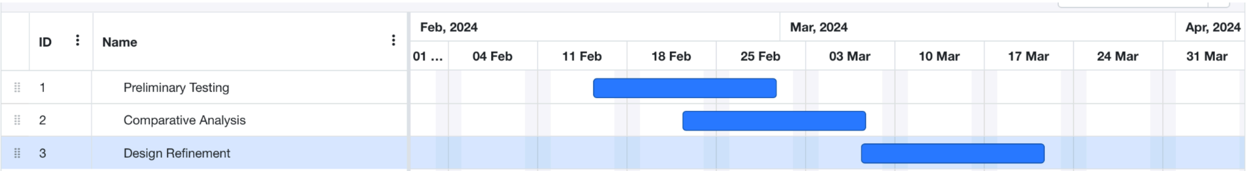
To address the setbacks and realign with the project timetable, the team is implementing corrective measures. These measures include reassessment of the timetable, with realistic deadlines, task prioritization, where team members are expected to allocate more time for more crucial tasks, and reviewed workload distribution, considering individual strength and availability of team members.

3.3 Effects of the Changes

The changes made in the project have a few important effects. First off, the new schedule is more practical and easier to follow, considering the challenges that were found in the literature and the task prioritization. This helps with managing the time and resources better. Being more adaptable lets us handle unexpected issues more effectively. Also, the reviewed workload gives us a deeper understanding of how indoor localization with images works. The updated plan aims to prepare us for potential issues in advance.

4. TASKS TO BE COMPLETED UNTIL PROGRESS REPORT II

Aligned with the capstone design process phases, the upcoming focus revolves around testing and refining proposed designs. The iterative process includes comprehensive testing of each design, refining based on results, comparative analysis, and optimization for out-of-sample data. The final design will be selected after thorough testing, and documentation will detail the testing procedures, design refinements, and the rationale behind the final design choice. The aim is to ensure the selected design is robust, adaptable, and consistently effective across diverse scenarios.



ID	Name
1	Preliminary Testing
2	Comparative Analysis
3	Design Refinement

Figure [4]: Gantt chart of testing phase

5. **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] 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

[7] 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

[8] Bach, T. B., Dinh, T. T., & Lee, J.-H. (2022). FeatLoc: Absolute pose regressor for indoor 2d sparse features with simplistic view synthesizing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *189*, 50–62. https://doi.org/10.1016/j.isprsjprs.2022.04.021

[9] 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