2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 4-7 October 2016, Alcalá de Henares, Madrid, Spain

Dual Scaling and Sub-model based PnP Algorithm for Indoor Positioning based on Optical Sensing using Smartphones

Arief Affendi Juri[1], Tughrul Arslan[1,2], Yichen Du[1], Zekun Wang[1]
[1]School of Engineering, University of Edinburgh, United Kingdom
[2]Sensewhere Limited, Edinburgh, United Kingdom
A.Juri@ed.ac.uk, T.Arslan@ed.ac.uk, s1439253@sms.ed.ac.uk, s1473591@sms.ed.ac.uk

***Abstract*— Traditionally, optical-based positioning has been an area that attracted high interest for specialised application such as robotic navigation. With the recent development and integration of cameras in mobile devices, optical positioning is gaining further interest in indoor positioning application for human navigation. Furthermore, the release of Google Tango has attracted the interest of several researchers for improving optical-based positioning using 3D mapping. However, estimating 3D pose based on PnP problem has been a challenge which affects the positioning accuracy. This paper proposes two novel strategies to improve PnP solution algorithms, and hence accuracy. The first "scaling" strategy, is based on minimising model size, with the second, "sub-model" strategy, involving selection of only the related area of the model to be used. The proposed strategies also have the advantage of limiting the error to the set scale size. The scale and sub-model strategies showed an average improvement of 1.61 and 3.33 m, respectively.**

Keywords—optical positioning, PnP, Google Tango,

I. Introduction

Location-based services (LBS) are widely used today. Most of them depend on Global Navigation Satellite Systems (GNSS) to provide accurate and reliable positioning. However, such positioning techniques cannot be supported in indoor environments, where the satellite signal is often blocked by walls and ceilings [1]. Accordingly, numerous technologies, such as RF, ultrasonic, optical, and infrared IR, have been used to detect users' indoor locations with convincing results.

Optical-based positioning is effective, as it require no additional infrastructure, such as beacons and access points [2]. Whilst many studies have focused on mobile robot navigation [2], such as [3][4] and [5], more recently there has been increased interest in human optical positioning, such as in [6][7] and [8], as smartphones have become more common.

The release of Google's Project Tango has further increased interest in optical-based positioning, particularly using 3D maps, such as [9][10] and [11]. Compare to normal vision sensors embedded in mobile phones, Project Tango brings a new spatial visualisation and localisation, by updating advanced computer vision and locomotion, and embedding a depth sensor within mobile devices [12].

Generally, optical positioning based on cameras is determined by comparing images with corresponding, pre-recorded images. The latter are found by searching similar images to the image captured at the current position. Optical positioning can also be used with a 3D map, which often employs 6-DOF localisation. The 3D map consists of 3D-point clouds, and is created using 3D sensors such as Microsoft Kinect, a stereo camera or the Project Tango device. An image-matching algorithm searches similarities between query images and the map. Once the corresponding image is found, a Perspective-n-Point (PnP) problem is solved, in order to determine the camera pose.

One major challenge of 3D visual indoor positioning is finding an optimal solution to solve the PnP problem. Existing solutions can be divided into two categories: iterative and direct [13]. Whilst the former are more accurate than the latter, they tend to have higher computing costs and are slower to compute [14]. This may be problematic for devices with limited computing capabilities, such as smartphones. As a less computing costs and faster computing solution, direct solutions use subsets of n points in calculations. For example, [15] introduced a new non-iterative solution to the PnP problem, utilising n 3D points as a weighted sum of four virtual control points. The problem is then simplified to just estimating the coordinates of these control points in the camera referential.

In this paper, in order to eliminate the problem of large model size and to lower computational complexity, we proposed two novel strategies to improve PnP without directly modifying the PnP algorithm. These two optimisation algorithms attempt to address this problem by reducing the room model size, or limiting the area worked on. The first strategy is based on direct solutions that scale down a large-sized model, in order to tackle the PnP problem more conveniently. The other strategy selects the related portion of the whole model to be computed, instead of processing the entire testing area, and is an efficient way to reduce computational complexity.

II. Perseptive-n-Point

PnP is defined as the problem to estimate the 3D pose or the 6 degree-of-freedom (DOF), given the 3D location features and their corresponding 2D image. [16] derived PnP as finding the length of the line segments that join the centre of perspective (camera origin point) to each of the control points (features in 2D image), given that the relative spatial location of n control points, and the angle of every pair control points to the centre of perspective, is known.

The PnP problem with n=1 or n=2 does not provide constraining information, thus an infinite number of solutions are possible for these conditions. For PnP problem with n=3, three lengths of tetrahedron are formed by the three points a, b, c and the centre of perspective. [16] states that the solution to this problem is implied by:

$$(Rab)^2 = a^2 + b^2 - 2 * a * b[\cos(\theta ab)] \tag{1}$$

$$(Rac)^2 = a^2 + c^2 - 2 * a * c[\cos(\theta ac)] \tag{2}$$

$$(Rbc)^2 = b^2 + c^2 - 2 * b * c[\cos(\theta bc)] \tag{3}$$

As there is no global optimal solution which is accurate and applicable to any problem with n > 3, a variety of solutions have been proposed for the PnP problem, such as [17], [18] or a more recent [19]. These solutions vary, with some focusing on specific point configurations, whilst others tackle a more general case.

III. Proposed Strategies for PnP Improvement

Two strategies are proposed to improve PnP results. These may improve PnP without actually modifying the PnP algorithm. Errors of 3D pose detection increase with room model size. The proposed strategies endeavour to counter this problem by reducing room model size, or limiting the area worked on.

A. Scale Strategy

This strategy uses a direct approach to tackle a large sized model by scaling it down. A specific size limit is set as the default value for the scale size. Currently, 3 m is chosen as the scale size, as this is about the width of a small office room. Ideally, the aim is to limit the error to within the scale size.

This strategy is performed by taking the model's largest dimension value, typically the length, and dividing it by 3. This value will be the scale ratio for this specific model's scale process. The model's other dimension size is then scaled down based on the calculated scale ratio. For example, a model with the dimension of 10 x 3 m will have a scale ratio of 1:3.33. Scaling down the model using this ratio produces a model of 3 x 0.9 m.

All point clouds' coordinates for the model are also scaled down, using the same scale ratio as for the entire model, and the information contained within it. The PnP algorithm is then performed as usual, using the new scaled-down model point clouds as the inputs.

B. Sub-model Strategy

This strategy attempts to minimise the area to be worked on. Out of the original model area, only a portion of the model is related to the positioning process. This strategy focuses only on the area which is useful for the positioning process. This is performed by creating a sub-model and ideally, the positioning process will only focus on the new sub-model area.

The sub-model is assigned based on the result of a 2D-to-3D correspondences algorithm. When an image is captured for positioning, it is compared with the database. The 2D-to-3D correspondences are calculated, and will provide 3D correspondences which show the parts of the model that are being used as the reference. This is the area which will be used as the sub-model.

The steps to assign the sub-model are executed by initially determining the 3D correspondences of the captured image. Then, the centre point of the 3D correspondences are calculated. All 3D correspondence points are then deducted with the value of the centre points. The idea behind this step is to move the model's origin point to the centre of the 3D correspondences. This creates a smaller model, or a sub-model. The PnP is then calculated using the new 3D correspondence values. Finally, the centre point is added back to the result point, to return it to the position within the actual model.

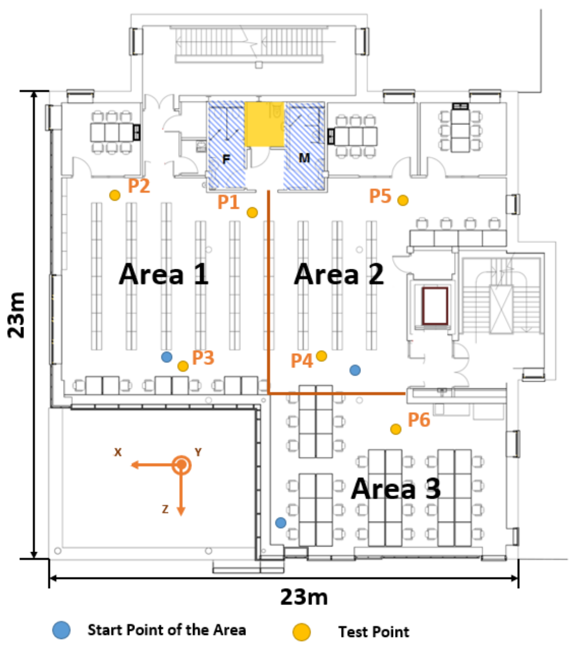

Fig. 1 Noreen and Kenneth Murray Library floor plan

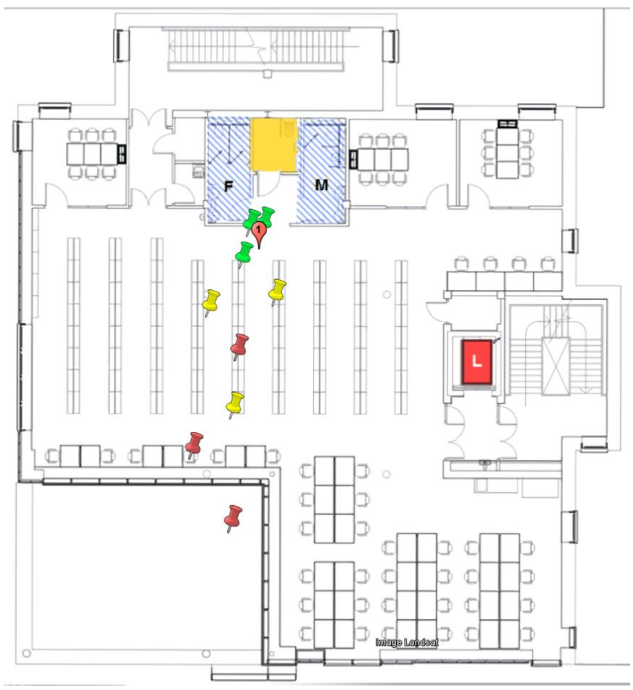

Fig. 2 Visual result of test point 1

TABLE 1. RESULTS FOR TEST POINT 1

Strategy		Error (m)	Mean Error (m)
Standard	1	2.60	4.72
	2	6.24	
	3	5.33	
Scale	1	2.57	2.58
	2	3.85	
	3	1.33	
Sub-model	1	0.57	0.60
	2	0.36	
	3	0.88	

IV. IMPLEMENTATION

The strategies are implemented and tested in the Noreen and Kenneth Murray Library in Kings Building, University of Edinburgh. The model and database are built and created using Google's Tango Tablet. An RGBDSLAM algorithm is used to create the model and database for this implementation.

Positioning is performed using the same device, but with a single camera. Six test points are taken and in each, three images are captured, each with a different image. The images are then processed using a standard algorithm, scale strategy and sub-model strategy. The feature-matching algorithm used in this test is the Brute Force matcher. This was found to be adequate for the purpose of this paper as this is the basic feature-matching algorithm with a good detection rate. The PnP solution used for this implementation is an iterative method based on Levenberg-Marquardt optimisation.

V. RESULTS

As this paper focuses on the difference made by the proposed strategies, any errors due to mismatched images are left out. This way, the only modifiers that can affect the results are the proposed strategies.

Fig. 1 shows the Noreen and Kenneth Murray Library floor plan and the six test points' locations. The floor plan is divided into three areas to minimise mismatching. Blue dots indicate the starting point of the corresponding area. The results for test points 1, 3 and 6 will be shown individually, followed by overall results for the entire six test points.

Fig. 2 shows the results for test point 1. The numbered red marker indicates the actual test point location. The red pins signal use of the standard algorithm, yellow pins indicate the scale strategy and green pins the sub-model strategy. The results clearly show the accuracy of the three strategies used for this test point. The standard algorithm results scattered further away from the actual test point, whilst the scale strategy results shows improvement compared to the standard algorithm. The sub-model strategy showed the best results, which lay in close proximity to the actual test point. Table 1 displays the errors for all strategies used for test point 1. The mean error for the scale strategy was well within the scale size set during the scaling down process. The sub-model strategy shows a significant improvement, with only a 0.60 m mean error.

Fig. 3 shows the results for test point 3. Almost all results clustered in proximity to the actual test point. Nevertheless, the sub-model strategy still showed better results, although the scale strategy was slightly less accurate than the standard algorithm as shown in Table 2. Despite this, results were all still well within the scale size set at 3 m. Notice the distance between test points 1 and 3 compared to the area origin point. The closer the test point is to the origin point, the more likely results tend to be more accurate. This is the error that the sub-model tries to tackle and improve.

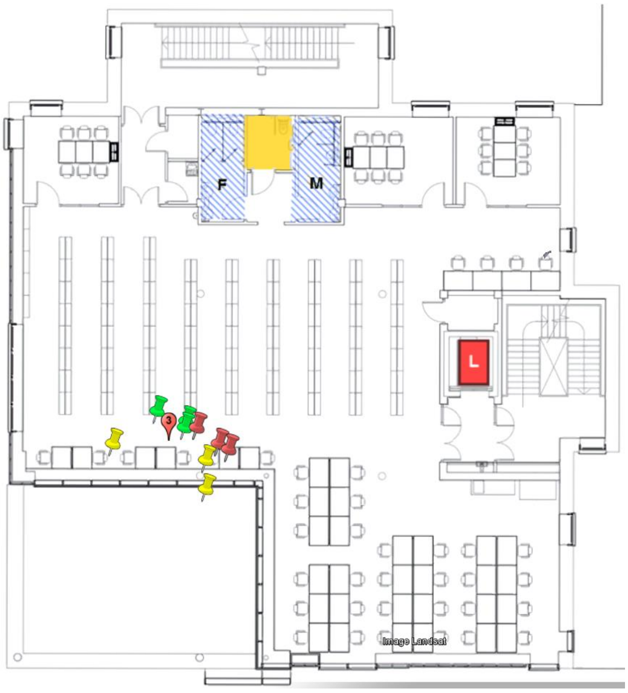

Fig. 3 Visual result for test point 3

Table 3 shows the overall test results. In most tests, the scale strategy was able to improve results compared to the standard algorithm. It attempts to limit errors within the scale size but in some cases, it does have errors larger than this size. The sub-model, however, managed to improve results in all test points. It had an accuracy to within 1 m for all test points except test point 6. On average, the sub-model strategy improved the standard algorithm by 3.33 m, whilst the scale strategy improved accuracy by 1.61 m.

VI. Conclusion

This paper has presented two new strategies to be used in conjunction with PnP algorithms. The proposed strategies tackle the problem of large model size. In doing so, they also limit the error within the set scale size.

The proposed strategies were implemented and compared to the same PnP algorithm but without the proposed strategies. The results showed a significant improvement to the standard implementation of PnP, with an average improvement of 1.61 m for the scale strategy and 3.33 m for the sub-model strategy.

The best improvement can be seen in conditions where the camera position is further away from the model's origin point, where the positioning error is more obvious, as in the example of test point 1.

TABLE 2. Results for test point 3

Strategy		**Error (m)**	**Mean Error (m)**
Standard	***1***	1.87	1.69
	2	2.24	
	3	0.95	
Scale	***1***	1.49	1.88
	2	2.32	
	3	1.84	
Sub-model	***1***	0.52	0.57
	2	0.70	
	3	0.47	

TABLE 3. Overall Implementation Results

Test Point		**Mean Error (m)**
Test Point 1	***Standard***	4.72
	Scale	2.58
	Sub-model	0.60
Test Point 2	***Standard***	5.23
	Scale	2.07
	Sub-model	0.78
Test Point 3	***Standard***	1.69
	Scale	1.88
	Sub-model	0.57
Test Point 4	***Standard***	1.35
	Scale	0.48
	Sub-model	0.56
Test Point 5	***Standard***	6.56
	Scale	4.51
	Sub-model	0.73
Test Point 6	***Standard***	5.07
	Scale	3.46
	Sub-model	1.38

References

[1] T. King, H. Lemelson, A. Färber, and W. Effelsberg, "BluePos: Positioning with bluetooth," in *WISP 2009 - 6th IEEE International Symposium on Intelligent Signal Processing - Proceedings*, 2009, pp. 55–60.

[2] X. Li, J. Wang, A. Olesk, N. Knight, and W. Ding, "Indoor positioning within a single camera and 3D maps," in *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service, UPINLBS 2010*, 2010, pp. 1–9.

[3] Y. Kim, J. Lee, M. S. Aldosari, M. S. Altokhais, and J. Lee, "Vision-Based Corridor Path Search of a Mobile Robot," in *2011*

11th International Conference on Control, Automation and Systems, 2011, pp. 700–705.

[4] E. Rivlin, I. Shimshoni, and E. Smolyar, "Image-based robot navigation in unknown indoor environments," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, 2003, vol. 3, pp. 2736–2742.

[5] W. T. Huang, C. L. Tsai, and H. Y. Lin, "Mobile robot localization using ceiling landmarks and images captured from an RGB-D camera," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, 2012, pp. 855–860.

[6] H. Hile and G. Borriello, "Positioning and orientation in indoor environments using camera phones," *IEEE Comput. Graph. Appl.*, vol. 28, no. 4, pp. 32–39, Aug. 2008.

[7] N. Ravi, P. Shankar, a. Frankel, a. Elgammal, and L. Iftode, "Indoor Localization Using Camera Phones," in *Seventh IEEE Workshop on Mobile Computing Systems & Applications (WMCSA'06 Supplement)*, 2006, vol. Supplement, pp. 1–7.

[8] H. Morimitsu, M. Hashimoto, R. B. Pimentel, R. M. Cesar, and R. Hirata, "Keygraphs for sign detection in indoor environments by mobile phones," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6658 LNCS, pp. 315–324.

[9] C. Sweeney, J. Flynn, B. Nuernberger, M. Turk, and T. Hollerer, "Efficient Computation of Absolute Pose for Gravity-Aware Augmented Reality," in *2015 IEEE International Symposium on Mixed and Augmented Reality*, 2015, pp. 19–24.

[10] H.-C. Wang, C. Finn, L. Paull, M. Kaess, R. Rosenholtz, S. Teller, and J. Leonard, "Bridging text spotting and SLAM with junction features," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 3701–3708.

[11] W. Winterhalter, F. Fleckenstein, B. Steder, L. Spinello, and W. Burgard, "Accurate indoor localization for RGB-D smartphones and tablets given 2D floor plans," 2015, pp. 3138–3143.

[12] "Project Tango." [Online]. Available: https://developers.google.com/project-tango/. [Accessed: 23-May-2016].

[13] E. Deretey, M. T. Ahmed, J. A. Marshall, and M. Greenspan, "Visual Indoor Positioning with a Single Camera Using PnP," in *2015 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2015*, 2015, pp. 1–9.

[14] S. Zhang, X. Cao, F. Zhang, and L. He, "Monocular vision-based iterative pose estimation algorithm from corresponding feature points," *Sci. China Inf. Sci.*, vol. 53, no. 8, pp. 1682–1696, Jul. 2010.

[15] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, Jul. 2008.

[16] M. a. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[17] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 930–943, Aug. 2003.

[18] C. P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 610–622, Jun. 2000.

[19] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom, and M. Okutomi, "Revisiting the PnP problem: A fast, general and optimal solution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2344–2351.