2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN) | 978-1-6654-0402-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IPIN51156.2021.9662575

BIPS: Building Information Positioning System

Max Jwo Lem Lee, Hiu Yi Ho, Li-Ta Hsu, Stephen Ling Ming Au

Abstract— **With the rise of digital twins and smart cities, Building Information Modelling have been widely adopted by the construction industry from design, construction to operation & maintenance. We present a BIPS (Building Information Positioning System) method which integrates a smartphone VPS (visual positioning system) based on the BIM models, and VO (visual odometry) for the indoor positioning. Firstly, the smartphone images and sensor measurements are sent to a server. In the server, the VPS utilizes computer vision algorithms to extract semantics from the smartphone images. Then, the smartphone image semantics are compared with the BIM semantics. The hypothesized position candidates are distributed in the BIM model. The candidate with the maximum likelihood is regarded as the VPS heading and position estimation. An extended Kalman filter is then used to integrate the VPS with VO, where the former and latter provide measurement and propagation models, respectively. According to the simulation result, the proposed BIPS proves effective in an indoor environment, being capable of improving indoor positioning accuracy to about 1 meter.**

Index Terms— **Visual Positioning System, Indoor Positioning, Kalman Filter, BIM, Semantics**

I. Introduction

INDOOR positioning systems have attracted a great interest from the researchers over the past decade. These systems can provide positioning, navigation, tracking services where Global Navigation Satellite Systems (GNSSs) could not reach [1]. To overcome the limitation of GNSS in indoor environments, various indoor positioning systems have been developed using Wi-Fi [], Bluetooth (BLE), ultra-wideband (UWB), and radio-frequency identification (RFID). Typically, beacon nodes are a prerequisite to localize in the indoor environment [2]. At a minimum, three non-collinear beacons are required to define a global coordinate system in two dimensions. The localization is usually performed by three main techniques: triangulation, received signal strength (RSS) scene analysis (fingerprinting) and proximity based. However, the radio ranging measurements contain noise on the order of several meters. This noise occurs because radio propagation tends to be highly non-uniform. Physical obstacles such as walls, furniture, etc. reflect and absorb radio waves. As a result, the distance estimation is deteriorated by the diffractions and reflections. In addition, the resource constraints like battery life and transmitter power, sensor network node density, installation cost and dynamic obstacle interference have increase challenges toward scaling in indoor environments. Therefore, to be sustainable, a software-based solution is essential.

In the recent years, 3D building information modeling (BIM) has matured to a stage where it is playing an important role in project management in the architectural, engineering and construction (AEC) industry [3]. In addition, the BIM models provide a reliable basis for setting up a digital twin (DT) because they can integrate information ranging from geometric changes in the building layout to the occupancy and use of rooms and spaces. DT for buildings can be seen as BIM models extended to capture real-world data and feed it back into the model, thus neatly closing the information loop of asset lifecycle management. Standing at the point of view of how a site foreman capture the realistic world back to the virtual digital mockup, we, human beings, locate based on the visual landmarks that consists of different semantic information. These semantic information are also stored in BIM model, and through the synchronization of the BIM visualizations and the smart device camera view, a position relative to the BIM model can be obtained. Therefore, the main objective of the study presented in this paper is to develop a building information positioning system (BIPS) that enables navigation of the indoor environment using smart device in-built cameras and connecting the BIM model to facilitate the digital twin operation.

The proposed BIPS attempts to make full use of the smartphone camera by utilizing the objects that are widely seen and continuously distributed in the indoor environment for positioning. The proposed method offers several major advantages over the existing radio-based positioning methods.

- Firstly, the proposed BIPS eliminates the need for infrastructure (beacons, wires, etc.) installed in the indoor environment.
- Secondly, the BIM model is post-processed to generate 2D semantics, hence the image matching of the BIM model is in the 2D domain. Eliminating the need for the depth estimation from smartphone camera images.
- Thirdly, the 2D domain comparison is performed in the equirectangular frame, greatly reducing computational load in comparison to the perspective frame.
- Lastly, the proposed BIPS provide absolute positioning and orientation within the building to initialize and correct the incremental positioning and orientation of the VO.

This work is supported by PolyU RISUD on the project – BBWK "Resilient Urban PNT Infrastructure to Support Safety of UAV Remote Sensing in Urban Region."

M.J.L. Lee, H.Y Ho, and L-T. Hsu are with the Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University (PolyU). L-T. Hsu is also with Research Institute for Sustainable Urban Development (RISUD). Corresponding author: Li-Ta Hsu (e-mail: lt.hsu@polyu.edu.hk).

978-1-6654-0402-0/21/$31.00 ©2021 IEEE

II. The Proposed BIPS

The flowchart of the proposed BIPS (Building Information Positioning System) is shown in **Error! Reference source not found.**.

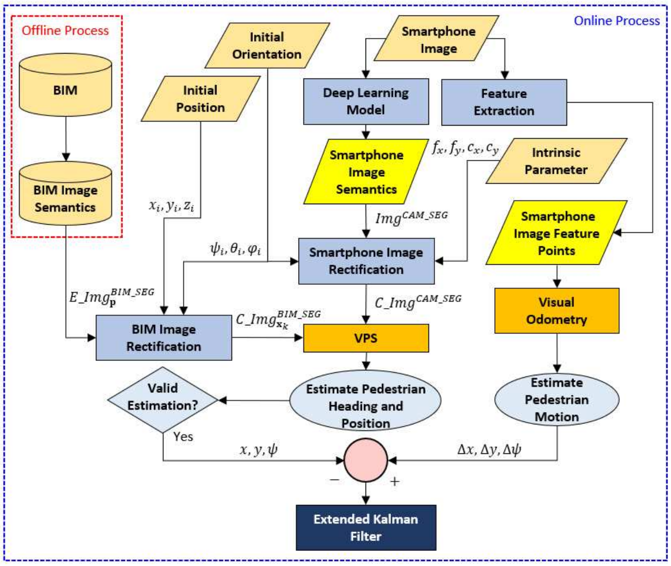

Fig. 1. Overview of the proposed BIPS.

The algorithm can be divided into the online and offline processes. The offline stage includes processing the BIM model to extract 2D semantics at each position (Sub-Section 3.1). At the online stage, the real-time smartphone images are processed using a deep learning model to extract semantics (Sub-Section 3.2), and a feature detection model to extract feature points in the smartphone image. The intrinsic parameter and initial orientation of the smartphone camera are used to rectify the smartphone image (Sub-Section 4.2). Then, based on an initial position, the hypothesized positioning candidates can be distributed (Sub-Section 4.1). Each candidate position stores the pre-computed 2D semantics of its environmental surrounding in the equirectangular frame. Using the initial orientation, the candidate images are rectified to be compared with the rectified smartphone image using the VPS method (Sub-Section 4.3).

After the image processing, the BIPS is composed of two principal parts: 1) VPS [4, 5] is used to compare the pre-computed BIM images with the smartphone image, which then estimates an absolute 2D position and heading for the EKF. The VPS position and heading are checked if they are valid or invalid by calculating the likelihood of the candidate. If valid position and heading are available, they are applied as the input to correct the visual odometry (VO) prediction with EKF. 2) VO [6] is used to produce an incremental motion estimate. The predicted incremental 2D position and heading are used as the input of the EKF.

The paper is organized as follow: Section 3 introduces the semantic extraction for BIM model and smartphone images. Section 4 presents the VPS method. Section 5 presents the VO method. Section 6 details the BIPS which is the VPS and VO integration using EKF. Section 7 tests the proposed method with data obtained by a simulation. Finally, conclusions and future perspectives are presented in Section 8 and 9, respectively.

III. BIM Model and Smartphone Semantics Extraction

3.1 BIM semantic extraction

The BIM model represents the use of computer-generated model to simulate the construction and operation of a facility, from which views and data (semantics) can be extracted and analysed to generate information. A BIM model carries all information related to the building, including its physical characteristics, in a series of “smart objects” [7]. For example, a door within a BIM model would also contain a label “door”, which can then be used to generate semantics. As many indoor semantic segmentation deep learning networks are trained using the NYU v2 dataset, this research uses the NYU v2 40 classes to classify all objects in the BIM model [8]. As shown in Fig. 2, the semantics of the BIM model are extracted as a 2D equirectangular image at each position.

$$\begin{gathered}\mathbf{p} = [x, y, z] \\ \mathbf{P} = \{\mathbf{p}_0 \cdots \mathbf{p}_s\} \\ E\_Img_{\mathbf{p}}^{BIM\_SEG} = SI(\mathbf{u}, \mathbf{v})\end{gathered} \tag{1}$$

Where $\mathbf{p}$ is a three-dimensional position, and the subscript s is the index of $\mathbf{P}$, which are all the positions inside the BIM model. Position $\mathbf{p}$ is extracted from database $\mathbf{P}$, where $\mathbf{p} \in \mathbf{P}$. SI is the function that assigns each pixel $(\mathbf{u}, \mathbf{v})$ an indexed number to represent a class. A segmented equirectangular image for a position is denoted as $E\_Img_{\mathbf{p}}^{BIM\_SEG}$. The equirectangular images are stored in a database for the online comparison.

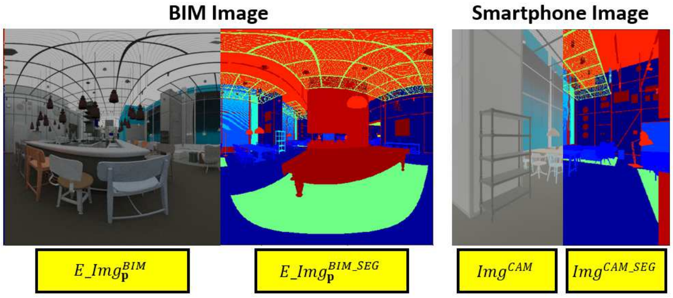

Fig. 2. Semantic extraction from BIM model and smartphone images.

3.2 Smartphone image semantic extraction

The data synthesis using the BIM model can be used to train a deep learning neural network for real-time semantic segmentation to classify the classes in the smartphone images [9]. In this paper, however, we assumed a perfect segmentation to test the feasibility of the proposed BIPS as shown in Fig. 2.

$$Img^{CAM\_SEG} = SS(Img^{CAM\_RAW}) \tag{2}$$

Where SS is the semantic segmentation, and Img^{CAM_SEG} is the segmented smartphone image. From here on, the equirectangular and smartphone images also denotes the semantics contained in the images.

IV. Visual Positioning System

4.1 BIM candidate generation

In the online process, candidate positions are spread across an initial guess position in a 40-meter radius with a 1-meter separation. An assumption is made that the ground truth position is within 40-meter from the initial position, this initial guess can be estimated using BLE, Wi-Fi and/or GNSS measurements in the indoor environment. Each candidate position has one corresponding equirectangular semantic image pre-computed from BIM model and stored in a database as shown in Fig. 3. Candidate position $\mathbf{p}_k$ is extracted from the database $\mathbf{P}$, and the subscript k is the index of the candidate positions. An equirectangular image for a candidate position is denoted as $E\_Img_{\mathbf{p}_k}^{BIM\_SEG}$.

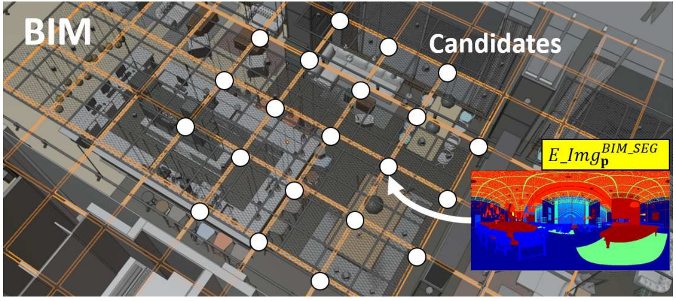

Fig. 3. Candidates with 1m separation in the BIM model.

4.2 Smartphone image rectification

To compare the smartphone image with the equirectangular image, the former is rectified from perspective to the equirectangular frame. This transformation requires the use of the smartphone camera intrinsic parameters and initial orientation. The intrinsic parameters can be computed through image calibration with a checkerboard [10], whereas the initial orientation can be obtained from IMU measurements.

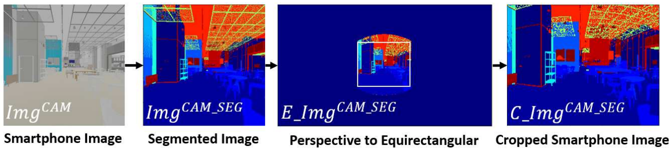

Fig. 4. Smartphone image rectification.

The transformed projection is neither equal area nor conformal and can result in different shapes. Therefore, the projected image is cropped to a rectangle (white outline) to achieve consistency in the comparison stage as shown in Fig. 4.

$$\begin{aligned} E\_Img^{CAM\_SEG} &= P2E(Img^{CAM\_SEG}) \\ C\_Img^{CAM\_SEG} &= E2C(E\_Img^{CAM\_SEG}) \end{aligned} \tag{3}$$

Where $P2E$ is the perspective to equirectangular transformation, and $E2C$ is the function to crop the transformed region to a rectangle. $C\_Img^{CAM\_SEG}$ is the cropped equirectangular smartphone image.

4.3 Candidate image rectification

Since each BIM equirectangular image stores information about the entire surrounding of the environment at a position, the equirectangular image is reduced to a window the same size of the cropped smartphone image as shown in Fig. 5.

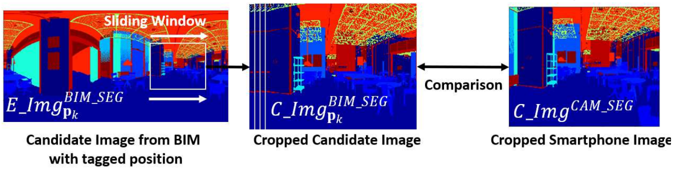
Sliding Window
$E\_Img_{\mathbf{p}_k}^{BIM\_SEG}$
$C\_Img_{\mathbf{p}_k}^{BIM\_SEG}$
Comparison
$C\_Img^{CAM\_SEG}$
Candidate Image from BIM with tagged position
Cropped Candidate Image
Cropped Smartphone Image

Fig. 5. Comparison of the candidate images with the smartphone image.

The center pixel position of the window is calculated based on the orientation estimated by the initial orientation. This orientation can be obtained from the smartphone's IMU sensor. The pitch angle is used to calculate the center pixel height position, whereas the heading angle is used to calculate the center pixel width position. Once positioned, the candidate headings are spread across the initial heading, and a sliding window algorithm is used to slide over the equirectangular to capture different portions (at different heading angles) for comparison with the smartphone image.

$$\begin{gathered} \mathbf{r} = [\varphi, \theta, \psi] \\ \mathbf{x} = \{\mathbf{p}, \mathbf{r}\} \\ C\_Img_{\mathbf{x}_k}^{BIM\_SEG} = E2C\left(E\_Img_{\mathbf{p}_k}^{BIM\_SEG}, \mathbf{r}_k\right) \end{gathered} \tag{4}$$

Candidate headings are spread across 90-degrees around the initial guess heading with 5-degree separation. An assumption is made that the ground truth heading is within 90-degrees from the initial heading.

4.4 Smartphone and candidate image comparison

The "frequency weight intersection over union" (FWIoU) score is used to estimate the heading and position as shown in Fig. 6.

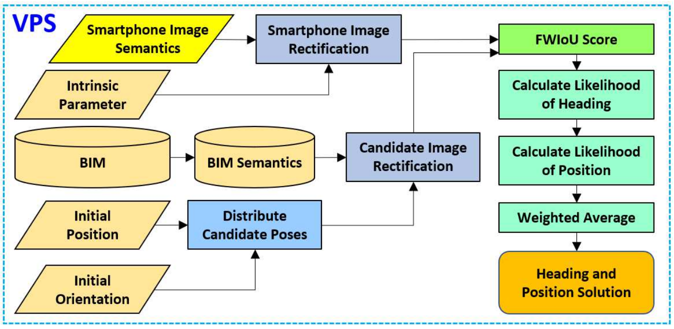
VPS
Smartphone Image Semantics
Smartphone Image Rectification
Intrinsic Parameter
BIM
BIM Semantics
Candidate Image Rectification
Initial Position
Distribute Candidate Poses
Initial Orientation
FWIoU Score
Calculate Likelihood of Heading
Calculate Likelihood of Position
Weighted Average
Heading and Position Solution

Fig. 6. Flowchart of the VPS

FWIoU is an improved version over the raw "Mean IoU" which weights each class importance depending on their appearance frequency [11]. We remark the following notation details: we assume a total of $n + 1$ classes and p_{ij} is the number of pixels of class i inferred to belong to class j. In other words, p_{ii} represents the number of true positives, while p_{ij} and p_{ji} are interpreted as false positives and false negatives.

$$\begin{aligned} &FWIoU \\ &= \frac{1}{\sum_{i=0}^{n}\sum_{j=0}^{n} p_{ij}} \sum_{i=0}^{n} \frac{p_{ii}}{\sum_{i=0}^{n} p_{ij} + \sum_{j=0}^{n} p_{ij} - p_{ii}} \\ &score(\mathbf{x}_k) \\ &= FWIoU(C\_Img_{\mathbf{x}_k}^{BIM\_SEG}, C\_Img^{CAM\_SEG}) \end{aligned} \tag{5}$$

$FWIoU$ score of 0 indicates full match, and 1 is no match. We considered the $FWIoU$ score of 600 smartphone images when compared to their BIM image counterpart at the same position, orientation and intrinsic parameters to calibrate two CDFs shown in Fig. 7 and Table I. In theory, the semantics should match entirely, and the similarity score should be 0. However, it is likely some pixel noise exist that will reduce the similarity score. Based on the CDF, the score of each candidate is used to calculate their likelihood.

978-1-6654-0402-0/21/$31.00 ©2021 IEEE

$$\alpha(\mathbf{x}_k) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{score(\mathbf{x}_k)} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \tag{6}$$

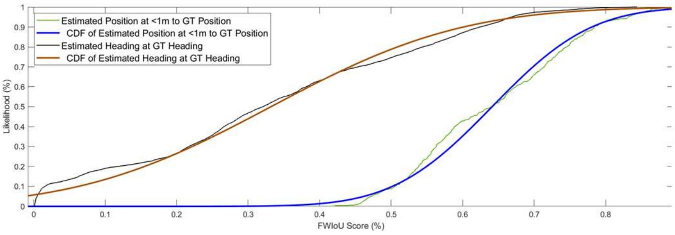

Fig. 7. CDF of FWIoU Score.

TABLE Ⅰ. Parameters of the likelihood modelled by the Gaussian distribution.

Method	Standard Deviation (m)	Mean (m)
Estimated Position at <1m to ground truth Position	0.109	0.641
Estimated Heading at ground truth Heading	0.210	0.332

We first assume that the semantics at the same heading does not differ significantly at a candidate position that is close to the ground truth position. Then, the FWIoU score of each heading at each candidate position is used to calculate a likelihood of the heading based on the heading CDF. For the candidate k, a weighted average is used to estimate its heading.

$$\hat{\psi}_k = \frac{\sum_{\psi} \alpha^{\psi}(\mathbf{x}_k)\psi(\mathbf{x}_k)}{\sum_{\psi} \alpha^{\psi}(\mathbf{x}_k)} \tag{7}$$

The estimated heading is then used to calculate the likelihood of the position of each candidate based on the positioning CDF. In order to exclude the anomalous candidate, this study heuristically defines a constant threshold, C_{FWIoU}, which is adjusted to 0.4.

$$\alpha_k^{\mathbf{p}}(\mathbf{p}_k) = \begin{cases} if\ score(\mathbf{p}_k) > C_{FWIoU} & \text{(Valid)} \\ 0 & \text{(Invalid)} \end{cases} \tag{8}$$

Finally, the positioning result of the proposed BIPS is calculated by:

$$\hat{\mathbf{p}} = \frac{\sum_k \alpha^k(\mathbf{x}_k)\mathbf{p}(\mathbf{x}_k)}{\sum_k \alpha^k(\mathbf{x}_k)} \tag{9}$$

V. Visual Odometry

The state-of-the-art ORB_SLAM3 is used to estimate the 2D position velocity and heading orientation velocity of the camera [6]. The keyframes selected from the tracking thread was used to compute a real-time trajectory and sparse 3D reconstruction of the scene. In addition, its loop closure is used to correct its pose drift. When there are enough features, the motion-only bundle adjustment (BA) optimizes the position in this map to minimize the error. Please find the detail in [6]. As shown in Fig 8, the VO demonstrates the key frames, feature points and the trajectory of the camera frames with 10 Hz frequency.

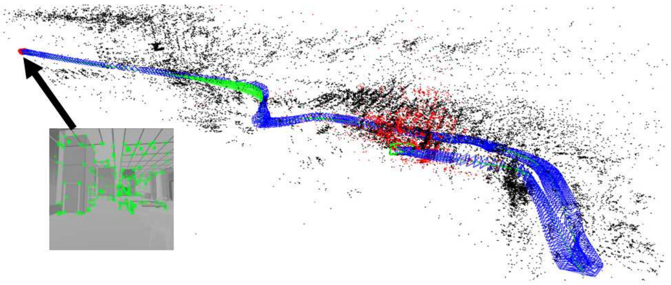

Fig. 8. Demonstration of the ORB SLAM 3 visual odometry using the simulated images from the BIM.

VI. Extended Kalman Filter

An extended Kalman filter is used to integrate VPS and VO. The state vector comprises of 2D position and heading of the smartphone. To calibrate the measurement noise covariance matrix of the VPS, we need to understand the VPS measurement uncertainty using the only available FWIoU similarity score. Therefore, we observed the correlation of the positioning and heading error in relation to the score (within the valid score) as shown in Fig. 9.

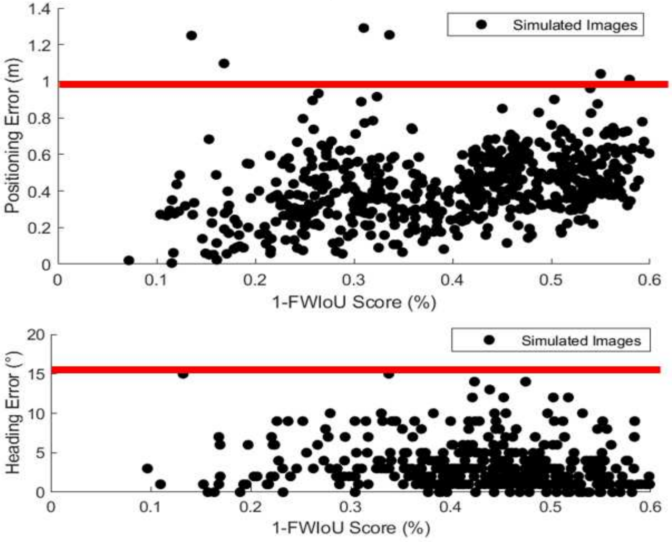

Fig 9. Monte Carlo simulation of positioning and heading error with various level of FWIoU score.

As shown in Fig 9, there are several outliers with more than 1m positioning error. The outliers were identified to understand its discrepancies. This poor result can be explained by the lack of distinctive semantics (when the entire image is dominated by one or two classes). Therefore, we have analyzed the number of classes in relation to the positioning error in Sect. VII. The observed maximum positioning error of 1m and heading error of 15° will be used as the measurement noise covariance matrix:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 15 \end{bmatrix} \tag{10}$$

The two 1m and one 15° are the x position, y position and heading uncertainty of the VPS, respectively. If there are no valid candidates, then there will be no measurement, which will directly increase the covariance matrix of the next estimated state, thus increasing the search radius of the VPS.

VII. Experimental Results

5.1 Simulation setting

In this study, the experimental trajectory was simulated within a hotel BIM model in Blender as shown in Fig. 10 [12].

Fig. 10. Hotel BIM Model.

The hotel is a conventional indoor environment in a highly developed city area which contains numerous objects. A major benefit of using the proposed BIPS is to make use of these objects. The captured virtual camera images along a simulated trajectory were used to test the feasibility of the proposed BIPS. The ground truth positions of the trajectory were recorded, and the positioning quality of the proposed method was analyzed based on the ideal segmentation of the images [13]. The point positioning VPS has a frequency of 0.5Hz. Whereas the VO has a frequency of 10Hz.

5.2 BIPS positioning and heading results

The positioning results are plotted onto a bird's-eye view of the hotel in Fig. 11. There are three performance metrics used: mean, standard deviation of the 2D positioning error and the availability of the positioning solution. Availability means the number of solutions in a fix period. For example, if the VPS outputs 48 epochs in a 100s period, the VPS availability is 48%. If the VO outputs 1000 epochs in a 100s period, the VO availability is 1000%.

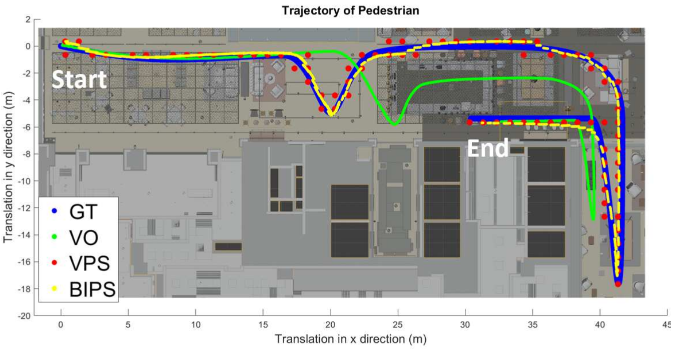

Fig. 11. Positioning results of the proposed BIPS and other visual positioning systems.

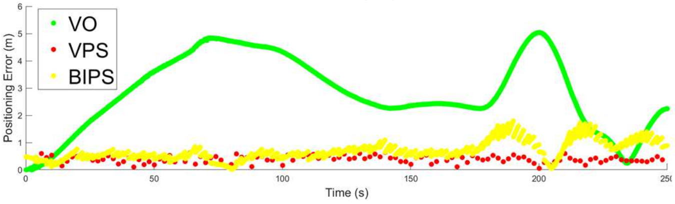

Fig. 12. Positioning error against time of the proposed BIPS and other visual positioning systems.

The results shown that the point positioning error of the VPS (red marker) is on average below 0.5m to the ground truth (blue marker) as analyzed in Fig. 12 and Table III. This is mainly due to the abundance of semantics in the BIM model, which provides the matching of more than 5 classes. The 0.5m accuracy also signifies that the candidate closest to the ground truth is often chosen. Therefore, it can be expected that by increasing the resolution of the candidates, the positioning accuracy will increase. Nonetheless, the VPS is inhibited by its low availability. The monocular VO (green marker) begins with great positioning accuracy but starts to accumulate drift error as it progresses as shown during camera rotations. This has led to a significant cumulative positioning mean error of 2.8m and heading mean error of 45°. The proposed BIPS (yellow marker) is to integrate the VPS and VO via EKF. Consequently, the VPS can correct the positioning and heading error of the VO, whereas the VO can increase the availability of the positioning system. The result of the proposed BIPS provides an accurate and continuous positioning of 0.7m mean error, and 10° heading error. Although the BIPS error has increased slightly from the VPS, it has still achieved meter level accuracy enough for navigation, with availability increasing to 10 Hz.

Table III. Accuracy of BIPS in the presence of few classes

Method	2D Position		Heading		Avail.
	Mean (m)	SD (m)	Mean (°)	SD (°)	$(\frac{Solution\ Frame}{Time\ (sec)})$
BIPS	0.69	0.34	9.87	32.07	1000/100
VPS	0.42	0.18	3.48	3.18	48/100
VO	2.83	1.35	44.87	42.23	1000/100

5.3 Number of classes vs positioning accuracy

We then experimented with the VPS for three distinctive cases, where semantics can be categorized as: 1) numerous, 2) limited, to 3) insufficient. Fig. 13 shows typical cases of position results in different scenarios. If the smartphone image contains numerous semantics, it can be matched to the candidate closest to the ground truth position within 1m accuracy. Where adjacent candidates will have decreasing level of likelihood. In the case 2, if the smartphone image contains limited semantics, there will be more valid candidates, as more candidates share the same semantics. Thus, it can then be weighted to consistently produce a positioning result of 1-2m accuracy. In the case 3, when there are only one semantic, the positioning result will be poor, and the positioning result of the proposed BIPS decreases significantly.

Table II notes the positioning accuracy of the following cases. In the ideal case, the more classes, the closer the estimated position is to the ground truth, given that there are valid candidates. Hence, we measured the positioning accuracy at various number of classes. It is shown that having 5+ classes will significantly improve the positioning results to within 0-1m. This can also be used as a threshold to improve the robustness of the BIPS in the future.

978-1-6654-0402-0/21/$31.00 ©2021 IEEE

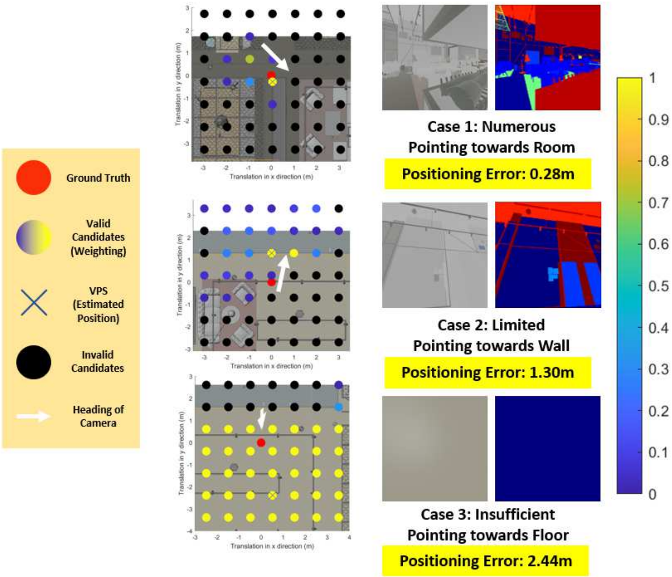

Fig. 13. Positioning candidates of the proposed BIPS in the case of numerous, limited, and insufficient semantics.

Table II. Accuracy of the proposed BIPS in the condition of different number of classes

Number of Classes (Each class occupy at least 10% of image)	Range accuracy in meters	Valid candidate in percentage (10m radius)
1-2	2-10+	50-100
3-4	1-2	10-50
5+	0-1	<10

VIII. Conclusions

This paper proposes a camera and BIM-based indoor positioning solution for heading and 2D position estimation by introducing semantics as a source of information. In short, the semantic information is extracted from the smartphone images and compared to the BIM model. The measured heading and point positioning are then combined with the incremental pose of the visual odometry to provide a seamless BIPS positioning solution within 1m positioning accuracy according to the simulation result. The potential advantages of the proposed BIPS method are:

- The use of BIM model for positioning eliminates the need for radio frequency positioning infrastructure.
- The formulation of positioning as a semantic-based problem enables us to apply existing wide variety of advanced matching metrics to this problem.
- Objects are diverse, distinctive, and distributed everywhere, hence the semantic information in an image is easy to recognize.

In this paper, synthetic data was used to test the proposed BIPS, however, in the real world there would be additional challenges that needs to be addressed including:

- Real-time semantic segmentation to classify the classes in the smartphone images.
- Detection and exclusion of dynamic objects to prevent false measurements.
- Up-to-date BIM model that reflects the building.

Considering the preliminary results presented in this paper, we believe the proposed BIPS can provide accurate positioning and heading estimation to support various indoor applications.

IX. Future Works

We will conduct a real-time experiment in Hong Kong to validate the proposed BIPS. We will integrate the state-of-the-art deep learning model to classify the semantics for the BIPS. Several potential future developments on the propose BIPS are suggested.

- To maximize all available visual information, the VPS can make use of the features identified from the VO to match with the BIM model to yield better positioning performance.
- The VPS and visual odometry can also be integrated with the IMU for better positioning performance and availability.
- The batch integration such as, factor graph optimization which is proven to outperform EKF will be a great candidate to improve the proposed BIPS.

References

[1] A. Hameed and H. A. Ahmed, "Survey on indoor positioning applications based on different technologies," in *2018 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, 24-25 Nov. 2018 2018, pp. 1-5, doi: 10.1109/MACS.2018.8628462.

[2] V. Renaudin *et al.*, "Evaluating Indoor Positioning Systems in a Shopping Mall: The Lessons Learned From the IPIN 2018 Competition," *IEEE Access,* vol. 7, pp. 148594-148628, 2019, doi: 10.1109/ACCESS.2019.2944389.

[3] G. Xiaodong, H. Jiwei, L. Siyu, L. Jianhua, and D. Mingyi, "Indoor localization method of intelligent mobile terminal based on BIM," in *2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, 22-23 March 2018 2018, pp. 1-9, doi: 10.1109/UPINLBS.2018.8559731.

[4] M. Jwo Lem Lee and L.-T. Hsu, "Semantic 3D Map Change Detection and Update based on Smartphone Visual Positioning System," *arXiv e-prints,* p. arXiv: 2103.11311, 2021.

[5] M. J. L. Lee and L. T. Hsu, "A feasibility study on smartphone localization using image registration with segmented 3d building models based on multi-material classes," presented at the ION ITM 2021, 2021.

[6] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM," 2020. [Online]. Available: https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.1109%2FTRO.2021.3075644&v=a8cc9408.

[7] X. Xu, "Research and Application of Operation and Maintenance 5D-BIM Data Integration and Sharing System Based on IFC," in *2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, 17-19 April 2020 2020, pp. 4-7, doi: 10.1109/CIBDA50819.2020.00009.

978-1-6654-0402-0/21/$31.00 ©2021 IEEE

[8] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, *Indoor Segmentation and Support Inference from RGBD Images*. 2012, pp. 746-760.

[9] M. Alawadhi and W. Yan, "BIM Hyperreality: Data Synthesis Using BIM and Hyperrealistic Rendering for Deep Learning," 2021. [Online]. Available: https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.1109%2FTRO.2021.3075644&v=a8cc9408.

[10] Matlab. "Single Camera Calibrator App." The MathWorks Inc. https://www.mathworks.com/help/vision/ug/single-camera-calibrator-app.html (accessed.

[11] I. Ulku and E. Akagunduz, "A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images," 2019. [Online]. Available: https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.1109%2FTRO.2021.3075644&v=a8cc9408.

[12] B. O. Community. "Blender - a 3D modelling and rendering package." Blender Foundation. http://www.blender.org (accessed.

[13] M. Denninger *et al.*, "BlenderProc," 2019. [Online]. Available: https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.1109%2FTRO.2021.3075644&v=a8cc9408.

Max Jwo Lem Lee is currently a graduate of Bachelor of Engineering (Honours) in Aviation Engineering from the Hong Kong Polytechnic University. He has previously interned in Boeing and Cathay Pacific as an engineer for 1 year and will be starting his PhD study in September 2021 with the Interdisciplinary Division of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University. He has been awarded the Hong Kong PhD Fellowship Scheme 2021/2022 and have won the Hong Kong Techathon 2021. His other research interests include positioning in urban environments, indoor positioning, and unmanned aerial vehicle.

Hiu Yi Ho is currently an undergraduate student of Bachelor of Engineering (Honours) in Air Transport Engineering from the Hong Kong Polytechnic University. She will be starting her MPhil study in September 2021 with the Department of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University. Her research interests include visual odometry in indoor and outdoor environments, and inertial navigation system for unmanned autonomous systems.

Li-Ta Hsu (S’09-M’15) received the B.S. and Ph.D. degrees in aeronautics and astronautics from National Cheng Kung University, Taiwan, in 2007 and 2013, respectively. He is currently an assistant professor with the Division of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University, before he served as post-doctoral researcher in Institute of Industrial Science at University of Tokyo, Japan. In 2012, he was a visiting scholar in University College London, U.K. He is an Associate Fellow of RIN. His research interests include GNSS positioning in challenging environments and localization for pedestrian, autonomous driving vehicle and unmanned aerial vehicle.

Stephen Ling Ming Au received Higher Diploma of Applied Science from Hong Kong Polytechnic University in 1982, MBA of Strategic Marketing from University of HULL , UK in 1999 and Master of Advance Business Practices from University of South Australia in 2007 respectively. He is the Managing Director and founder of MTECH ENGINEERING CO.,LTD since 1995 specific on PLM, BIM and Digital Construction. He got the award of China 2007 Top 100 Innovative Enterprise Leader in 2008 and the Outstanding PolyU Alumni Award in 2017.

978-1-6654-0402-0/21/$31.00 ©2021 IEEE