Using images to avoid collisions and bypass obstacles in indoor environments

David Silva de Medeiros*, Thiago Henrique Araújo*,
Elias Teodoro da Silva Júnior† and Geraldo Luis Bezerra Ramalho†
* † Federal Institute of Education, Science and Technology of Ceará,
Fortaleza, CE, Brazil
Email: *{david.silva.medeiros05, thiago.henrique.araujo07}@aluno.ifce.edu.br † {elias, gramalho}@ifce.edu.br

***Abstract*—Convolutional Neural Network (CNN) has contributed a lot to the advancement of autonomous navigation techniques, and such systems can be adapted to facilitate the movement of robots and visually impaired people. This work presents an approach that uses images to avoid collisions and bypass obstacles in indoor environments. The constructed dataset uses information from forward and lateral speeds during walks to determine collisions and obstacle avoidance. VGG16, ResNet50, and Dronet architectures were used to evaluate the dataset. Finally, reflections on the dataset characteristics are added, and the CNNs performance is presented.**

I. Introduction

Autonomous navigation systems have been studied to find ways to achieve maximum performance and safety, and the advancement of Machine Learning techniques has contributed to improving such systems. Independent navigation systems can be expanded to facilitate the movement of robots and people who have visual impairments.

Vision is people's principal sense of orientation and navigation. Losing it totally or partially implies difficulties in performing many tasks. The World Health Organization (WHO) estimates that there are 252.6 million people with visual impairment, of which 36 million have a total visual impairment, and 216.6 million have mild or moderate impairment [1].

Among the problems that visually impaired people face, mobility is one of the most studied by researchers, because it allows greater independence to navigate indoors and outdoors. Some strategies have been proposed using cameras as wearable devices to help the visually impaired to be safe while walking, as shown in [2] and [3].

This work presents an approach that uses images to avoid collisions and bypass obstacles in indoor environments. The proposed method consists of: (1) Capture of walking images; (2) Image labeling with forward and lateral speed categories; (3) Training a CNN with the generated dataset, applying transfer-learning methods; (4) Evaluation of the dataset. The approach uses Convolutional Neural Networks (CNN) to classify forward and lateral speeds in intensity levels. This method can be used for orientation in terrestrial robotics and accessibility, guiding visually impaired people. In the future, trained models can be deployed on embedded computers to be adopted in robots or as wearable devices.

The orientation of visually impaired people during walks on aisles was used as a problem to study the proposed method. This problem was chosen because it includes all stages of the process, from image signal capture to classifier evaluation.

II. Related Works

A. Assistive Technology

In [4] a wearable device is proposed to guide the visually impaired to walk and run using computer vision. The device is equipped with a monocular camera. Digital Image Processing (DIP) techniques are used for image stabilization and detection of regions of interest. Line detection is used to guide the visually impaired to stay within the limits of the walking region. This approach has been evaluated on athletics tracks and establishes a restriction of having a line along the path.

Other approaches to navigation were explored in the literature. [5] used Fuzzy control logic to suggest obstacle avoidance. The authors used indoor (office) and outdoor (grassy parks) images and focused only on detecting collisions. Other research using eyeglasses cameras is described in [6], which used a variety of urban images, where objects are marked to aid navigation.

B. CNN applied to navigation problem

Many publications studied CNN with transfer learning. Part of these works explored the use of pre-trained models in datasets with few samples to improve their generalization, as in [7], which used a VGG16 for terrain classification for autonomous robots. In [8] transfer learning was also used as a feature extractor for UAV (Unmanned Aerial Vehicle) navigation. The approach used achieved 99.97% accuracy using kNN, MLP, OPF and SVM. The study does not consider detours and collisions, and the dataset is quite different from what is being proposed.

To assist the visually impaired navigation in outdoor environments, [9] used some CNN architectures to conduct training and evaluate models to classify different floors. The dataset used is composed of images of floors with different characteristics. An interesting perspective, although the experiments were limited to only two classes, clear and non-clear path. The authors did not comment on how this approach could face the obstacle detour problem, as well.

III. Background

A. Navigation problem for impaired people

According to [10] Visual Substitution can be subdivided into Electronic Travel Aids (ETAs), Electronic Orientation Aids (EOAs), and Position Locator Devices (PLDs). Each of these subdivisions is responsible for solving specific navigation problems for the visually impaired.

From definitions and characteristics of Visual Substitution shown in [11], the approach and dataset presented in this paper combine properties from ETA and EOA. It detects obstacles close to the user's body, makes the user aware of the distance between him and the obstacles, and guides the user by providing path signs and instructions.

B. Transfer Learning

For some purposes, it is unfeasible to collect a database to train a CNN rightly [12]. The concept of Transfer Learning emerged to solve this problem [13] and has been widely used in several applications [12]. According to the concept of Transfer Learning, the knowledge acquired during the resolution of a problem can be used in a similar one [13].

This work used this technique to train CNNs with the proposed dataset in order to improve the adjustment of weights during the training phase.

IV. Methodology

A. The dataset

The Dataset for navigation with obstacle avoidance is one of the contributions of this work (available upon request). All images have a resolution of 2160x3840 pixels, captured from a smartphone camera connected to a 4-point belt attached to the pedestrian's chest. The images correspond to the vision of those who follow a path and were set in categories according to the decisions made during the walk. Decisions are based on obstacle avoidance and stopping intentions. During the walks, the pedestrian passes by static objects and people in transit.

Particularly for this article, the dataset was taken indoors. The illumination is provided by artificial lights of the buildings. After the acquisition, all images are resized to the input dimensions for each CNN used.

B. The Problem modeling

The navigation problem for the visually impaired will be explored in order to predict detours. The system must be able to inform the visually impaired the need for a left or right sidestep, as well as its intensity. Furthermore, the system must be able to infer imminent collision (suggesting deceleration) or an obstacle-free path (suggesting acceleration).

The images of aisles were labeled according to the intensities of the forward and lateral speeds. Figure 1 shows the representation of speeds during walking. Estimations of these velocities were used to label the images, according to values obtained from accelerometers and gyroscopes. The x-axis represents forward velocities, and the y-axis represents laterals. Although there are other movements related to walking, in this work only the speeds of the x and y axes were considered.

The dataset was subdivided into subsets of forward and lateral speeds. The set of forward speeds is composed of images that represent the intentions to stop and accelerate during the walk. The set of lateral speeds is composed of images that represent the sidesteps from obstacles. For each subset, the images were categorized into five classes with different speeds. A larger number of classes could be used if it is needed a smoother transition between classes. Figure 2 shows some examples of images from dataset.

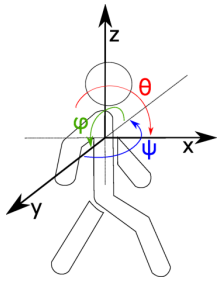
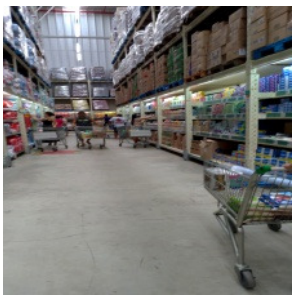


Fig. 1. Representation of existing movements throughout walking. The x, y, and z axes are linear velocities, and θ, ϕ, and ψ are angular velocities.

(a) Forward speed - class 4.



(b) Lateral speed - class 2.

Fig. 2. Example images from aisles dataset.

1) Forward speeds dataset: The subset of forward velocities was categorized so that it was possible to infer variations between stopping and walking acceleration. Table I shows the categorization of this subset and the number of samples per class that is available in each set. Velocities range from stop (Class 0) to fast (Class 4). In the aisles dataset, 572 samples per class were used. Figure 2(a) shows an example image that was labeled class 4 for forward speeds.

TABLE I
Forward speeds classes and the number of samples for each class.

Class	Class description	Samples per class
		Aisles
0	Stop	572
1	Slow speed	
2	Medium speed	
3	Moderately fast speed	
4	Fast speed	

2) Lateral speeds dataset: In the lateral speeds subset, sidesteps to the left and the right were considered during walking. Table II shows the categories of this set and the number of samples per class after the mirroring operation. The middle class was defined as no sidestep (class 2). Classes 1 and 0 represent the velocity of sidestep to the left, with class 0 being the highest intensity. Similarly, classes 3 and 4 represent the right shift.

TABLE II
LATERAL SPEEDS CLASSES AND THE NUMBER OF SAMPLES FOR EACH CLASS.

Class	Class description	Samples per class
		Aisles
0	A strong sidestep to the left	232
1	A slight sidestep to the left	
2	No sidestep	
3	A slight sidestep to the right	
4	A strong sidestep to the right	

3) Extension of lateral speeds dataset: There is a symmetry relationship in the classes that represent sidesteps. Images that represent left shifts can be mirrored to obtain images that belong to right shift classes and vice versa. The number of images indicated in Table I already includes this extension.

One question that can be asked is why not make a dataset for a regression network? Considering the problem of guiding a person on a path, the granularity of values given by a regression system is not necessary. In fact, the original dataset contains velocity values on a continuous scale. So, a test was performed with a regression network whose output was discretized into five classes. Its performance was compared with a network trained as a 5-classes classifier, and the latter had better performance.

C. Training and parameters

The CNN used for training were VGG16, ResNet50, and Dronet. VGG16 and ResNet50 are networks explored in the literature for multiclass problems. Dronet is a network applied to street navigation problems [14] that have similarities with the one studied here.

All the CNN were trained using Adam optimization with learning rate of 10^{-4}, exponential decay rate, and with a batch size of 64. The datasets were divided into 70% for training, 15% for validation, and 15% for testing. All samples were normalized using min-max normalization. AWS GPU instances were used for all the trainings.

D. Transfer Learning

All the experiments presented were performed using transfer learning. The weights of the ImageNet training [15] were used for the VGG16 and ResNet50 networks. The weights of the convolutional layers were frozen for 20 epochs and then thawed for 80. For Dronet, the weights provided by the authors were used [14]. Sixty epochs with convolutional weights frozen and 240 thawed.

TABLE III
EVALUATION OF FORWARD SPEEDS FOR AISLES.

Aisles Dataset		
CNN	**Accuracy (%)**	**F1 (%)**
VGG16	90.60 ± 01.17	90.60 ± 01.17
ResNet50	91.14 ± 01.40	91.12 ± 01.41
Dronet	86.80 ± 01.56	86.78 ± 01.58

E. Evaluation Metrics

Two metrics were used to assess the classifier performance in the presented problem: Accuracy and F1; being all of them obtained from the confusion matrix. F1 consists in a harmonic mean between Precision and Recall. For each experiment (a combination of dataset and CNN), ten runs were performed with random rearrangement of the dataset. The results presented are the average of these runs.

V. RESULTS

A. Forward speed

Table III shows the results of forward velocities inference for aisles dataset. The VGG16 and ResNet50 architectures achieved best results, with accuracies greater than 90%, being ResNet50 with the highest accuracy result, 91.14%. The Dronet architecture achieved the worst results, but still acceptables. Since Dronet is the shallowest of the three, it is understandable.

In order to evaluate performance by class, the models of each architecture that achieved the highest accuracy were selected and results are presented in Table IV. The results for the aisles present good uniformity in the results by class, no matter the network used. Classes 1 and 3 have the lowest hit rates. Apparently, the dataset offers some confusion in these classes, suggesting that the intensity of the speed is not very well differentiated, especially at the threshold between some classes.

TABLE IV
BEST ACCURACY RESULTS BY CLASS FOR THE FORWARD SPEEDS.

Class	Accuracy (%)		
	Aisles		
	VGG16	**ResNet50**	**Dronet**
Stop	92.94	91.76	97.67
Reduce speed a lot	86.04	89.53	83.72
Slightly reduce speed	96.51	96.51	85.88
Keep speed	90.69	94.18	82.55
Speed up	98.83	98.83	94.18

B. Lateral speed

Table V shows the results of lateral velocities inference for aisles. VGG16 and ResNet50 tied with more than 94% accuracy and little advantage of 0.34% in favor of ResNet50.

To evaluate performance by class, the models of each architecture that achieved the highest accuracy were selected,

TABLE V
EVALUATION OF LATERAL SPEEDS FOR AISLES.

Aisles Dataset		
CNN	**Accuracy (%)**	**F1 (%)**
VGG16	94.02 ± 01.43	94.03 ± 01.43
ResNet50	94.36 ± 01.88	94.38 ± 01.86
Dronet	89.36 ± 02.38	89.27 ± 02.45

TABLE VI
BEST ACCURACY RESULTS BY CLASS FOR THE LATERAL SPEEDS.

Class	**Accuracy (%)**		
	Aisles		
	VGG16	**ResNet50**	**Dronet**
A strong sidestep to the left	100	100	94.28
A slight sidestep to the left	100	100	94.11
No sidestep	96.51	91.42	82.85
A slight sidestep to the right	90.69	97.05	100
A strong sidestep to the right	98.83	97.14	88.57

and results are shown in Table VI. The class that represents no sidesteps has the lowest hit rates in most cases. Images that indicate some sidestep always contain an easy-to-detect pattern, with the presence of an obstacle. On the other hand, images from the 'no sidestep' class do not present any objective pattern, making their classification difficult.

Comparing Tables III and V, one thing can be highlighted: It is easier to classify sidesteps (lateral movements) than the forward speeds (forward movements). Usually, more images improve a classifier's hit rate, and Tables I and II show datasets of different sizes. However, one should note that even with a smaller dataset, the sidestep problem was solved more successfully.

VI. CONCLUSIONS

This work presents a study on using computer vision to guide the walking of visually impaired people. It can also be extended to other application domains, like terrestrial robotics. A dataset was created which allows to recommend the walking speed and obstacle avoidance. The dataset with indoor (aisles) images were evaluated. Sidesteps are modeled by the lateral speeds, while the forward speeds indicate the stops and accelerations. Tests on VGG16, ResNet50, and Dronet architectures were performed using transfer learning.

The ResNet50 stood out for getting the best results across all datasets. For forward speeds, ResNet50 presented an average accuracy superior to 91% for indoors images. For lateral speeds, the ResNet50 achieved average accuracy bigger than 94%. Additionally, it is easier to categorize lateral movements than the forwards. On the other hand, VGG16 achieved a very similar result, offering a lower computational cost. This is particularly important, considering the possibility of the algorithm being deployed in a wearable system.

The provided dataset is relevant in three contexts: (1) detecting obstacles close to the user, (2) making the user aware of obstacles in the path, and (3) guiding the user by providing path signs and instructions.

As future works, it is suggested:

- Enlarge the dataset by including external images to improve its generality.
- Evaluate the performance of less-complex architectures models (VGG16 and DroNet) in embedded platforms aiming at their use as a wearable device.

REFERENCES

[1] World Health Organization, "Global cooperation on assistive technology (gate)," 2021, https://www.who.int/disabilities/technology/gate/en/, Last accessed on 2021-03-08.

[2] A. Aladren, G. Lopez-Nicolas, L. Puig, and J. J. Guerrero, "Navigation assistance for the visually impaired using RGB-d sensor with range expansion," *IEEE Systems Journal*, vol. 10, no. 3, pp. 922–932, Sep. 2016. [Online]. Available: https://doi.org/10.1109/jsyst.2014.2320639

[3] R. Cheng, K. Wang, L. Lin, and K. Yang, "Visual localization of key positions for visually impaired people," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, Aug. 2018. [Online]. Available: https://doi.org/10.1109/icpr.2018.8545141

[4] A. Mancini, E. Frontoni, and P. Zingaretti, "Mechatronic system to help visually impaired users during walking and running," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 649–660, Feb. 2018. [Online]. Available: https://doi.org/10.1109/tits.2017.2780621

[5] W. M. Elmannai and K. M. Elleithy, "A novel obstacle avoidance system for guiding the visually impaired through the use of fuzzy control logic," in *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, Jan. 2018. [Online]. Available: https://doi.org/10.1109/ccnc.2018.8319310

[6] B. Jiang, J. Yang, Z. Lv, and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1375–1383, Apr. 2019. [Online]. Available: https://doi.org/10.1109/jiot.2018.2842229

[7] F. Schilling, X. Chen, J. Folkesson, and P. Jensfelt, "Geometric and visual terrain classification for autonomous mobile navigation," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Sep. 2017. [Online]. Available: https://doi.org/10.1109/iros.2017.8206092

[8] S. P. P. da Silva, P. H. Filho, L. B. Marinho, J. S. Almeida, N. M. M. Nascimento, A. W. de O. Rodrigues, and P. P. R. Filho, "A new approach to navigation of unmanned aerial vehicle using deep transfer learning," in *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, Oct. 2019. [Online]. Available: https://doi.org/10.1109/bracis.2019.00047

[9] F. Breve and C. N. Fischer, "Visually impaired aid using convolutional neural networks, transfer learning, and particle competition and cooperation," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Jul. 2020. [Online]. Available: https://doi.org/10.1109/ijcnn48605.2020.9207606

[10] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, p. 565, Mar. 2017. [Online]. Available: https://doi.org/10.3390/s17030565

[11] K. Manjari, M. Verma, and G. Singal, "A survey on assistive technology for visually impaired," *Internet of Things*, vol. 11, p. 100188, Sep. 2020. [Online]. Available: https://doi.org/10.1016/j.iot.2020.100188

[12] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient inference," 2017.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010. [Online]. Available: https://doi.org/10.1109/tkde.2009.191

[14] A. Loquercio, A. I. Maqueda, C. R. del Blanco, and D. Scaramuzza, "DroNet: Learning to fly by driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1088–1095, Apr. 2018. [Online]. Available: https://doi.org/10.1109/lra.2018.2795643

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2009. [Online]. Available: https://doi.org/10.1109/cvpr.2009.5206848