Outdoor and Indoor Navigation Assistant for Visually Impaired People

Mansi Phute
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
mansi.phute18@vit.edu

Khushi Bharti
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
khushi.bharti18@vit.edu

Swarali Purandare
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
swarali.purandare18@vit.edu

Ashutosh Lahoti
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
ashutosh.lahoti18@vit.edu

Yash Nikhare
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
yash.nikhare18@vit.edu

Prof Jyoti Madake
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
jyoti.madake@vit.edu

Abstract— This paper presents a unique assistive navigation system based on simultaneous localization and mapping (SLAM) , to assist visually challenged people in both indoor and outdoor environments. The system uses a stereo camera and ORB to detect features in images and gives instructions to guide the user to the intended location by matching the detected landmarks against the corresponding features on the digitised floor map. Because of incorrect data association caused by moving objects, the accuracy of localization and mapping is reduced. Mapping utilising visual simultaneous localization and mapping (V-SLAM) is reduced in dynamic environment conditions. To address this issue, a stereo SLAM approach based on dynamic object rejection is developed, in which only the static portion of the image is extracted and used in mapping.This improves robustness as accuracy of the model in dynamic conditions

Keywords—

I. Introduction

Navigation in outdoors environments is a great challenge in the lives of visually impired people. Google Maps and other pedestrian navigation systems have grown in popularityAs smartphones and portable tablet devices have evolved and become more widespread. However, because almost all of these systems are designed for healthy people, visually impaired people have a hard time using these systems.

There has recently been a surge in interest in wearable blindness-assistive devices. Wearable blindness-assistive technologies can distinguish between people, messages, traffic signals, and currency. Wearable navigation devices (WNDs) for blind people still rely on traditional methods, such as use of ultrasonic sensors, GPS, inertial odometry, IMU and other indoor localization sensors, which are unable of meeting the precision and accuracy requirements of walking navigation and have major limitations in the indoor environment

A moving part in SLAM has at least one sensor which gathers information about its environment . It could be a camera,a sonar or a laser scanner . Other than that moving part can have other sensors to measure its own motion (encoders, accelerometers, gyrometers, odometer, IMU). Every SLAM system consists of at least one moving part connected to a computer.

In this paper, therefore, we have detailed the process of the development of a navigation assistant system for visually impaired people using the visual SLAM algorithm using a head-mounted stereo camera.

The paper majorly focuses on :

- Real time performance using ORB feature extraction.
- A stereo SLAM algorithm including dynamic object rejection to improve the robustness and localization accuracy .
- Reverse trajectory mapping

II. Literature Review

The localization and mapping are dependent on each other as to navigate, the system needs to locate itself along with map generation . The algorithm used to do so is Simultaneous Localization and Mapping (SLAM). SLAM algorithm is considered as the backbone of autonomous manoeuvring. SLAM uses 3 major sensors: laser, monocular and binocular. It is a method of obtaining motion of the sensor and reconstruction of the structure in an unspecified environment. The Simultaneous Localization and Mapping (SLAM) which uses camera is called visual SLAM (vSLAM) because it is based only on visual data . vSLAM has various applications in fields like augmented reality computer vision, and robotics . The use of SLAM methods has resulted in semi-dense or dense 3D area reconstruction. Feature-based monocular SLAM methods, on the other hand, can give a more accurate trajectory than other direct methods. It is possible to achieve both a dense or semi dense 3D reconstruction from depth prediction and an accurate trajectory from ORB-SLAM . Laser SLAM is not good for highly dense or crowded environments or object detection. It is very unsuitable for humans, and as that is our main objective we focus on monocular and stereo SLAM. However, getting depth information from the images is very computationally expensive. This led to the use of depth sensors along with a camera or RGB-D Sensor. The objective of the SLAM algorithm is to create an accurate 3D map of the surroundings that can be used for reliable navigation. While graph SLAM techniques started to be used as batch techniques offline , recently the advancement in optimization of incremental graphs allows their application for online localization and mapping. Currently this is the most reliable method for 6D slam. Their graph symbolises poses of robots and landmarks as nodes, interconnected by their related evaluation.

There are a few steps taken in all SLAM implementations, i.e., those of feature extraction, Transformation estimation, Aligning Point Clouds, Graph Optimization, and then Map Representation.

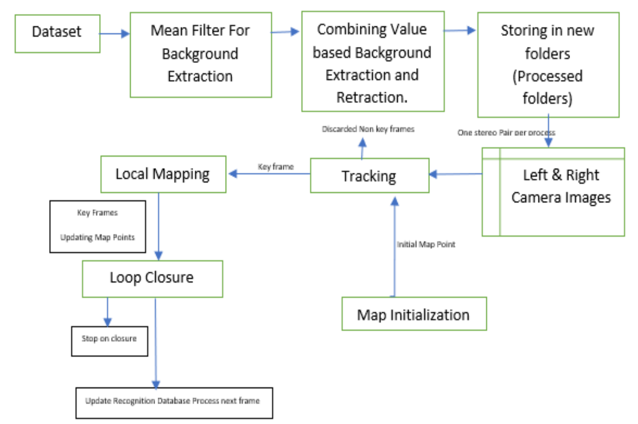
Scale Invariant Feature Transform (SIFT), its derived descriptor Speeded Up Reverse Features (SURF) and ORB are often used as feature descriptors. SIFT extracts a lot of features and the error is small however it has a slow speed. SURF has quicker speed than SIFT, it extracts fewer features, robust performance is good. ORB gives the best real-time performance, although its robust performance is bad. ORB is a combination of FAST and BRIEF descriptors. FAST detects corner detection. It utilizes Harris Corner Detection. EKF approaches model their landmark-based maps as multivariate Gaussians. The major disadvantage of EKF is higher computation with the number of landmarks, however lower complexity variation still exists. In EKF slam one can fuse data from multiple sensors. ORB-SLAM obtains ORB features from frames and uses feature matching to estimate the pose of the camera and construct a sparse map. However, the features obtained from the common frame (relative to the keyframe) are not used again except in this estimation of camera pose stage ,, which will be computationally expensive.

Feature Extraction Method	Attributes
SIFT	More features extracted, not useful in real time
SURF	Better time performance, Lesser Features, more robust
ORB	Best Real time performance, poor robust performance
EKF	high convergence, handle uncertainty, slow in high dimensional maps

For initial posture estimation, Random Sample Consensus (RANSAC) is utilised. To remove incorrect matching, the PROSAC algorithm, a version of RANSAC, is used. The PROSAC algorithm reduces outliers through iterative operations and estimates the parameter matrix using the least square approach. PROSAC outperforms RANSAC in terms of robustness and computing efficiency. To align two sets of point clouds, the Iterative Closest Point (ICP) technique is commonly employed. For precise estimate and motion transform, the ICP algorithm is used. This is referred to as Direct Image Alignment and Affine Lighting Correction. The graph has both nodes and edges, with nodes representing the user's posture and edges representing the transformation of two neighbouring moments. The graph optimization comes at the back-end of the graph. The assignment in the back-end module is to find the state vector that maximises the probability function of geometric constraints. The trajectory may then be used to project all point measurements into a global 3D coordinate system, resulting in a point cloud representation of the environment. However, the map is redundant and necessitates a huge number of calculational and memory resources. The 3D map of the surrounding environment is shown using Octomap or rviz. Sub-mapping is a strategy for improving performance. A deep learning-based SLAM system and several submaps can also be used to navigate. In a somewhat large environment, it offers mapping and localization information for the following planning and control system. Google Cartographer and the corresponding image library may be used to create several consecutive submaps, and Deep LCD and the adaptive Monte Carlo localization (AMCL) Robot Operating System (ROS) package may be used to do on-line map switching and navigation.

III. METHODOLOGY

1. FLOWCHART



2. ALGORITHM

Visual SLAM was chosen as opposed to other methods due to reasons mentioned. In Visual SLAM algorithms Stereo SLAM is preferred over RGB-D SLAM in this application as depth in RGB-D is measured using IR sensors. However, in the outdoor environment the ambient infrared radiation affects the reading of the sensor. This causes unreliable information that gives us no information about the depth. The implementation of stereo vision gives us a lot more reliable information about the environment.

Dataset: The dataset used here is UTIAS In the Dark datasets which has 159,549 stereo image pairs. These images are from the University of Toronto Institute for Aerospace Studies. In the chosen dataset GPS data is also available to compare the obtained result with the ground truth

The first step in SLAM is processing the dataset to remove dynamic objects. Then initialization of the map is done. Relative camera pose is extracted by the 3D stereo points. 2D feature correspondences are used in the triangulation process. Tracking of features is done in each frame, poses are created by matching features in current and previous frames. Key Frames are the frames captured that help us in tracking and localization process. Whenever the current frame matches the keyframe a 3-D map of points is created and bundle adjustment is done to refine the camera poses. And then detection of the loop and optimization of pose graphs is done. And reverse trajectory mapping is done by reconstructing the trajectory in the opposite direction.

3. Getting Background From Image

A major issue in the localization and mapping especially when it comes to outdoors mapping as we have little to no control over the environment. The environment is dynamic and the same scene will change over time. This will interfere with loop closure. However the background will be constant over time. Extracting features from the background for localization and mapping will give us a robust map that won't be affected by the changes of foreground objects in the image.

There are multiple methods of background extraction like gaussian estimation, Kalman filter, adaptive background estimation or foreground detection. The types used are given in detail below

A. Mean Filter and Median Filter

We use a mean or median filter as we need the background relatively intact and also want it to work in low illumination conditions. Due to daylight alteration of the light intensity, color in the frame changes. This is why we opt for the HIS (Hue Saturation, Intensity) or HSV (Hue Saturation Value) colour model as opposed to the RGB (Red Green Blue) colour model. This can be done by taking the mean or median of three different parameter of image hue, value and saturation respectively, over a range of consecutive images. This removes the dynamic object in the frames and keeps the static background. It gives quite good features for mean filter compared to the median filter.

Fig 2. The Original Input Image

Fig 3. Output of value-based background extraction

Fig 4. Output of saturation-based background extraction

B. CVS method

Secondly, a method which is a combination of hue, saturation and value is used for background extraction.In Fig.3. value-based background extraction recognizes the white and bright color but not dark color on the other hand In Fig.4. saturation-based background extraction recognizes dark color but not white and bright color. This method is used to combine both the methods to recognize both the dark and light color as shown in Fig.5.

Fig 5. Extracted Background

4. SLAM

Visual simultaneous localization and mapping (vSLAM), is the method in which visual input is used to map the environment and the surroundings. We do not use any other input except the images from the camera. We use vSLAM in areas like areas, autonomous vehicles, etc. Monocular vSLAM is possible though it is not accurate as depth is unreliable and causes the map to drift over time. Using a stereo camera provides us solutions to these issues and gives us a more reliable solution. The steps to build a map from the stereo images is as follows

A. Map Initialization

We initialize a 3D map and store the first left image as the key frame

B. Tracking

After initialization when the system receives a stereo pair the new pose is calculated by matching the ORB features of current left frame to the previous key frame. After the pose is estimated it is then refined.

.

C. Local Mapping

When the left image is a key frame, new 3D points are calculated using the disparity in the stereo left and right images. We use bundle readjustment so as to minimize the reprojection errors.

D. Loop Closure

The keyframes are then compared to the previous keyframes by the bag of features method and loops are detected. On loop closure the optimization of all the poses is carried out

4. Reverse Trajectory Mapping

To navigate the people successfully to their starting position it is necessary to reconstruct the trajectory in the opposite direction. To do this the pose graph is extracted from the traversed path. This graph is looped through in the opposite direction after a 180 degrees turn. A new pose graph is generated by using the original pose graph in the opposite direction.

IV. Results

1. Foreground Removal

2. SLAM implementation feature matching

Fig 5 : Feature mapping in processed data

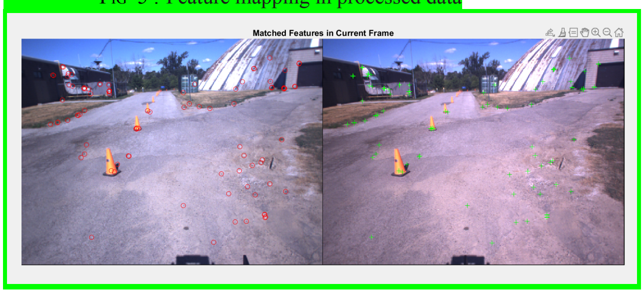
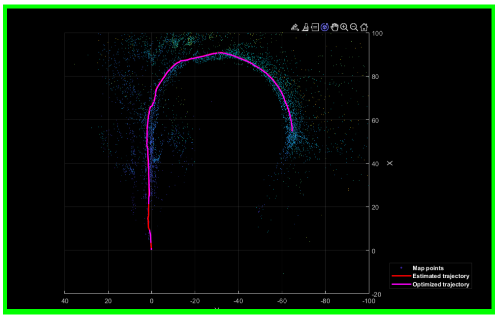
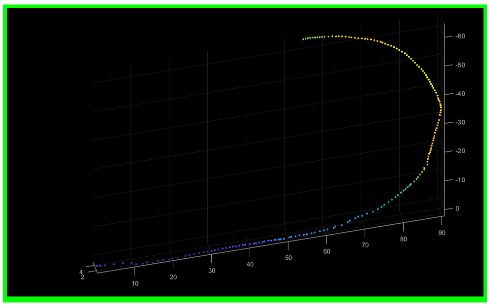


Fig 6 : Feature mapping in original data

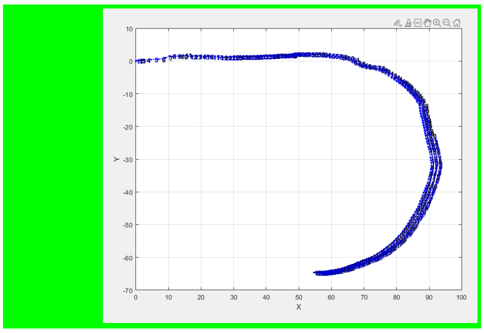
3. Optimized graph



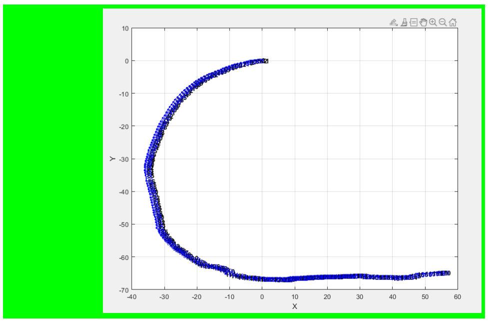
4. Extracted 3D worldpoints graph



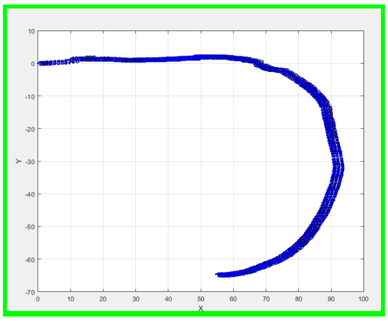
5. Extracted Posegraph

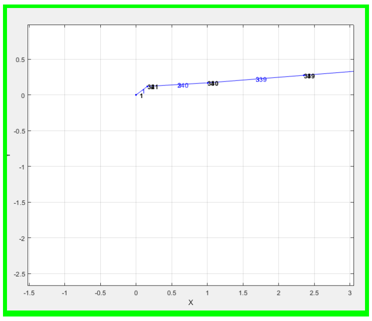


6. Reversed Posegraph



7. Superimposed extracted and reverse posegraph





V. Future Work

We are successfully able to get the forward and reverse trajectory, the later part could be navigation. i.e., to direct the visually impaired people, for this a feedback system is needed such as text to speech conversion in order to make users understand where they have to move. A different approach for background extraction could be used in order to obtain more accurate features which will indeed develop a robust and accurate system .

VI. Conclusion

We have written about multiple approaches utilized in the implementation of the SLAM algorithm. The resultant conclusion found from the survey was that the ORB SLAM algorithm is the most well suited to our application as it has a good balance of accuracy, real time output and accuracy. Along with this we will also be using GPS to map the generated map with real world coordinates, and to get the ground truth conditions. For the use of the visual input, we will be using a stereo camera to get the RBG as well as depth input. Later ,background extraction is done using CVS method to remove dynamic object from the frame , for which HSV color spaces are used to resist the daylight alterations.Stereo slam is implemented in MATLAB using UTIAS In the Dark dataset which gives us forward as well as reverse trajectory which can be given as input to a navigation system for visually impared people .

Acknowledgment

We thank Prof Jyoti Madake for her valuable guidance and assistance throughout the course of the project.

References

[1]Mohd Nadhir Ab Wahab , Ching May Lee, Muhammad Firdaus Akar And Fadratul Hafinaz Hassan Path Planning For Mobile Robot Navigation In Unknown Indoor Environments Using Hybrid PSOFS Algorithm

[2] Lili Mu, Pantao Yao , Yuchen Zheng, Kai Chen , Fangfang Wang, And Nana Qi Research On SLAM Algorithm Of Mobile Robot Based On The Fusion Of 2D Lidar And Depth Camera

[3] Yongbo Chen , Brenton Leighton, Huishen Zhu, Xijun Ke, Songtao Liu, And Liang Zhao Submap-Based Indoor Navigation System For The Fetch Robot

[4] Robust Stereo Visual SLAM For Dynamic Environments With Moving Object Gang Li, Xiang Liao, Huilan Huang , Shaojian Song , Bin Liu, Yawen Zeng

[5] Lingdong Zeng , Shuai Guo, Zhen Xu, And Mengmeng Zhu An Indoor Global Localization Technique For Mobile Robots In Long Straight Environments

[6] Duan, Y., Zhang, J., & Yang, L. (2018). Dense Mapping from Feature-Based Monocular SLAM Based on Depth Prediction. In 2018 IEEE CSAA Guidance, Navigation and Control Conference, CGNCC 2018. Institute of Electrical and Electronics Engineers Inc

[7] He, H., Jia, Y., & Sun, L. (2018). Simultaneous Location and Map Construction Based on RBPF-SLAM Algorithm. In Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018 (pp. 4907–4910). Institute of Electrical and Electronics Engineers Inc

[8] Almansa-Valverde, S., Castillo, J. C., & Fernández Caballero, A. (2012). Mobile robot map building from time-of-flight camera. Expert Systems with Applications, 39(10), 8835–8843.

[9] Whelan, T., Johannsson, H., Kaess, M., Leonard, J. J., & McDonald, J. (2013). Robust real-time visual odometry for dense RGB-D mapping. In Proceedings - IEEE International Conference on Robotics and Automation (pp. 5724–5731).

[10] Daniel, H. C., Kim, K., Kannala, J., Pulli, K., & Heikkilä, J. (2015). DT-SLAM: Deferred triangulation for robust SLAM. In Proceedings - 2014 International Conference on 3D Vision, 3DV 2014 (pp. 609–616). Institute of Electrical and Electronics Engineers Inc.

[11] Master of Science Thesis in Datorteknik, Fordons system Department of Electrical Engineering, Linköping University, 2017

[12] Fernando A Auat Cheein1*, Natalia Lopez2, Carlos M Soria1, Fernando A di Sciascio1, Fernando Lobo Pereira3,Ricardo Carelli1" SLAM Algorithm applied to robotic assistance for navigation in unknown environments ".Auat Cheein et al. Journal of NeuroEngineering and Rehabilitation 2010

[13] Udo Frese · René Wagner · Thomas Röfer" A SLAM Overview from a User's Perspective "Springer Verlag 2010 ,DOI 10.1007/s13218-010-0040-4

[14] Takafumi Taketomi ,Hideaki Uchiyama ,Sei Ikeda “ Visual SLAM algorithms :survey from 2010-2016” , IPSJ Transactions On Computer Vision and Applications ,2017

[15] Chunguang Li,1,2 Chongben Tao,3 and Guodong Liu1” SLAM Algorithm applied to robotic assistance for navigation in unknown environments “.Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 943510, 11 pages http://dx.doi.org/10.1155/2015/943510

[16] Raul Mur-Artal and Juan D. Tard ´ os” ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras“IEEE Transactions on Robotics DOI: 10.1109/TRO.2017.2705103

[17] Y. Cheng, J. Bai and C. Xiu, "Improved RGB-D vision SLAM algorithm for mobile robot," 2017 29th Chinese Control And Decision Conference (CCDC), Chongqing, 2017, pp. 5419-5423, doi: 10.1109/CCDC.2017.7979460.

[18] G. Xin, X. Zhang, Xi Wang and Jin Song, "A RGBD SLAM algorithm combining ORB with PROSAC for indoor mobile robot," 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), Harbin, China, 2015, pp. 71-74, doi: 10.1109/ICCSNT.2015.7490710..

[19] Aguilar-González, Abiel. (2019). Monocular-SLAM dense mapping algorithm and hardware architecture for FPGA acceleration.

[20] J. Engel, J. Stückler and D. Cremers, "Large-scale direct SLAM with stereo cameras," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015, pp. 1935-1942, doi: 10.1109/IROS.2015.7353631.

[21] Z. Wang, Q. Jia, P. Ye and H. Sun, "A depth camera based lightweight visual SLAM algorithm," 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, 2017, pp. 143-148, doi: 10.1109/ICSAI.2017.8248279.

[22] Yang A., Luo Y., Chen L., Xu Y. (2017) Survey of 3D Map in SLAM: Localization and Navigation. In: Fei M., Ma S.,

Li X., Sun X., Jia L., Su Z. (eds) Advanced Computational Methods in Life System Modeling and Simulation. ICSEE 2017, LSMS 2017. Communications in Computer and Information Science, vol 761. Springer, Singapore. https://doi.org/10.1007/978-981-10-6370-1_41

[23] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An Open

Source SLAM System for Monocular, Stereo, and RGB-D Cameras," in IEEE Transactions on Robotics, vol. 33, no.