Blind navigation using ambient crowd analysis

*Avaneep Gupta
Department of Mathematics
IIT Delhi
New Delhi, India
mt1150592@iitd.ac.in

*Sumeet Khandelwal
Department of Mathematics
IIT Delhi
New Delhi, India
mt1150615@iitd.ac.in

Tapan Gandhi
Department of Electrical Engineering
IIT Delhi
New Delhi, India
tgandhi@ee.iitd.ac.in

***Abstract*—Research in assistive technology has been on the rise over the last decade. Numerous solutions and consumer products have flooded the market to guide visually impaired making use of beacon technology, depth cameras and many more. Though certain products and solutions are available for highly structured and regular indoor environments, we are still a long way from an industrial level product for unstructured, dynamic and irregular outdoor environments. Our work harnesses the decision making power of sighted individuals and crowd as a group surrounding the visually impaired. This information extraction from the crowd along with coarse terrain mapping of major obstacles like footpath edges, walls and large pot holes will help the subject to navigate dynamic and irregular environments. This out of the box approach provides us the margin to use low grade equipment and develop algorithms with low computational complexity. The paper explains the theoretical aspects of this approach along with its proof of concept and some remarkable results achieved in real life implementation.**

***Index Terms*—Assistive technology, navigation, Computer vision, Path planning, Gait recognition, Object detection**

I. Introduction

An increased interest in assistive technology over the last decade has helped ease navigation for the visually impaired [1] [2]. Efficiency of indoor navigation using beacon technology [3] and hand held cameras is reaching within the same domain as that of a sighted human being. On the other hand outdoor navigation with assistive tech still requires an additional sighted human for smooth traversal. Indoor landscapes can be mapped with high accuracy and along with learning algorithms via user end feedback can perfect navigation for specific environments. Navigation in outdoor environments offer bigger challenges due to the following two reasons. Primarily, in countries like India the terrain is highly unstructured and irregular. Mapping minute details of irregular footpaths require high end devices and computationally expensive algorithms both of which are major challenges to overcome for affordable hand held devices. Secondly, majority of existing assistive technology in this domain view crowd as an obstacle and tries to navigate around them, and fails for countries as crowded as India. With every human being tagged as an obstacle, terrain mapping gets affected, generating very small data set to find the path of least resistance.

There are numerous innovative products currently available like sonar based SmartCane [4] ,UltaCane [5], Sunnu Smart Watch by MIT [6], MiniGuide [7] and visual cue based OrCam [8] ,OxSight [9]. Both these categories of products fail to deliver results in an Indian setting due to the latter reason. This paper presents a novel approach to the problem of helping blind navigate in crowded environments, assuming traversal is required in a particular known direction. It is inspired from [10], [11], [12], [13], [14], [15], which uses 3D cameras like Bumblebee and Microsoft Kinect. Their philosophies and approach as mentioned fail in crowded settings and are designed for structured environments.

Sighted human beings subconsciously or consciously follow the path of least resistance. Their ability to detect small aberrations on footpaths, deflect oncoming human traffic and walk at a safe distance from major obstacles like walls, edge of footpaths and potholes is unparalleled and cannot be mimicked by a computer.

Taking motivation from the fact that a visually impaired person is aided by a sighted human being to traverse in the absence of an assistive device, our model estimates the ability of a stranger in the current field of view to be a good *leader* for the visually impaired to follow, an approach somewhat similar to [23] and [22]. This *leader* is updated dynamically to not let visually impaired deviate from his route. In case no *leader* is found our model along with coarse object detection highlights major obstacles, harnesses the decision making capabilities of every individual through detection and tracking algorithms and assign *walkability scores* to each point on the ground in the current field of view. This so called *walkability score* metric enables us to compute the path of least resistance quantitatively for the subject to follow. This approach provides a margin to use low grade equipment while at the same time extract crucial information regarding the dynamic (e.g. oncoming human traffic) and static conditions (e.g potholes) of the environment. The proposed model very intelligently circumnavigates numerous problems of terrain mapping, obstacle detection and segregation of surrounding humans into categories of assets and non-assets.

II. Overview

In order to guide the subject we present a pipeline with two parts that are based on two possible scenarios in a crowded environment :

- A sighted human being can be identified walking in a similar direction whose trajectory can be safely followed

* - *equal contribution*

by the blind person. We call this sighted person *Leader*. This *leader* will be updated in real time for the subject to follow.
- No sighted human being can be identified as a *leader*, but there are people around the subject whose motion can provide with crucial information about the surroundings.

In order to extract useful information from crowd and identify a *leader* every person in the visible field of view is characterized by the following parameters extracted using an RGB-D feed -

- Velocity vector with respect to the subject.
  (Subject will be able to follow a *leader* if he/she has an optimum speed)
- Displacement vector with respect to the subject.
  (Large distances from a *leader* can be harmful in dynamic scenarios where obstacles might come up in between the subject and *leader*)
- Angle between the subject and the pedestrian with reference to the direction of motion of subject.
  (Important to ensure the position of the *leader* falls almost in straight line to that of the motion of the subject)
- Angle between velocity vector of the subject and the pedestrian.
  (Important to keep note that the *leader* is moving in the desired direction)

It is always favourable to be able to identify a *leader* in the crowd and follow him/her. In the case when no *leader* is identified, it becomes important to self compute the path of least resistance based on the information gathered from crowd movements and environment. To be able to quantify the ease of traversal through a point on ground we introduce the concept of *Walkability Score* defined as - A numerical value between -10 and 100 assigned to a square grid on the ground at a particular time instant which indicates the ease of walking on that grid point while moving in the subject's desired direction (100 being the most walkable grid, -10 being not walkable). It is a dynamic score and makes use of the following factors

- Presence of an obstacle at that grid point.
- Distance of grid point from the closest obstacle.
- Movement of people through that grid and their direction of motion.
- Minimum of the time elapsed since a person last moved through that grid point and the expected time remaining for another person to move through that grid point.

It becomes important to note that after a person walks through a point in the same direction a high *walkability score* is assigned, but the assurance of high walkability will decay as time passes and thus its score must be gradually reduced over time to the value that would have been assigned in case of no pedestrian. Similarly, for a point through which a person walks in opposite direction a low walkability score is assigned but its score is gradually increased.

While crowd sensing algorithms and computer vision paradigms form the base of our model, first scenario is implemented using machine learning techniques to identify a *leader* while the second scenario uses real time environmental data, crowd movements and self constructed metrics refined via numerous simulations to identify a deterministic path of least resistance frame by frame.

The model maps the complexities of decisions taken by a pedestrian onto its few gait parameters and along with coarse obstacle detection identifies a *leader* and assigns *walkability scores*. It thus becomes important to ensure whether or not this approach is good enough to deliver practical and real life implementable results. Therefore, as a proof of concept, the pipeline is first simulated on test data. Fig. 1 represents the work flow overview of our framework.

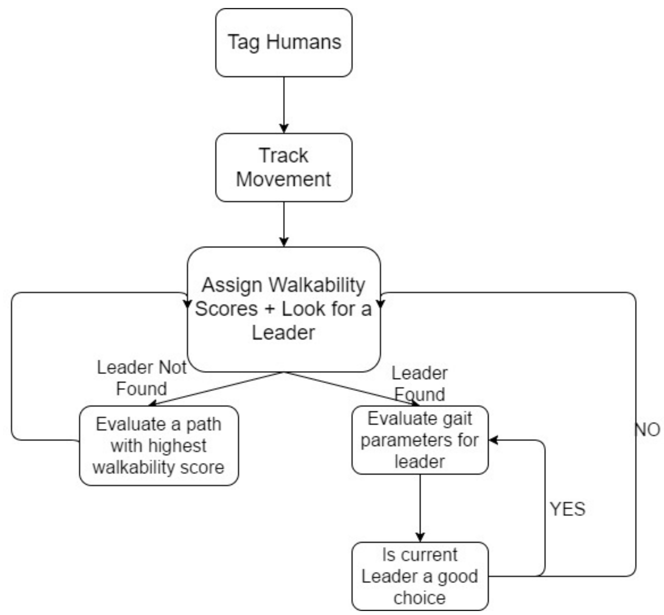

Fig. 1. Overview of the pipeline

III. Proof of Concept

For simulation we use the data set present in [1]. A snapshot of the recorded data is shown in Fig. 3. This data set has been recorded from a hotel room and provides a bird's eye view of the pedestrians in the street below. It's a stationary view but if the pipeline works in stationary scenarios, it would work when the person is in motion too as every gait parameter is taken relative to blind person. The data set provides the pedestrians position and velocity for every 5 frames. The occupancy map indicating the obstacles on the ground was also provided. It also provides group behaviour which is very useful and can be considered in future works.

A. Classifier for leader identification

For given data, a classifier was trained. For classification, the pedestrians were labelled suitable *leaders* by checking their relative velocity and distance w.r.t to a point (blind person position) in the frame. Numerous classifier such as decision trees, K-Nearest Neighbour (KNN) were trained using five fold cross validation with varying parameters. The

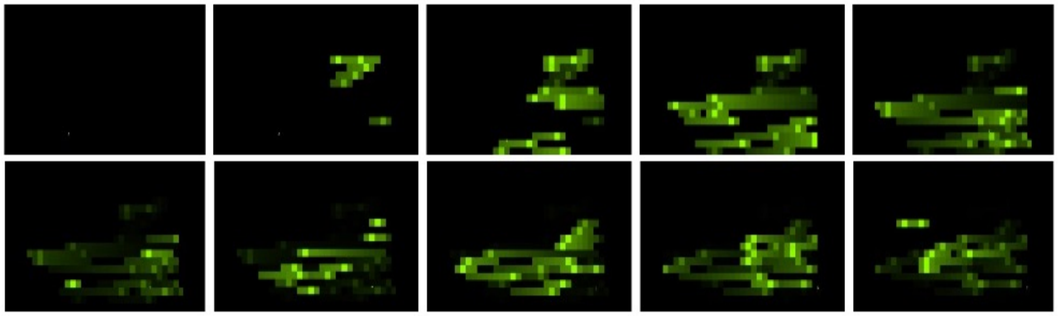

Fig. 2. Snapshot of every 3 seconds indicating Walkability Score proportional to the brightness of color for the simulation on test data

Fig. 3. Snapshot of image used for simulation

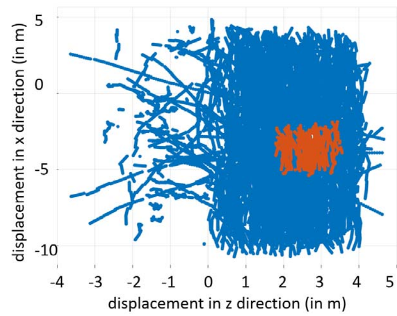

Fig. 4. Scatter plot of displacement of pedestrian in X and Z direction w.r.t visually impaired

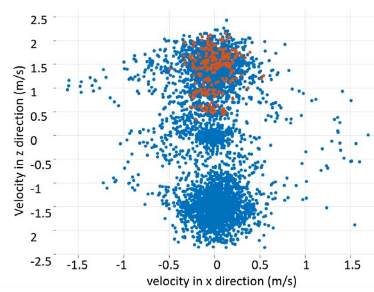

Fig. 5. Scatter plot of velocity of pedestrian in X and Z direction w.r.t visually impaired

split on testing of training data was 1:10. The best results were obtained from classifier based on the Gaussian SVM model,with training accuracy 98.6% and testing accuracy of 99.35%. Visual representation in image Fig. 4 and Fig. 5 of our test data shows very clear clustering of *leader* (Red) and non-*leader* (Blue). True positive rates at 99% for non *leader* and 72% for *leader*. Positive predictive value at 99% from non-*leader* and 80% *leader*.

B. *Results of simulation*

The results of the simulation justifies the pipeline's use for navigation. At many crucial moments, the decisions taken are very intuitive and can be used to guide a blind person. The classifier was able to pick an appropriate *leader* and also leave or switch to another *leader* whenever situation became unfavourable. The path taken then proved to be without obstacles and one which most people took. Some of the special features of such a path was that they were generally at a safe distance from the obstacles and there was a tendency of people moving in opposite directions to move in separate lines, which changed dynamically over time. It was this line of flow on which the person is guided which according to our intuition is best for him. Fig. 4 and Fig. 5 represent the gait parameter values of *leader* (red) and a non-*leader* (blue) as definite clusters, further validating the existence of certain characteristics of a good *leader*.

IV. Real time implementation of pipeline

For practical implementation of framework, 3 dimensional data is supplied by Intel ZR300 RGB-D camera [16]. One of the limitations of this class of camera is that the results are not very accurate in certain conditions especially in extreme sunlight making it unsuitable for use. We discovered that the camera performs best in grassy or muddy surface as compared to concrete and tarmac road surfaces. Moreover even in favourable conditions due to excessive noise only 59% of pixels recorded had similar values for the same scene under identical conditions, making pre-processing an important step.

A. *Pre-processing used*

For pre-processing, one of the biggest issue was to deal with noisy pixels with null values as portrayed in Fig. 6. A fast and

Fig. 6. *a* RGB Test image. *b* Corresponding (extremely noisy) Depth Image

efficient algorithm was needed for an effective implementation of our model. Hence the following algorithm was implemented

- Fill the null point with modal (non-zero) depth value of the grid

From the pre processed data, the ground plane is extracted using a standard RANSAC algorithm. This algorithm functions by randomly sampling and fitting a plane to the points and verifying if it's a good fit for the data give. It is very robust to noise and outliers. Next in line, after the ground plane is detected, major obstacles are detected which are within the scope of camera. Every pixel in the depth map whose distance from the ground is above a certain threshold is classified as a part of an obstacle. This leaves out the finer irregularities and obstacles as the data obtained is very coarse, but serves the purpose of identifying major obstacles.

B. Detection

Pedestrian detection is an active area of research. A review paper [17] gives a fair idea of the state of art development. Different algorithms for crowd detection, tracking and extracting gait parameters were tested. Detection using upper body depth map [18], [19] and HOG based SVM for pedestrian detection [20] gave poor results with accuracy close to 5% and 10% due to the noise in data. Finally a color based detector was used, which showed promising results in detection as seen in Fig. 7.
Front towards the viewer - detection rate 66%.
Back towards the viewr - detection rate 88%.

This detector is based on representation of any color in HSV system [21]. The system implements color segmentation and identification based on 3 metrics namely Hue, Saturation and Value. It's use is advantageous as color of clothing provides a definitive way of identifying pixels corresponding to a particular pedestrian especially when depth maps are accurate near fringes of clothes. One more advantage is that in a scenario with multiple people wearing different colored clothing, every person can be detected separately and tracking becomes much easier given that multiple tracking is a fairly challenging problem. A limitation with this detection technique is an increased false positive rates and incomplete elimination of background image due to similar color of clothing. This algorithm delivered the expected high accuracy with very low false positives and false negative rates, with results accurate to every pixel, in certain cases with an accuracy of 100% and no false positives and false negatives.

Thus this detector was used to build a prototype for our navigation system in a semi simulated environment where the clothes color of pedestrian were already known.

Fig. 7. *a* ROI in a color based detector. *b* Accurate detection using color based detector

C. Algorithm

For scenarios when no *leader* is detected it becomes important to have an efficient path finding algorithm which makes use of the *walkability scores* and provides a path of least deviation as well as least resistance for the visually impaired to reach the farthest point in the current field of view. For this a customized algorithm was designed to keep note of the following points.

- The run time complexity of it should not be exponential or high order polynomials in M and N (for a grid of dimensions M and N)
- The algorithm should be flexible enough to accommodate the needs of different people. Not every visually impaired person is completely blind, they have varying sense of sight some sensitive to source of light while some with vague sense of sight that can make out objects very faintly. Thus the threshold of walkability score above which it is safe for the subject to follow a path will be different for everyone. For example threshold for a sighted human being would be 0 while for a completely blind person would be above 80.
- Given a particular threshold the algorithm must find a path of least deviation from a direct path, causing less confusion to the people around the blind and also minimizing information transfer between the system and the blind, creating less confusion and easier for the subject to follow instructions.
- The algorithm must try to find a path with a certain walkability threshold, if a path is not found it must reduce the threshold and once again try computing a path until either a path is obtained or threshold becomes 0 (highly unfavourable and a rare condition).

D. Results

A number of classifiers were trained for real time implementation. One of the best results were obtained on using

Decision Tree classifiers, with max split = 100 using Gini Diversity Index as the splitting criteria. Our data set comprised of a total of 16 videos (600 frames each) and processed upon by the algorithms and computer vision techniques discussed above. The results are summarized in the following points. Decision Tree Results:

- Training accuracy : 96.3%
- Testing accuracy : 92.19% (90:10 split)(Training data : 153 positive and rest negative)

Gaussian SVM : Testing accuracy : 90.95%, recall = 0 K-Nearest Neighbour : Testing accuracy : 90.95%, recall = 0 Validation of the results was made by cross referencing the results of *leader* detection algorithm and the tags of *leader* or not assigned by volunteers who were made to watch 16 videos from the same perspective as that of the visually impaired person.

V. Discussion

The novel approach to make use of fellow pedestrians without bothering them in navigating through dynamic and irregular environments works in principle as presented in the simulation and its proof of concept. The proposed framework is designed to be implemented from the frame of reference of visually impaired. Hence real life implementation testing was done using a stationary camera, which shows promising results. The model highlights various limitations for a practical real time implementation but offers deep insights and a starting point for future work and development. Its hardware implementation would require developing efficient techniques to transmit directional information to the visually impaired, either in the form of sensory vibrations or audio signals. Next in line work will include methods to eliminate noise during image capture and devising ways to make the camera more robust in varied surroundings.

VI. Conclusion

The novelty of our concept and the practical validity can be easily seen through our results achieved on semi-simulated models. Through the results we have validated the existence of certain characteristics of a good *leader* that can be quantified through limited gait parameters as seen in Fig. 4 and Fig. 5. Real life testing provided intuitive results for leader identification. An accuracy of 90% in pedestrian detection and tracking was achieved despite using cheaper noise elimination and ground plane detection algorithms. Our model can be broken down into three major components - ground plane detection, obstacle detection and pedestrian detection and tracking. Our work has greatly reduced the need of high quality data in finding optimum path by introducing the concept of *leader* and *walkability score*.

References

[1] Marion A. Hersh and Michael A. Johnson. *Assistive Technology for Visually Impaired and Blind People* . Springer Publications, 2008. ISBN : 9781846288678

[2] Kun (Linda) Li. Electronic Travel Aids for Blind Guidance. PhD thesis. Berkley Engeneering, 2015

[3] S. A. Cheraghi, V. Namboodiri, and L. Walker. Guide- Beacon: Beacon-based indoor wayfinding for the blind, visually impaired, and disoriented. In: *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)* . Mar. 2017, pp. 121130. DOI : 10.1109/PERCOM.2017.7917858.

[4] Vaibhav Singh, Rohan Paul, Dheeraj Mehra, et al. SmartCane for the Visually Impaired: Design and Controlled Field Testing of an Affordable Obstacle Detection System. In: *TRANSED 2010: 12th International Conference on Mobility and Transport for Elderly and Disabled PersonsHong Kong Society for Re- habilitationS K Yee Medical FoundationTransportation Research Board* . 2010

[5] Hoyle B. and Waters D. Mobility AT: The Batcane (UltraCane). In: Springer, London, 2008. Chap. 6.

[6] Albertorio. *sunu smart watch assisitve technology*. URL: https://www.sunu.io/index.html?.

[7] GDP Research. *The Miniguide mobility aid. 2005; Adelaide, Australia: The Miniguide*. URL: http://www.gdpresearch. com.au/minig 1.htm.

[8] Yonatan Wexler. *orcam assitive device.* URL: https://www.orcam.com/en/.

[9] *smart specs for blind*. URL: http://smartspecs.co/.

[10] V. Pradeep, G. Medioni, and J. Weiland. Robot vision for the visually impaired. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. Apr. 2010, pp. 1522. DOI: 10.1109/CVPRW.2010.5543579.

[11] A. Aladrn, G. Lpez-Nicols, L. Puig, et al. Navigation Assistance for the Visually Impaired Using RGB-D Sensor With Range Expansion. In: *IEEE Systems Journal* 10.3 (Sept. 2016), pp. 922932. ISSN: 1932-8184. DOI: 10.1109/JSYST. 2014.2320639.

[12] Young Hoon Lee and Gerard Medioni. RGB-D camera based navigation for the visually impaired. In: *Proceedings of the RSS.* 2011.

[13] Gu-Young Jeong and Kee-Ho Yu. Multi-section sensing and vibrotactile perception for walking guide of visually impaired person. In: *Sensors* 16.7 (2016), p. 1070

[14] Paulo Costa, Hugo Fernandes, Paulo Martins, et al. Obstacle detection using stereo imaging to assist the navigation of visually impaired people. In: *Procedia Computer Science* 14 (2012), pp. 8393.

[15] Huy-Hieu Pham, Thi-Lan Le, and Nicolas Vuillerme. Real-time obstacle detection system in indoor environment for the visually impaired using microsoft kinect sensor. In:*Journal of Sensors*2016 (2016).

[16] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, et al. IntelR RealSenseTM Stereoscopic Depth Cameras.

[17] Detection: An Evaluation of the State of the Art. In:*IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.4 (Apr. 2012), pp. 743761. ISSN: 0162- 8828. DOI: 10.1109/TPAMI.2011.155.

[18] O. H. Jafari, D. Mitzel, and B. Leibe. Real-time RGBD based people detection and tracking for mobile robots and head-worn cameras. In:*2014 IEEE International Conference on Robotics and Automation (ICRA)* . May 2014, pp. 56365643. DOI: 10 . 1109 / ICRA . 2014 . 6907688

[19] O. H. Jafari and Michael Ying Yang. Real-time RGBD based template matching pedestrian detection. In: *2016 IEEE International Conference on Robotics and Automation (ICRA).* May 2016, pp. 55205527. DOI: 10.1109/ICRA.2016.7487767.

[20] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Vol. 1. IEEE. 2005, pp. 886893.

[21] soham gandhi. *Color based object detection*. URL: https: //www.codeproject.com/Articles/850023/Color-Based- Object-Detection.

[22] Procopio Stein, Anne Spalanzani, Vitor Santos, et al. Experiments in Leader Classification and Following with an Autonomous Wheelchair. In: *Experimental Robotics*. Springer. 2016, pp. 245260.

[23] M. Gupta, S. Kumar, L. Behera, et al. A Novel Vision- Based Tracking Algorithm for a Human-Following Mobile Robot. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47.7 (July 2017), pp. 1415 1427. ISSN: 2168-2216. DOI: 10 . 1109 /TSMC. 2016 . 2616343.