RESEARCH ON INDOOR POSITION BASED ON AI IMAGE RECOGNITION AND PROTOGRAMMETRY

Chin-Huang Hsu [1], Tien-Yin Chou *[2], Mei-Ling Yeh *[3]

[1] Graduate Student , College of Construction and Development, Feng Chia University
No. 100, Wenhua Road, Xitun District, Taichung City (3rd Floor, Qiu Fengjia Memorial Hall)
Email: p1068757@o365.fcu.edu.tw

[2] Distinguished Research Chair Professor, Department of Urban Planning and Spatial Information,
Feng Chia University
Room 413-1, 4th Floor, Zhongqin Building, No. 100 Wenhua Road, Xitun District, Taichung City 407102
Email: jimmy@gis.tw

[3] Assistant Professor, College of Construction and Development, Feng Chia University
No. 100, Wenhua Road, Xitun District, Taichung City (3rd Floor, Qiu Fengjia Memorial Hall)
Email: milly@gis.tw

Keywords : Indoor Positioning System (IPS), AI image recognition, Photogrammetry, 3D Lidar Point cloud

ABSTRACT: Smart Navigation Technology, with the development of 3D spatial information, the Indoor Positioning System (IPS) is a hot research topic. The convenient and accurate Global Positioning System-GNSS can be used for outdoor positioning, but in indoor spaces, GNSS wireless signals are affected by building barriers and cannot be used for indoor positioning. Over the years, there have been many methods for IPS positioning, such as using wireless local area network (WLAN) radio wave fingerprints for positioning, or using RFID tags and deploying Locator Nodes devices for positioning or using Bluetooth Low Energy (BLE) beacons, Ultra -Wideband (UWB) and other positioning technologies, but the above-mentioned related positioning technologies require a huge hardware cost. The positioning accuracy is also limited by the transparency of positioning signal transmission and the ability to resist interference.

In recent years, AI image recognition technology has developed rapidly, and indoor public areas are based on security and management considerations, and all indoor spaces are monitored by surveillance camera equipment at any time. AI-based image recognition can quickly identify image objects, especially for human body posture recognition. After various learning calculations, the accuracy is greatly improved. This research intends to use the surveillance camera equipment deployed indoors, the human body posture recognition technology of AI images, and the technology of photogrammetry and geographic information positioning as an IPS method. The purpose is to identify people or target objects in the indoor space, not only for the positioning of geospatial location information but also to record their moving trajectory and direction. This research method is different from the technology of wireless positioning. It does not need to increase the cost of indoor positioning hardware, but it can also improve the positioning accuracy to 0.5m within or higher.

1. Instructions

The Global Navigation Satellite System (GNSS) is composed of a group of positioning satellites deployed in space, such as the Global Positioning System (GPS) developed by the United States, Europe's Galileo satellite navigation system), GLONASS (Global Navigation Satellite System) developed by Russia, and BeiDou Navigation Satellite System (BeiDou Navigation Satellite System) developed by China. Global positioning technology has been developed for over 30 years and has gradually matured, especially in transportation systems. , intelligent positioning and navigation systems have become an indispensable tools.

Intelligent navigation technology develops with the development of three-dimensional space information, among which indoor positioning system (IPS) is a hot research topic. Because GNSS wireless signals are affected by building obstructions, they cannot be used for positioning indoors. Over the years, IPS positioning technology and principles have been developed in many ways, such as using wireless network architectures such as Wifi, Bluetooth, Zigbee, etc., or positioning technology using RFID tags and deploying Locator Nodes, or positioning methods such as Ultra-wideband (UWB). Its positioning signal processing operation model includes signal fingerprinting (Fingerprinting) sensing type, calculation of the time difference of the signal (called Time of Arrival; TOA), RSSI (Received Signal Strength Indicator) calculation of the signal strength indication, and differential calculation using time, which is called It is TDoA (Time Difference of Arrival). Although each has its advantages and disadvantages (F. Zafari et al., 2019), the expensive construction cost and the anti-interference ability of wireless signal transmission that affect positioning accuracy still need to be improved.

2. Research materials and methods

To improve the efficiency and accuracy of indoor positioning, UWB devices are used and imitated GPS positioning mode to improve indoor positioning accuracy (Luca et al. 2021); in dynamic 2D images, real-time methods can be implemented to detect the 2D postures of multiple people in the image, regardless of the image how many people in the room can achieve a combined body and foot keypoint detector, including joint points of the body, feet, hands and face (Zhe et al., 2018); Omsri 2017 is based on automated images to conduct indoor dynamics Positioning of the vehicle.

Comprehensive research related to indoor positioning, there are various approaches to the discussion of indoor sensing, communication transmission, and positioning accuracy. This study is based on the fact that surveillance camera equipment has been deployed in indoor public areas. In recent years, AI image recognition technology has can identify the posture of the human body and the classification of various joints. In addition to making the best judgment and image positioning for the position-related foot joints, this research also combines the principles of photogrammetry, indoor LiDAR point cloud modeling of the experimental site and Geographic coordinate positioning as the basic core architecture of IPS. This method not only does not increase the cost of proprietary indoor positioning hardware, but also improves indoor positioning accuracy. The coordinate system of its positioning information will also be consistent with the world coordinate system. For consistency, the indoor positioning system can be integrated with the outdoor global satellite positioning system to reduce the construction cost of indoor positioning facilities and avoid the problems of indoor positioning signal shielding and interference.

(1) Research design and process

The main design methods and process steps of this study are shown in Figure 1:

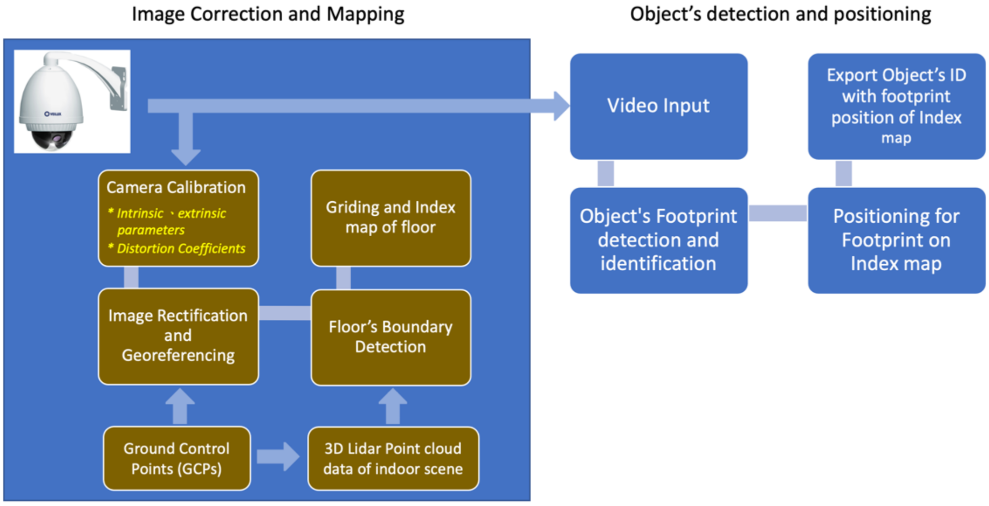


Image Correction and Mapping
Object's detection and positioning
Camera Calibration
* Intrinsic 、extrinsic parameters
* Distortion Coefficients
Griding and Index map of floor
Image Rectification and Georeferencing
Floor's Boundary Detection
Ground Control Points (GCPs)
3D Lidar Point cloud data of indoor scene
Video Input
Export Object's ID with footprint position of Index map
Object's Footprint detection and identification
Positioning for Footprint on Index map

Figure 1

- Camera Calibration: The purpose of camera calibration is to obtain the camera's parameters, including Intrinsic Matrix, Extrinsic Matrix, Lens Distortion, etc.

- Georeferencing: Use camera Intrinsic and Extrinsic parameters and ground control points (GCP) for image rectification and spatial georeferencing.

- Floor's Boundary Detection: After geolocating the 3D Lidar point cloud data scanned in the indoor scene and fusing it with the image, the floor image data has depth information, and the image is segmented into pixels belonging to the floor range.

- Griding and Index map of floor: Grid segment the range image belonging to the floor and create an index map. The center point of each grid has geographical coordinate information.

- Object's Footprint detection and identification: identification, detection and encoding processing of the footprints of the image target object

- Calculation for Footprint on Index map: Calculate the index map position where the target's footprint is located.

- Output Object's ID with Index map ID: Output the target object ID and the index map ID of the location of the footprint.

(2) Indoor positioning experimental area

(a) Camera location selection: This experimental area is the indoor area of the teaching building as the measurement area. The indoor positioning camera is a commercial model used for general indoor surveillance photography. It has a wide-angle fixed focus lens and a shooting range covering indoor entrances and exits, the camera images are shown in Figure 2. Two cameras were used at the experimental site. The overlap rate of the images taken by the two cameras was 45 degrees. The other captured images occupied more than 60% of the entire image with the floor space. , the photography angle is based on human body recognition.

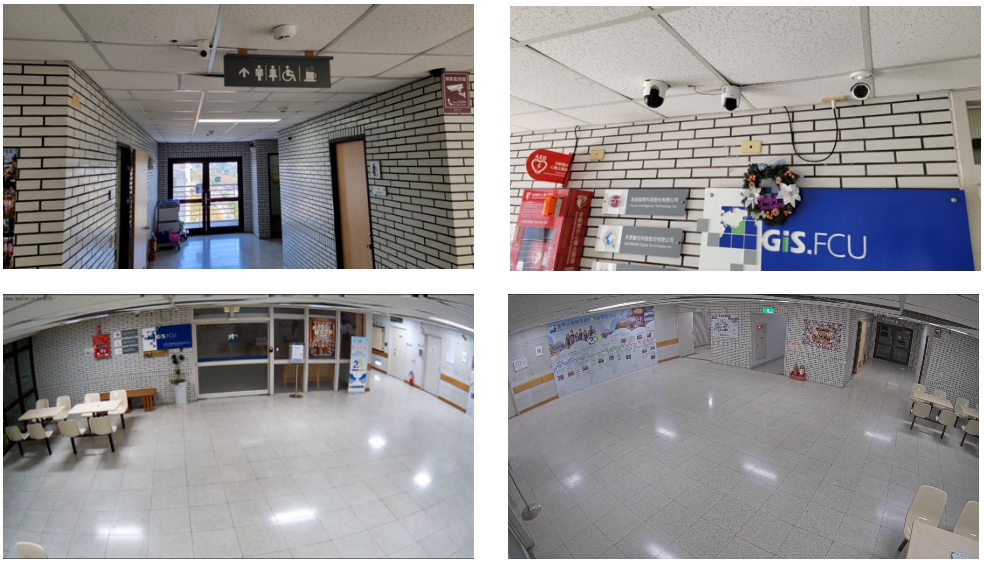
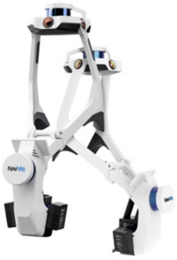


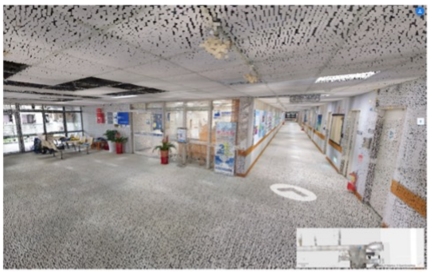
Figure 2 (Cam-1 location and view) (Cam-2 location and view)

(b) 3D LiDAR Point Cloud Modeling and Geospatial Coordinate Localization in Experimental Field: Utilizing the Navvis VLX 3D LiDAR scanner as shown in Figure 3.1, to create a 3D model of the experimental field, obtaining a comprehensive and detailed 3D point cloud and panoramic images (Figure 3.2 and Figure 3.3). In addition to serving as a means for calibrating and validating the extrinsic parameters of the cameras, these precise spatial data also serve as a tool for error checking in the outcomes of this research.



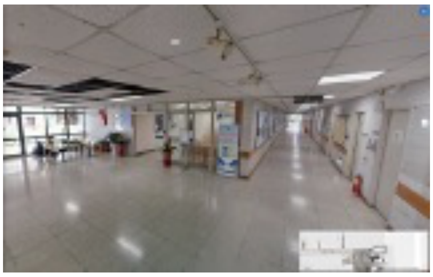
Navvis VLX

Figure 3.1



3D point cloud

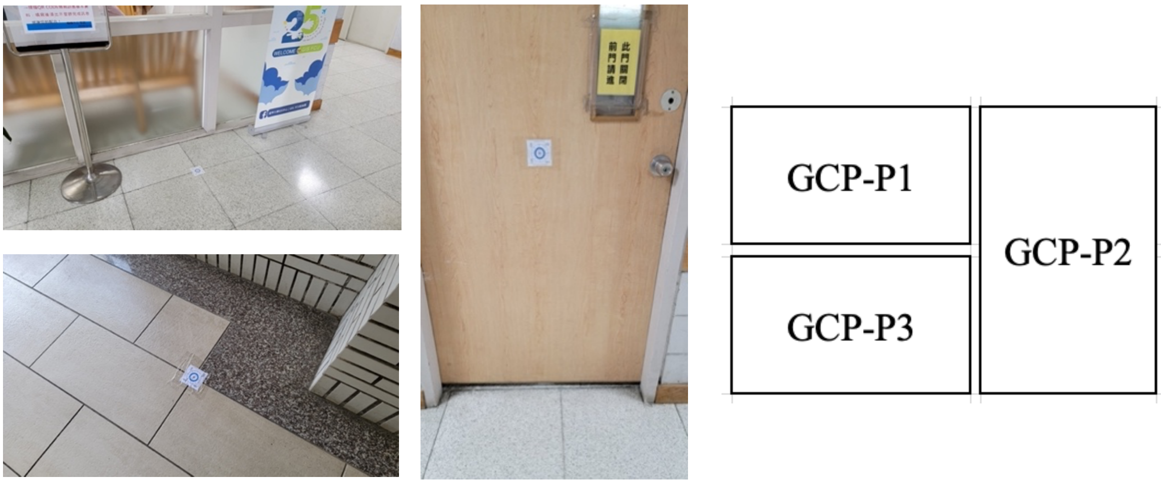
Figure 3.2



Panoramic images

Figure3.3

(c) Ground Control Point (GCP) Deployment: To establish the spatial information coordinate system for the experimental area that employs GPS receivers and theodolites at known control points outdoors near the experimented field, it allows to performance of precise coordinate measurements on the three ground control points within the experimental area, obtaining accurate ground coordinates as depicted in Figure 4.



Coordinate system	TWD-97			WGS84		
GCP	E	N	Height	Longitude	Latitude	Height
P1	214433.708	2674816.926	96.51	120.649942669655	24.1783257778351	96.51
P2	214409.991	2674820.441	97.56	120.649709154461	24.1783569794081	97.56
P3	214386.904	2674819.696	96.496	120.649481944836	24.1783497304924	96.496

Figure 4

(d) Pinhole camera calibration (Intrinsic and Extrinsic parameters)

■ Camera Intrinsic Matrix (Camera-to-Image, Image-to-Pixel): Converts points from the camera 3D coordinate system to the 2D pixel coordinate system.

■ Camera Extrinsic Matrix (World-to-Camera): Converts 3D points from the world coordinate system to the camera 3D coordinate system.

(I) Pinhole camera model :

$$sm' = A[\mathrm{R}|\mathrm{t}]\mathrm{M}' \quad \text{or} \quad s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r32 & r33 & t3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Where:

- (X,Y,Z) are the coordinates of a 3D point in the world coordinate space
- (u,v) are the coordinates of the projection point in pixels
- A is a matrix of intrinsic parameters
- $(c_x, c_y)$is a principal point that is usually at the image center
- f_x,f_y are the focal lengths expressed in pixel units.
- [R|t] is called a matrix of extrinsic parameters
- R matrix (ω,ϕ κ) :

  $r11=\cos\phi\ \cos\kappa$

  $r12=-\cos\phi\ \sin\kappa$

  $r13=\sin\phi$

  $r21=\cos\omega\ \sin\kappa+\sin\omega\sin\phi\cos\kappa$

  $r22=\cos\omega\ \cos\kappa-\sin\omega\sin\phi\sin\kappa$

  $r23=-\sin\omega\cos\phi$

  $r31=\sin\omega\sin\kappa-\cos\omega\sin\phi\ \cos\kappa$

  $r32=\sin\omega\cos\kappa+\cos\omega\sin\phi\ \sin\kappa$

  $r33=\cos\omega\ \cos\phi$

The transformation above is equivalent as below (when z≠ 0):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$

$$x' = x/z$$

$$y' = y/z$$

$$u = f_x * x' + c_x$$

$$v = fy * y' + cy$$

(II) Due to the lenses have some radial and tangential distortion, the above model is extended as below:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$

$$x' = x/z$$

$$y' = y/z$$

$$x^{"} = x'(1+k_1r^2+k_2r^4+k_3r^6)/(1+k_4r^2+k_5r^4+k_6r^6)+2p_1x'y'+p_2(r^2+2x'^2)$$

$$x^{"} = y'(1+k_1r^2+k_2r^4+k_3r^6)/(1+k_4r^2+k_5r^4+k_6r^6)+p_1(r^2+2y'^2)+2p_2x'y'$$

$$where \quad r^2 = x'^2 + y'^2$$

$$u = f_x * x^{"} + c_x$$

$$v = f_y * y^{"} + c_y$$

- k_1, k_2, k_3, k_4, k_5, k_6 are radial distortion coefficients
- p_1 and p_2 are tangential distortion coefficients

(e) Utilizing the camera calibration module of OpenCV and directly compute the radial distortion coefficients (K1, K2, K3) and tangential distortion coefficients (P1, P2) of the camera lens by processing images of a checkboard and 3D objects.

(f) more than 6 Ground Control Points need to be arranged on the ground of the experimental site, and the GCP positions of the GCPs should be evenly distributed within the floor area captured by the camera image (Figure 5). OpenCV can use these GCPs to obtain the extrinsic parameters of the camera image. , including 3 rotation angles (Omega-w, Phi-f, Kappa-k), and 3 translation offsets (tx, ty, tz) along the X, Y, and Z axes.

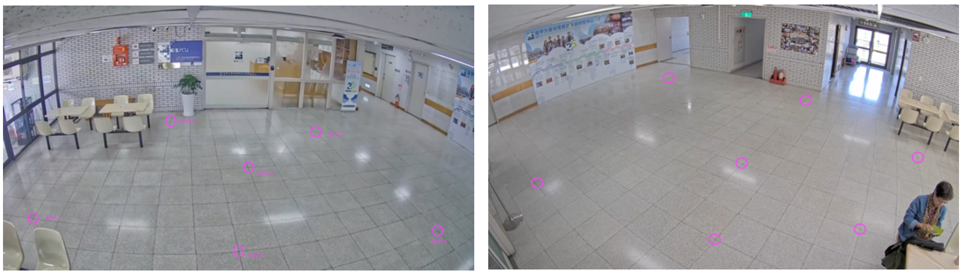


Figure 5 (Cam-1 GCP's location) (Cam-2 GCP's location)

(g) After merging the geospatially positioned 3D LiDAR point cloud data with the properly rectified images, the floor image data is enriched with depth information. Subsequently, through the application of deep learning techniques that segmented the pixels corresponding to floor area and divided the floor image region into a grid of 50cm x 50cm squares oriented in the north-south direction, using this gridded layout to construct an indexed map, referred to as the "Index map."

(h) Pose estimation is a computer vision technique used to detect people or objects' position or movement trajectory in images or videos. Different methods and theories are developed for pose estimation depending on the specific object of study and the intended purpose, in this research that primarily utilize Openpose as our primary method. Openpose is particularly suitable for real-time, multi-person scenarios, as it can identify the location of the heels of human bodies. This study, it uses real-time footage from surveillance cameras for indoor localization, making Openpose a suitable choice. The image below shows the detection of human body heels using Openpose in the experimental field image (Figure 6). Calculating the position of the human body's footprint, obtaining the Index Map's ID and time associated with the footprint location.

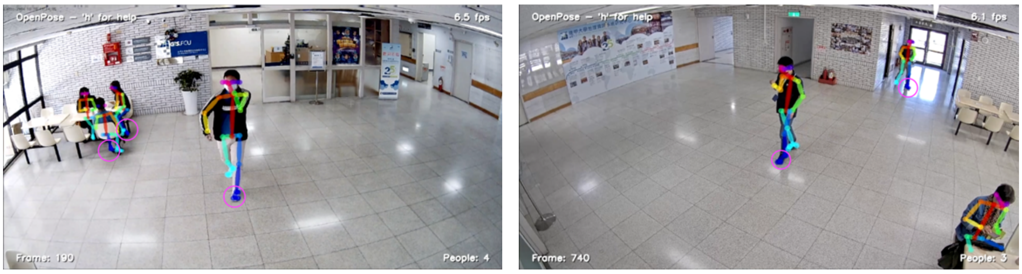


Figure 6 Detection of footprint (Cam-1 and Cam-2)

3. Initial results and conclusions of the research

This research is focused on the development of an indoor positioning system that utilizes image measurements, 3D LiDAR point clouds, coordinate transformations, and AI image-based positioning methods. The primary objectives are to reduce the hardware costs associated with indoor positioning systems and enhance indoor positioning accuracy. Additionally, the study explores and analyzes the integration of deep learning-based human pose recognition as the primary method for indoor positioning of mobile targets (humans).

This preliminary report outlines the framework and concept of the research methodology. Subsequent phases of the research will involve a comparative analysis of floor area detection methods and validation of indoor positioning accuracy.

References from Other Literature:

1. F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568-2599, third quarter 2019, doi: 10.1109/COMST.2019.2911558.

2. L. Santoro, M. Nardello, D. Brunelli and D. Fontanelli, "Scale up to infinity: the UWB Indoor Global Positioning System," *2021 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, FL, USA, 2021, pp. 1-8, doi: 10.1109/ROSE52750.2021.9611770.

3. Omsri K.A., 2017 Automatic Image-Based Positioning, Master Thesis, Department of Applied Signal Processing Blekinge Institute of Technology, Sweden, P6-P8

4. Cao, Z., Hidalgo, G., Simon, T., Wei, S., & Sheikh, Y. (2018). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. ArXiv. /abs/1812.08008

5. Z. Zhang, "A flexible new technique for camera calibration," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, Nov. 2000, doi: 10.1109/34.888718.

6. Sergey S., A., Oleninilov, M.F. Demirci, S. Almas, 2020, Deep Learning-Based Object Classification and Position Estimation Pipeline for Potential Use in Robotized Pick-and-Place Operation, Robotics, 9, 63.

7. C. Chun, D. Park, W. Kim and C. Kim, "Floor detection based depth estimation from a single indoor scene," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 2013, pp. 3358-3362, doi: 10.1109/ICIP.2013.6738692.