Visual Positioning System for Automated Indoor/Outdoor Navigation

Anup S
PES University
Bangalore, India 560085

Abhinav Goel
PES University
Bangalore, India 560085

Suresh Padmanabhan
PES University
Bangalore, India 560085

***Abstract*—This paper proposes a proof-of-concept for a novel automated indoor/outdoor navigation system. Our proposed method shall enable an object/user equipped to be able to navigate through closed environments using an automatically generated Spatial Map Graph (SMG) with the aid of pre-placed visual markers. The system is robust to dynamically changing complex environments, through adaptive reconfigurations in the SMG during execution. We show that it is possible to find optimal routes among several interconnected paths without relying on an external positioning system. Points of interest in a passive environment are detected from real time first-person- view input, which is then processed to pinpoint the current location, and extended further for navigation. Through this paper we explore the advantages and the challenges faced in implementing such a navigation system in a practical scenario. The way forward and future work has also been discussed.**

***Keywords* - indoor navigation, localisation, computer vision, navigational guidance**

I. Introduction

The problem associated with simplifying human navigation has been pursued profusely over the decades[1]. Propositions have been made modelling a human's spatial knowledge[2]. Directions have been modelled as providing an action with reference to a landmark or a point-of-interest[3][4]. The process of navigation can be broken down into 3 main parts,

1) *Localisation* - identifying the current location accurately.
2) *Wayfinding* - identifying an ideal path to the destination based on required parameters (time, effort, safety)
3) *Guidance* - ensuring that the right route is being taken along the way.

Localisation has been pursued with the intent of increasing accuracy, by using *triangulation* methods involving GPS[5][6][7], Cell-tower[7], WLAN[8], RFID[9]. Triangulation methods can either be lateration(distance between user and 3 known points) or angulation(angular measurements between user and 3 known points)[10]. Apart from GPS - which can be used exclusively in outdoor applications - the other stated methods suffer from inaccuracy due to multipath reflections. Another class of methods commonly proposed are *dead-reckoning* methods. These estimate a user's location based on a previously known or estimated position by analysing odometry readings from gyroscopes, accelerometers, magnetometers and compasses[11] or using a walking pattern(average speed)[12]. It's usually combined with another system like GPS or RFID[11] to determine the initial location. Estimating the current location is a recursive process which results in errors building up over time. This inaccuracy and the need to combine with another method are major drawbacks. A common brute-force method is *direct-sensing* which determines the location of the user through sensing identifiers or tags installed in the environment. The location can either be stored on the tag itself or it is retrieved from a database using a unique identifier. One tag is sufficient to determine the location of the user and successive scans can be used to roughly determine the orientation of the user[13]. Multiple options are available in terms of choosing a technology for the tags, namely RFID[14], Infrared[15], Ultrasound Identification (USID)[16], Bluetooth beacons[17], Barcodes[18]. Normally, a single technology is chosen but multiple technologies have also been incorporated into a single system, albeit increasing accuracy, the additional expense and the extra equipment deems it ineffective.

We are proposing a *Visual Positioning System*(VPS) that is capable of providing an optimised route through the points of interest, by detecting the current location using passive visual markers and mapping the points of interest to an SMG. The SMG is used to enable cognitive path finding by providing appropriate instructions based on the current location and the tentative destinations. There are two possible situations,

1) *Single Point of Interest* - A shortest path algorithm can be used to navigate to the destination.
2) *Multiple Points of Interest* - This poses an NP-hard problem, which requires an efficient solution to the Travelling Salesman Problem.

This system will work in unknown environments with minimal setup, as the machine learning approach will remodel the interest graph and the corresponding SMG to obtain the most optimal route. The adaptive nature of the system will ensure robust performance even in dynamic environments(e.g., aisle configuration is often changed to prompt customers towards new products).

The main contributions of this paper can be described as a novel approach to attempt to improve the ease of implementing a visual navigation system based on passive visual markers placed near points of interest. This technology attempts to simplify indoor navigation in all closed spaces to provide a quick, easy and convenient method to interact with the environment.

978-1-5090-1134-6/17/$31.00 ©2017 IEEE

In the field of indoor navigation, different types of technologies have been introduced in recent years. We outline some of these implementations in *Section 2*. We move on to provide a proof of concept for our current proposed method in *Section 3*

II. Related Work

Some early methods of successful indoor/outdoor navigation include localisation methods using GPS and Ultrasound(L. Ran and S. Helal and S. Moore 2004[16]) in which GPS was used outdoors and a transition was made indoors to an OEM ultrasound position system achieving an indoor accuracy of 22cm. This was mainly focused towards helping visually impaired individuals to navigate in an environment. The transition to ultrasound indoors is needed because of the limitations of GPS. However, the system required a wearable computer, a differential GPS receiver, an ultrasound positioning device and several software components. More recent methods employing the use of triangulation methods using Bluetooth beacons(Ahmetovic et. al., 2017[19]) and localising by calculating the *Received Signal Strength Indicator*(RSSI) probability distribution and k-NN regression to receive the current location. This method achieves an accuracy that largely depends on the distance between the placed beacons and users were subjected to a single path between two points. Another approach proposed making use of only cameras [O. Koch and S. Teller 2008[20]] reduces the number of components carried by the end user to just a single camera with good accuracy. An exploration step is used to capture features along the entire path. This requires the use of a large database of features which introduces problems in scaling to larger, more complex environments. A more hybrid method combining Wi-Fi and cameras has been presented(Dao et. al., 2016][21]) and it was found that sufficiently accurate localisation can be performed with just cameras. In order to cope with the complexity, scale and the need for relative simplicity, we are attempting to solve this problem with the following methodology.

Method	**Characteristics**	**Remarks**
Ultrasound	Combatting the inaccuracy of GPS indoors using Ultrasound	Requires umpteen components on and off the object trying to navigate.
Bluetooth	Local triangulation using beacons	Relies heavily on the placement of the beacons. Multiple beacons contribute heavily to the total cost of the system.
Cameras	Single camera on the object trying to navigate the environment.	Requires a large database of features as the entire image is being used to localise.

TABLE I
Comparison of existing techniques.

III. Proposed Methodology

In this section we describe the fundamental and technical details of our proposed VPS implementation. The object that will be used for indoor navigation has to be equipped with a camera to perceive visual features from the environment. According to models of wayfinding[1], acquiring these levels of knowledge is necessary, namely survey knowledge(knowledge about the environment), route knowledge(knowledge about the paths between the nodes), destination knowledge(knowledge about the final destination).

We propose a system, which can extract the cognitive intelligence from its surroundings with minimal computation and high accuracy, using visual perception markers. It consists of:

1) *Priming the environment* - Initial set up, which involves manual identification of landmarks and affixing a visual marker in its vicinity.
2) *Training* - Learning and adapting to the dynamic maps of the environment autonomously.
3) *Localisation* - Extracting the exact location of the user in a complex environment.
4) *Path finding* - Computationally determine the most optimal path to the destination.
5) *Guidance* - Graded actions and responses to navigate the object along the path.

A. Priming the environment

In order to facilitate spatial cognition, the environment must be demarcated manually for landmarks and points of interest. Each point of interest would be considered to be a node in our *Spatial Map Graph* (SMG). Each node will be associated to a unique visual marker. Different shapes, colours, arrangements and signage can demarcate these markers[22]. QR Codes are viable markers to provide a distinct and acceptable representation of the nodes.

In order to reduce complexity, features of the QR Codes can be stored in a database instead of saving the whole image. To extract these features, two viable methods are *Scale Invariant Feature Transform* (SIFT) and *Speeded Up Robust Features* (SURF). The SIFT algorithm was proposed by Lowe[23] in 2004 to solve rotation, scaling and affine deformation. The SURF algorithm is based on multi-scale space theory and the feature detector is based on Hessian Matrix since is performs well. In an image I, x = (x,y) is the given point, the Hessian matrix in x at σ can be defined as

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{yx}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}$$

Where $L_{xx}(x,\sigma)$ is the convolution result of the second order derivative of Gaussian filter $\frac{\partial^2 g(\sigma)}{\partial x^2}$ with the image in point x, and similarly for $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$. This is an important part of the process and care has to be taken to avoid ambiguity with the markers as the training phase entirely depends on the markers affixed in this stage.

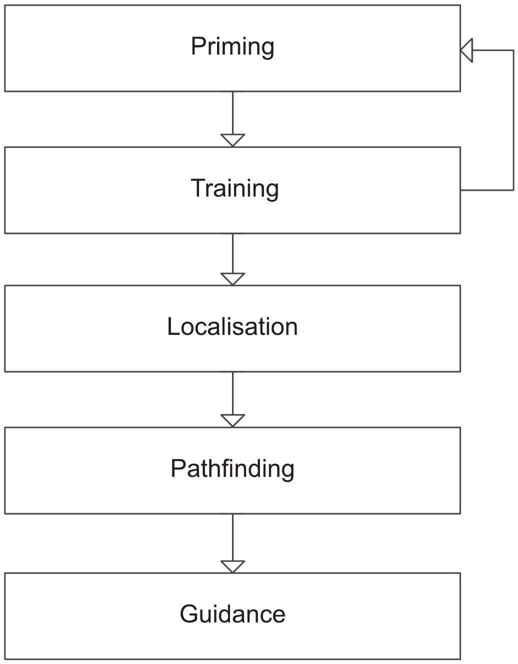

Fig. 1. Flow diagram.

The following actions would be required to prime the environment (Indoor):

1) Identify distinct nodes in the environment.
2) Generate QR codes for said nodes.
3) Affix these markers along the ceiling as the camera can obtain an unobstructed view.
4) Ensure markers are not obfuscated by scintillating light sources.
5) Camera should be oriented optimally to detect the markers.

The following actions would be required to prime the environment (Outdoor):

1) Identify distinct nodes in the environment.
2) Generate QR codes for said nodes.
3) Affix these markers on towering objects like lamp posts, trees etc.,.
4) Ensuring QR codes are appropriately sized for the application.
5) Camera should be oriented optimally to detect the markers.

Due to the easy accessibility and distinctness of QR Codes, their ability to demarcate separate nodes is optimal. They can be easily generated and can also be deployed efficiently. The uniqueness of QR codes ensures less matching error which leads to reliable localisation of the object, compared to matching features in an environmental image. The remaining steps are consistent for both indoor and outdoor systems.

B. Training

Once the nodes for the SMG have been defined, the interconnections are to be formed autonomously. This can be achieved by manoeuvering the object around the environment, such that it can identify adjacent nodes and construct an edge between them in the SMG. An example of a generated SMG is depicted in Fig. 2.

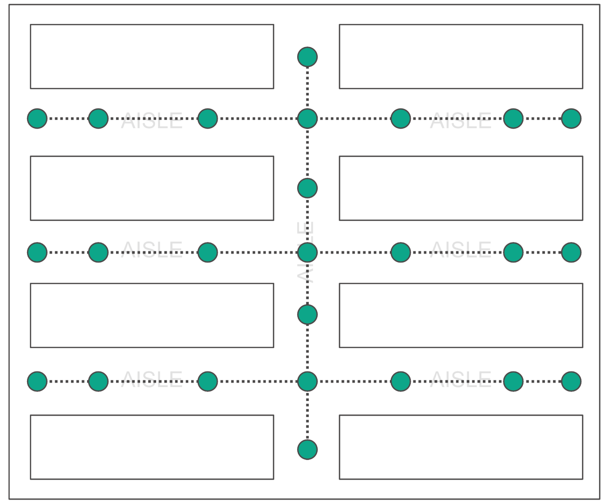

Fig. 2. Generated SMG.

Each node will contain the relative direction of translation to each adjacent node with respect to itself[20]. It would be stored as phrases like ”forward-left”, ”forward-right”, ”forward-center”, ”backward-left”, ”backward-right”, ”backward-center” etc. This process can be completed using a semi-supervised reinforced learning algorithm as described in [24], which works as a classifier. This will enable the graph to be reconfigured and reinforced during runtime. The wayfinding techniques will be described in subsection 3.4.

C. Localisation

For real time FPV captured by the object, we propose to localise an object by using distinct and informative visual markers present in the environment. We would first perform a scene analysis,

1) Convert each frame into grayscale.
2) Segment the image into small sections such that it is guaranteed that at least one section will be fully covered by a QR Code, if present.
3) Perform 2D Fourier Transform on all sections.

If there is a section with significantly high frequency values in both X and Y directions then, there is a high probability that there is a QR code present in the scene. Once a QR code is detected in the scene, we would extract the *Speeded Up Robust Features* (SURF)[25] Descriptors and compare it against the database of the SURF Descriptors of the aforementioned markers. We will use the Euclidean distances to filter out the outlying key points and then using the *Fast Library for Approximate Nearest Neighbours* (FLANN) [26] we can estimate the match coefficient. The marker that gives us the best match will allow us to pinpoint the current location of the object, and hence will provide a starting point for navigation.

As the feature matching is performed only after confirming the presence of a QR code in the scene, the complexity of the system is reduced drastically. The matching process can be optimised by ensuring the markers in adjacent nodes are highly distinguishable to avoid a misclassification.

D. Path Finding

With the current location of the object determined, the path to the destination must be computed. By minimising the path weight or cost associated with each edge of the SMG, the most optimal path can be identified. The weight or cost can be expressed in terms of distance, risk, crowd or other quantifiable parameters along the path depending on the user's preference. The number of nodes encountered and the path determined may vary between objects, due to different preferences. Shortest path routing algorithms [27] can be used to enable the object to find the best path towards the destination. These algorithms can be implemented with a worst-case time complexity of $O(Elog(V))$ [28] where E is the number of edges and V is the number of nodes in the graph. The algorithm would choose the most optimal path in a complex SMG. The SMG will be a dense graph when there exist several possible routes and paths between a pair of nodes.

E. Guidance

The final component of the system would allow the object to navigate the user through the intermediate nodes to the final destination. The object will prompt the user to move in a direction towards the next node in the SMG. The user will be subjected to text-based, speech-based or visual cues, as described in subsection 3.2. The object will continuously monitor the environment for presence of visual perception markers. The SURF Descriptors of the detected marker will be compared against the database of the SURF Descriptors of the known markers, as described in section 3.3. If the match coefficient obtained is greater than an acceptable threshold value, the object classifies the marker. To further optimise the matching process, the expected next node from the path finding algorithm will be matched first. This would be followed by matching the frequently coupled nodes, which the machine would learn from experience. There are 3 possible outcomes:

- *The marker detected is associated with the expected next node* - In this case, the user is on the suggested track. The current position will be updated to the new node and the user would be notified with the next prompt.
- *The marker detected is not associated with the expected next node* - The user may have taken a wrong turn or misinterpreted the prompt. In this case, the shortest path algorithm, described in subsection 3.4, would be implemented again with the current node as the starting location. The user will be alerted and a new set of prompts would be relayed.
- *No markers have been detected for a considerable amount of time* - In this case, the user may have strayed off course into an unknown or forbidden region. The user will be alerted and asked to retrace their steps back towards the last node.

IV. Future Work

The generation of the SMG is performed using machine learning. Data collection is critical for this process, and the training process must continue even after a preliminary SMG has been generated in order to dynamically adapt the SMG in an ever-changing environment. The efficacy of existing machine learning techniques is to be further investigated. A verification technique must be developed and made available to the user in order to deploy a robust system.

V. Conclusion

We have proposed a vision-based indoor/outdoor navigation system that is relatively easy to install with low set-up costs to assist people/objects to traverse in complex environments. Our method hopes to bridge the gap between unfamiliar environments and people. Using an approach that requires minimum calibration opens our method up to a wide variety of environments like campuses, retail locations, museums and collaborative workspaces.

We hope to solve the problems that might be associated with implementing this system in real world environments. Occlusion due to large crowds may be solved using some form of dead-reckoning. Marking points of interest and using a design-thinking approach to building interfaces to the system is an important task ahead. Also, we would like to extend this system to communicate with smart devices like lights and fans in a space and perform actions based on user navigation.

References

[1] Jan M. Wiener, Simon J. Bchner, and Christoph Hlscher. "Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach". In: *Spatial Cognition & Computation* 9.2 (2009), pp. 152–165. DOI: 10.1080/13875860902906496.

[2] Benjamin Kuipers. "Modeling Spatial Knowledge*". In: *Cognitive Science* 2.2 (1978), pp. 129–153. ISSN: 1551-6709. DOI: 10.1207/s15516709cog0202_3.

[3] Marie-Paule Daniel and Michel Denis. "Spatial Descriptions as Navigational Aids: A Cognitive Analysis of Route Directions". In: *Kognitionswissenschaft* 7.1 (1998), pp. 45–52. ISSN: 1432-1483. DOI: 10.1007/BF03354963.

[4] Kai-Florian Richter and Alexander Klippel. *A Model for Context-Specific Route Directions*. Ed. by Christian Freksa, Markus Knauff, Bernd Krieg-Brückner, et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 58–78. ISBN: 978-3-540-32255-9. DOI: 10.1007/978-3-540-32255-9_4.

[5] Simon Holland, David R. Morse, and Henrik Gedenryd. "AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface". In: *Personal and Ubiquitous Computing* 6.4 (2002), pp. 253–259. ISSN: 1617-4909. DOI: 10.1007/s007790200025.

[6] Wilko Heuten, Niels Henze, Susanne Boll, et al. "Tactile Wayfinder: A Non-visual Support System for Wayfinding". In: *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges*. NordiCHI '08 (2008), pp. 172–181. DOI: 10.1145/1463160.1463179.

[7] M. Arikawa, S. Konomi, and K. Ohnishi. "Navitime: Supporting Pedestrian Navigation in the Real World". In: *IEEE Pervasive Computing* 6.3 (July 2007), pp. 21–29. ISSN: 1536-1268. DOI: 10.1109/MPRV.2007.61.

[8] Vassileios Tsetsos, Christos Anagnostopoulos, Panayotis Kikiras, et al. "Semantically Enriched Navigation for Indoor Environments". In: *Int. J. Web Grid Serv.* 2.4 (Dec. 2006), pp. 453–478. ISSN: 1741-1106. DOI: 10.1504/IJWGS.2006.011714.

[9] T. Amemiya, J. Yamashita, K. Hirota, et al. "Virtual leading blocks for the deaf-blind: a real-time way-finder by verbal-nonverbal hybrid interface and high-density RFID tag space". In: *IEEE Virtual Reality 2004* (Mar. 2004), pp. 165–287. ISSN: 1087-8270. DOI: 10.1109/VR.2004.1310070.

[10] Pei Zheng and Lionel Ni, eds. Burlington: Morgan Kaufmann, 2006, pp. xvii –. ISBN: 978-0-12-088560-2. DOI: 10.1016/B978-012088560-2/50000-9.

[11] S. Koide and M. Kato. "3-D human navigation system considering various transition preferences". In: *2005 IEEE International Conference on Systems, Man and Cybernetics* 1 (Oct. 2005), 859–864 Vol. 1. ISSN: 1062-922X.

[12] Hua Wu, Alan Marshall, and Wai Yu. "Path Planning and Following Algorithms in an Indoor Navigation Model for Visually Impaired". In: *Proceedings of the Second International Conference on Internet Monitoring and Protection*. ICIMP '07 (2007), pp. 38–. DOI: 10.1109/ICIMP.2007.31.

[13] Scooter Willis and Sumi Helal. "RFID Information Grid and Wearable Computing Solution to the Problem of Wayfinding for the Blind User in a Campus Environment". In: (2005).

[14] Vladimir Kulyukin, Chaitanya Gharpure, John Nicholson, et al. "Robot-assisted wayfinding for the visually impaired in structured indoor environments". In: *Autonomous Robots* 21.1 (2006), pp. 29–41. DOI: 10.1007/s10514-006-7223-8.

[15] Jörg Baus, Antonio Krüger, and Wolfgang Wahlster. "A Resource-adaptive Mobile Navigation System". In: *Proceedings of the 7th International Conference on Intelligent User Interfaces*. IUI '02 (2002), pp. 15–22. DOI: 10.1145/502716.502723.

[16] L. Ran, S. Helal, and S. Moore. "Drishti: an integrated indoor/outdoor blind navigation system and service". In: *Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the* (Mar. 2004), pp. 23–30. DOI: 10.1109/PERCOM.2004.1276842.

[17] H. Huang, G. Gartner, M. Schmidt, et al. "Smart Environment for Ubiquitous Indoor Navigation". In: *2009 International Conference on New Trends in Information and Service Science* (June 2009), pp. 176–180. DOI: 10.1109/NISS.2009.16.

[18] Yao-Jen Chang, Shih-Kai Tsai, and Tsen-Yung Wang. "A Context Aware Handheld Wayfinding System for Individuals with Cognitive Impairments". In: *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*. Assets '08 (2008), pp. 27–34. DOI: 10.1145/1414471.1414479.

[19] Dragan Ahmetovic, Masayuki Murata, Cole Gleason, et al. "Achieving Practical and Accurate Indoor Navigation for People with Visual Impairments". In: (2017).

[20] Olivier Koch and Seth Teller. "A Self-Calibrating, Vision-Based Navigation Assistant". In: *Workshop on Computer Vision Applications for the Visually Impaired* (Oct. 2008).

[21] T. K. Dao, T. H. Tran, T. L. Le, et al. "Indoor navigation assistance system for visually impaired people using multimodal technologies". In: *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)* (Nov. 2016), pp. 1–6. DOI: 10.1109/ICARCV.2016.7838771.

[22] P. Arthur and R. Passini. *Wayfinding: People, Signs, and Architecture*. Limited Commemorative Edition. Focus Strategic Communications, 1992. ISBN: 9780973182200.

[23] D.G. Lowe. "Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image". In: (Mar. 2004). US Patent 6,711,293.

[24] Liyuan Li, Weixun Goh, Joo Hwee Lim, et al. "Extended Spectral Regression for efficient scene recognition". In: *Pattern Recognition* 47.9 (2014), pp. 2940 –2951. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2014.03.012.

[25] Herbert Bay, Andreas Ess, Tinne Tuytelaars, et al. "Speeded-Up Robust Features (SURF)". In: *Comput. Vis. Image Underst.* (2008). DOI: 10.1016/j.cviu.2007.09.014.

[26] Marius Muja and David G Lowe. "Fast approximate nearest neighbors with automatic algorithm configuration." In: *VISAPP (1)* 2 (2009).

[27] Chang Wook Ahn and Rudrapatna S Ramakrishna. "A genetic algorithm for shortest path routing problem and the sizing of populations". In: *IEEE transactions on evolutionary computation* 6 (2002).

[28] Edsger W Dijkstra. "A note on two problems in connexion with graphs". In: *Numerische mathematik* 1.1 (1959), pp. 269–271.