

Self-Localization of Guide Robots Through Image Classification

Muhammad S. Alam^{1,3}  , *Farhan B. Mohamed*^{*1,2}  , *AKM B. Hossain*^{1,4}  

¹Department of Emergent Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.

²Department of Media and Game Innovation, Centre of Excellence (MaGICX), Universiti Teknologi Malaysia, Johor Bahru, Malaysia.

³Department of Computer Science, University of Bisha, Bisha, Saudi Arabia.

⁴Department of Information Systems, University of Bisha, Bisha, Saudi Arabia.

*Corresponding Author.

PARS2023: Postgraduate Annual Research Seminars 2023.

Received 24/09/2023, Revised 10/02/2024, Accepted 12/02/2024, Published 25/02/2024



© 2022 The Author(s). Published by College of Science for Women, University of Baghdad.

This is an Open Access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The field of autonomous robotic systems has advanced tremendously in the last few years, allowing them to perform complicated tasks in various contexts. One of the most important and useful applications of guide robots is the support of the blind. The successful implementation of this study requires a more accurate and powerful self-localization system for guide robots in indoor environments. This paper proposes a self-localization system for guide robots. To successfully implement this study, images were collected from the perspective of a robot inside a room, and a deep learning system such as a convolutional neural network (CNN) was used. An image-based self-localization guide robot image-classification system delivers a more accurate solution for indoor robot navigation. The more accurate solution of the guide robotic system opens a new window of the self-localization system and solves the more complex problem of indoor robot navigation. It makes a reliable interface between humans and robots. This study successfully demonstrated how a robot finds its initial position inside a room. A deep learning system, such as a convolutional neural network, trains the self-localization system as an image classification problem. The robot was placed inside the room to collect images using a panoramic camera. Two datasets were created from the room images based on the height above and below the chest. The above-mentioned method achieved a localization accuracy of 98.98%.

Keywords: Convolutional Neural Network, Deep Learning, Guide Robot, Image Classification, Self-localization.

Introduction

The use of robots in various fields of artificial intelligence and computer vision has increased significantly in recent years. Among them, the guide robot is one that is worth mentioning, as it plays a very important role in navigating blind people and providing information. It guides people from one point to another, both indoors and

outdoors. This is useful for blind as well as sighted people. The robot must know the initial position to determine the optimal path. For example, on the floor of a building, if a robot is situated inside a room, it can decide the path to another room if the initial room position is known. A robot finds its own location and orientation in relation to its

environment through a process known as self-localization. The ability to operate efficiently and securely in changing surroundings is a crucial prerequisite for autonomous mobile robots. Expensive and sophisticated sensor systems such as GPS or laser range finders are frequently used in conventional self-localization techniques.

A glove specially designed for the benefit of people who are blind or visually impaired is likely to include navigational aids or other features to improve their everyday lives and mobility¹. The self-localization of guide robots was the subject of the deep-learning approach suggested by Smith and Johnson in 2022. This strategy produced encouraging outcomes and concentrated on enhancing the robots' capacity to ascertain their own position in varied surroundings². Creating a path-planning system suitable for indoor environments for visually impaired people³. This study looks at creative ways to improve accessibility and mobility for blind people. A particular robotic device has been released for blind people by the more famous automated organization ARIEL. When walking the visually impaired people with the computerized device, ten ultrasonic sensors produce the signals and help the blinds⁴.

The diverse user connection research has proposed a novel approach for autonomous shopping carts⁵. The various interactions of the shopping cart attract the attention of the customers. It's also a new experience for the customers. Another invention of the guided robotic system is the acoustic reflector localization system (ARLS), which is produced by using the localization of an acoustic reflector⁶. It identifies the source of the image and predicts the reflector position.

Materials and Methods

The proposed research uses an image classification system that thinks about a guide robot to localize indoor locations properly. The main objective is to help the visually impaired people of the indoors for accurate navigation¹⁶. An indoor robot navigation system must find a robot's precise position in the environment. In indoor environments, traditional approaches to robot localization, such as GPS or

A probabilistic framework created for a robot that functions in a room-like setting⁷. The robot uses semantic data about user activity to build an accurate metric map of the space. The motions of the user and robot were noticeably in completely different directions. The robot effectively and successfully locates the user within a room using this method. A convolutional neural network (CNNs)-based visual location identification technology was proposed⁸. It showed how CNNs might be used to extract distinguishing characteristics from images⁹⁻¹², enabling a robot to identify previously visited locations and self-localization. Deep learning techniques have been applied in vision-based localization approaches. To accomplish precise self-localization, even in difficult and dynamic situations, deep convolutional networks with pose regression are used⁹. Image-based indoor localization^{13,14} techniques use both visual odometry and deep neural networks. This illustrates how deep learning may help mobile robots to self-localize accurately¹⁵.

The primary objective is to create a deep learning model that can localize an indoor robot from surrounding images, which is a difficult image classification problem. Our main objective is to create and deploy an innovative and efficient deep learning model that allows a guiding robot to independently establish its own location inside a region using image classification techniques. This innovative solution makes use of the capabilities of computer vision and machine learning algorithms to allow the guide robot to understand its surroundings and correctly identify its position without relying on external infrastructure, such as GPS or beacons.

laser-based techniques, have limitations, making image classification a worthwhile alternative.

Guide Robot Concept

In today's world, an estimated 283 million visually impaired people suffer from considerable mobility and travel obstacles. Navigating unfamiliar places is difficult for these people. However, in recent years,

there has been a surge in the development of assistive robotic devices designed to assist visually impaired or blind people.

In an unusual approach, when a robot enters a room, a comprehensive 3D map of the surroundings is built using RGB-D video. Owing to this procedure, the robot can learn everything about the environment in which it works. The robot can precisely identify its present location inside the space by selecting certain frames from the recorded video and combining them with the 6D camera position data. After identifying the room and

determining the viability of several routes leading to the intended location, the robot applied image classification¹⁷.

The system's performance depends significantly on the accuracy and quality of the created 3D map. Four RGB-D sensors were positioned on top of the robot to provide a wide-angle and thorough input. With this configuration, the robot can observe its surroundings from various angles, which enhances its ability to reliably and robustly interpret its surroundings.

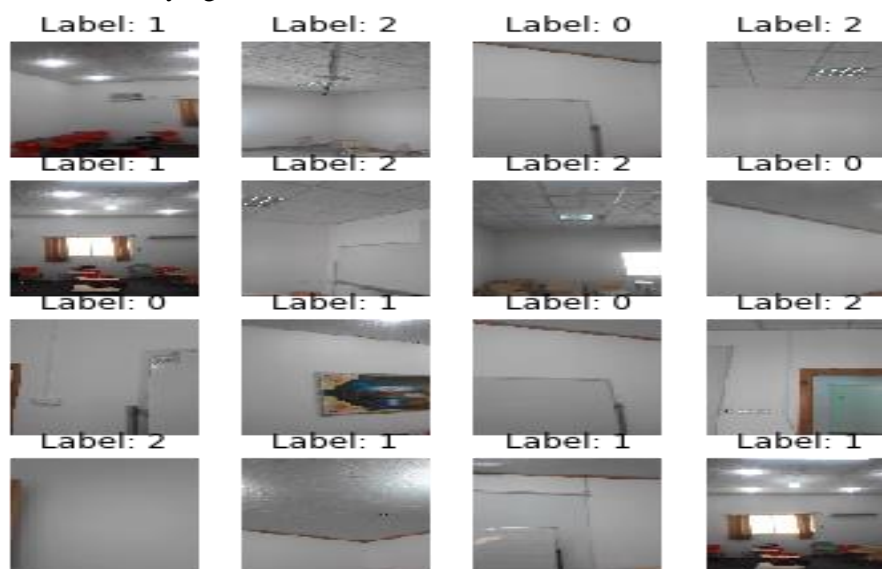


Figure 1. Sample rooms images.

Dataset

Gathered a dataset of annotated images captured by the robot's camera in indoor settings to train and evaluate the model. The collection includes images of several rooms in the structure labeled with the associated ground truth location coordinates. The dataset consists of several feature vectors taken from various types of room images. Each class of image depicts a certain type of room environment. Sample images data as shown in Fig.1. The dataset was divided into three subsets, each of which included images from the same three rooms: Rooms 1, 2, and 3. These rooms are all at the same level as the building. Took two recordings from the inside of each room, each lasting 30 seconds, to ensure complete coverage. The first video was captured above the chest level, focusing particularly on the upper portion of the environment, including the ceiling. The second video was captured below chest

height and covered the entire floor. Gathered a large number of images from three separate videos, totaling over 20 K images for the dataset.

Model Architecture

Used a deep learning-based strategy for self-localization using image classification. Given the impressive performance in image identification tasks, convolutional neural networks (CNNs) are an appropriate solution. The CNN architecture designed numerous convolutional layers with fully connected layers. The final output layer was created based on the identified locations to forecast the coordinates of the robot's position.

Model Training

The proposed CNN model was trained on the collected dataset using supervised learning. The

Loss functions were minimized when training the model. An optimization method is used while training the model, one of which is the hyperparameter optimization. The hyperparameter optimization modifies the model weights, batch size, number of epochs, learning rate etc. Also use a GPU to speed up computation and reduce the model training time.

Evaluation Metrics

To evaluate the performance of the self-localize guide robotic system, the training model was evaluated on two different datasets. The trained model was evaluated using several evaluation

metrics, such as precision, recall, and F-1 score. Thus, the accuracy of the model in determining a more accurate position of the guide robot was determined. A qualitative evaluation was performed to analyze how accurately the guide robot was localized in its environment. A proposed image classification system, a deep learning system for a guide robot, is finding the initial location of a guide robot and identifying the room correctly. The image classification model accuracy shows the result is more accurate and acceptable.

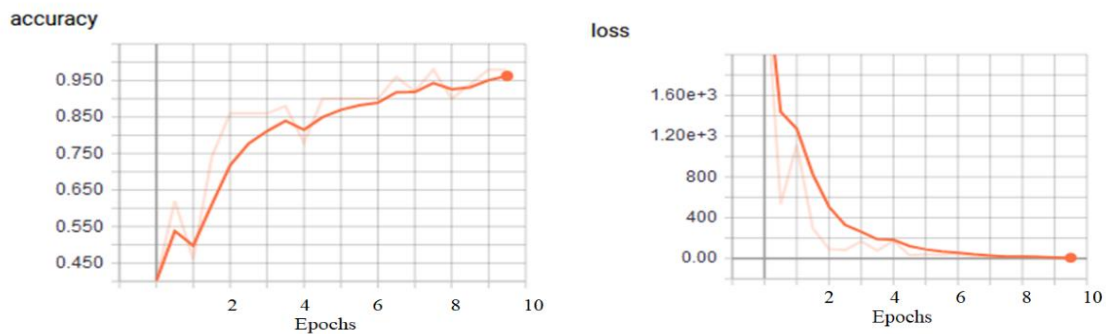


Figure 2. Training accuracy and training loss of dataset made above the chest height of room images

Results and Discussion

The objective of this study is to develop an image classification system for a guide robot to find its accurate position inside a room. The guide robot inputs the required images through the panoramic camera and finds the position by processing the photos. Convolutional neural network (CNN) is a deep learning strategy to extract the image features and classify the images. In this way, a guide robot finds the room position.

Experimental Results

This research study trains a deep learning model and finds a robot's accurate location inside a room. The main focus is for a robot to know its position in a room through an image classification process. The necessary image collected on the inside camera and featured extracted through the convolutional neural network.



Figure 3. Training accuracy and training loss of dataset made below the chest height of room images

Two self-created indoor datasets evaluated the proposed deep learning-based image classifier. The first dataset consisted of images from above the chest height of the room, and the second dataset consisted of images from below the chest height of the room. There were three groups of images in each dataset, including rooms 1, 2, and 3. The training, testing, and validation images were included in each dataset. The first dataset was used to train the model. The accuracy of the model is 95.74%. Fig. 2 illustrates the accuracy and training loss of the model. After running the deep learning model on the second dataset, the model accuracy was 98.98%. Fig. 3 shows the model accuracy graph and training loss. The model accuracy is substantially higher for the second dataset because it contains a lower part of the room images with diverse furniture and can classify the rooms more correctly.

Model Evaluation

The classifier's evaluation metric are the Precision, Recall, and average F1-score used to assess a classification model's performance. The following is its definition:

Precision is defined as accurate positive predictions divided by all predicted positives. For the calculation of precision, use the following Eq. 1:

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots 1$$

Recall is the ratio of accurate positive predictions to all actual positive predictions. To calculate recall, use Eq. 2 below:

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots 2$$

The F1-score is a single score that balances precision and recall, calculated as the balanced average of both measures. The following Eq. 3 is used to compute it:

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots 3$$

where TP is the true positive, FP is the false positive, FN is the false negative. The accuracy metric normally compares 1 and 0 and determine how many correct. Our classifier average F1-score comes in 1 that means classify the rooms properly and robot identify a room more accurately. The model's evaluation metric is shown in Table 1.

Table 1. Model evaluation metric

	Precision	Recall	F1-score	Support
Room 1	1.00	1.00	1.00	5
Room 2	1.00	1.00	1.00	2
Room 3	1.00	1.00	1.00	2
Avg. /Total	1.00	1.00	1.00	9

Discussion

This has various benefits for guiding robots to use image classification for self-localization. The robot can pinpoint its location without using external sensors by evaluating real-time video data to distinguish its surroundings properly. This strategy lowers hardware complexity and cost while offering strong and flexible localization capabilities. Occlusions and shifting illumination conditions, however, may lead to difficulties in reducing categorization accuracy. Despite these obstacles, image-based self-localization offers significant potential to augment the autonomy and

effectiveness of guide robots in challenging circumstances. This is because of developments in deep learning and computer vision. According to our findings from this study, multiplicative textures cause room images below chest height to vary more, and those above chest height to vary less. However, it is more likely that images beyond chest height will have fewer distinctive features. The key result of the experiment was that implanted the model in a robot, which accurately recognized a room from the images of the room.

Conclusion

This study investigated how a robot's self-localization could be guided by image classification. The robot showed a promising ability to detect its surroundings and establish its position by utilizing cutting-edge machine-learning algorithms. To enhance the technique's precision and flexibility to actual situations, further refinement and validation are required. Future research might improve the guide robot's ability to

self-localize through image categorization by investigating the integration of sophisticated deep-learning approaches, such as deep reinforcement learning or attention processes. Additionally, using multimodal sensor fusion, such as fusing information from cameras and range sensors, such as LIDAR, may enhance the robot's spatial awareness and navigational skills.

Acknowledgment

The authors thank the Faculty of Computing at Universiti Teknologi Malaysia in Malaysia for their

assistance.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for

- re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Universiti Teknologi Malaysia.

References

1. Faseeha U, Ghazal S, Rehmani B, Rafique R, Taba S. Impaired Glove for Blind and Impaired Person. *KJCIS*. 2020 Jan 1;3(1):07-07. <https://kjcis.kiet.edu.pk/index.php/kjcis/article/view/30/25>
2. Shaikh B, Faraz SM, Jafri SR, Ali SU. Self-localization of mobile robot using map matching algorithm. *J. Eng. Appl. Sci.* . 2021;40(1):69-77.
3. Najjar AB, Al-Issa AR, Hosny M. Dynamic indoor path planning for the visually impaired. *J. King Saud Univ. - Comput. Inf. Sci.* . 2022 Oct 1;34(9):7014-24. <https://doi.org/10.1016/j.jksuci.2022.03.004>
4. Jiang Y, Pan X, Li K, Lv Q, Dick RP, Hannigan M, Shang L. Ariel: Automatic wi-fi based room fingerprinting for indoor localization. In *Proceedings of the 2012 ACM conference on ubiquitous computing* 2012 Sep 5 (pp. 441-450). <https://dl.acm.org/doi/abs/10.1145/2370216.2370282>
5. Ryumin D, Kagirow I, Axyonov A, Pavlyuk N, Saveliev A, Kipyatkova I, Zelezny M, Mporas I, Karpov A. A multimodal user interface for an assistive robotic shopping cart. *Electronics*. 2020 Dec 8;9(12):2093. <https://doi.org/10.3390/electronics9122093>
6. Rosa S, Lu X, Wen H, Trigoni N. Leveraging user activities and mobile robots for semantic mapping and user localization. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 2017 Mar 6; (pp. 267-268). <https://doi.org/10.1145/3029798.3038343>
7. Remaggi L, Jackson PJ, Coleman P, Wang W. Acoustic reflector localization: Novel image source reversion and direct localization methods. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2016 Dec 1;25(2):296-309. <https://doi.org/10.1109/TASLP.2016.2633802>
8. Camara LG, Gäbert C, Přeučil L. Highly robust visual place recognition through spatial matching of CNN features. *ICRA*. 2020 May 31; (pp. 3748-3755). IEEE. <https://doi.org/10.1109/ICRA40945.2020.9196967>
9. AL-TAMEEMI, M. I. RMSRS: Rover Multi-purpose Surveillance Robotic System. *Baghdad Sci. J.* .2020; 17(3 (Suppl.)): 1049-1049. [https://doi.org/10.21123/bsj.2020.17.3\(Suppl.\).1049](https://doi.org/10.21123/bsj.2020.17.3(Suppl.).1049)
10. Alsaedi, E. M., & kadhim Farhan, A. . Retrieving Encrypted Images Using Convolution Neural Network and Fully Homomorphic Encryption. *Baghdad Sci. J.* . 2023; 20(1): 0206-0206. <https://doi.org/10.21123/bsj.2022.6550>

11. Hasan, A. M., Qasim, A. F., Jalab, H. A., & Ibrahim, R. W. Breast cancer MRI classification based on fractional entropy image enhancement and deep feature extraction. *Baghdad Sci. J.* .2023; 20(1): 0221-0221. <https://doi.org/10.21123/bsj.2022.6782>
12. Wenzel P, Schön T, Leal-Taixé L, Cremers D. Vision-based mobile robotics obstacle avoidance with deep reinforcement learning. *ICRA.* 2021 May 30 ;(pp. 14360-14366). <https://doi.org/10.1109/ICRA48506.2021.9560787>
13. Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. *Multimed. Tools. Appl.* . 2018 May; 77:10437-53. <https://link.springer.com/article/10.1007/s11042-017-4440-4>
14. Alam MS, Mohamed FB, Selamat A, Hossain AB. A review of recurrent neural network based camera localization for indoor environments. *IEEE Access.* 2023 May 2. <https://doi.org/10.1109/ACCESS.2023.3272479>
15. Alam MS, Hossain AK, Mohamed FB. Performance Evaluation of Recurrent Neural Networks Applied to Indoor Camera Localization. *IJETAE.*2022. https://doi.org/10.46338/ijetae0822_15
16. Vásquez BP, Matía F. A tour-guide robot: Moving towards interaction with humans. *Eng. Appl. Artif. Intell.* . 2020 Feb 1; 88:103356. <https://doi.org/10.1016/j.engappai.2019.103356>
17. Li J, Li W, Sisk A, Ye H, Wallace WD, Speier W, Arnold CW. A multi-resolution model for histopathology image classification and localization with multiple instance learning. *Comput. Biol. Med.* . 2021 Apr 1;131:104253. <https://doi.org/10.1016/j.compbiomed.2021.104253>

تحديد الموقع الذاتي للروبوتات الموجهة من خلال تصنيف الصور

محمد س. العالم^{3,1}، فرحان ب. محمد^{2,1}، أم ب. حسين^{4,1}

¹قسم الحوسبة الناشئة، الجامعة التكنولوجية الماليزية، جوهور باهرو، ماليزيا.

²مركز التميز لايتكار الوسائط والألعاب (MaGICX)، جامعة التكنولوجيا الماليزية، جوهور باهرو، ماليزيا.

³قسم علوم الحاسب، جامعة بيشة، بيشة، المملكة العربية السعودية.

⁴قسم نظم المعلومات، جامعة بيشة، بيشة، المملكة العربية السعودية.

الخلاصة

خلال السنوات الأخيرة، هناك تقدم كبير في تطوير أنظمة الروبوتات الذاتية، والذي يسمح لها بأداء مهام معقدة في سياقات مختلفة، أحد هذه التطبيقات هو استخدام الروبوتات الموجهة لمساعدة الناس الذين لديهم ضعف بصري، ولإنهاء هذه المهمة بنجاح فهناك ضرورة لتحديد الموقع الذاتي الدقيق للروبوتات الموجهة في محيطها، تقوم هذه الدراسة بتقديم طريقة لتحديد الموقع الذاتي للروبوتات الموجهة، ولهذا الإجراء تم تدريب شبكة عصبية تلافيفية على مجموعة من الصور المكانية المستلمة من ناحية الروبوت، إن تحديد الموقع الذاتي للروبوتات الموجهة الذي يعتمد على تصنيف الصور يوفر حلاً دقيقاً للتنقل في محيطات أو بيئات داخلية مختلفة، تنفيذ هذا النظام بشكل ناجح قد يمهد الطريق لتطبيقات روبوتات ذاتية متطورة بشكل كبير والتي ستساعد المجتمع من خلال تقديم تفاعلات أكثر أماناً وكفاءة بين الإنسان والروبوتات، تقوم هذه الدراسة بتوضيح كيفية تحديد موقع الروبوت بنفسه داخل أي غرفة (محيط)، كما قمنا كباحثين ببناء مشكلة تحديد الموقع كأنها مشكلة لتصنيف الصور وقمنا بتدريب شبكة عصبية تلافيفية لحل المشكلة، حيث يقوم الروبوت بالتقاط صور بانورامية من داخل الغرفة ويقوم بالتعرف على الغرفة لتحديد موقعه، ومن خلال البيانات التي تم إنشاؤها في هذه الدراسة، تبين أن نسبة الدقة كانت 98.98%.

الكلمات المفتاحية: الشبكة العصبية التلافيفية، التعلم العميق، الروبوت الموجه، تصنيف الصور، تحديد الموقع الذاتي.