Single and Double Reference Points Based High Precision 3D Indoor Positioning with Camera and Orientation-Sensor on Smart Phone

Article *in* Wireless Personal Communications · August 2015
DOI: 10.1007/s11277-015-2499-7

Honggui Li
Yangzhou University

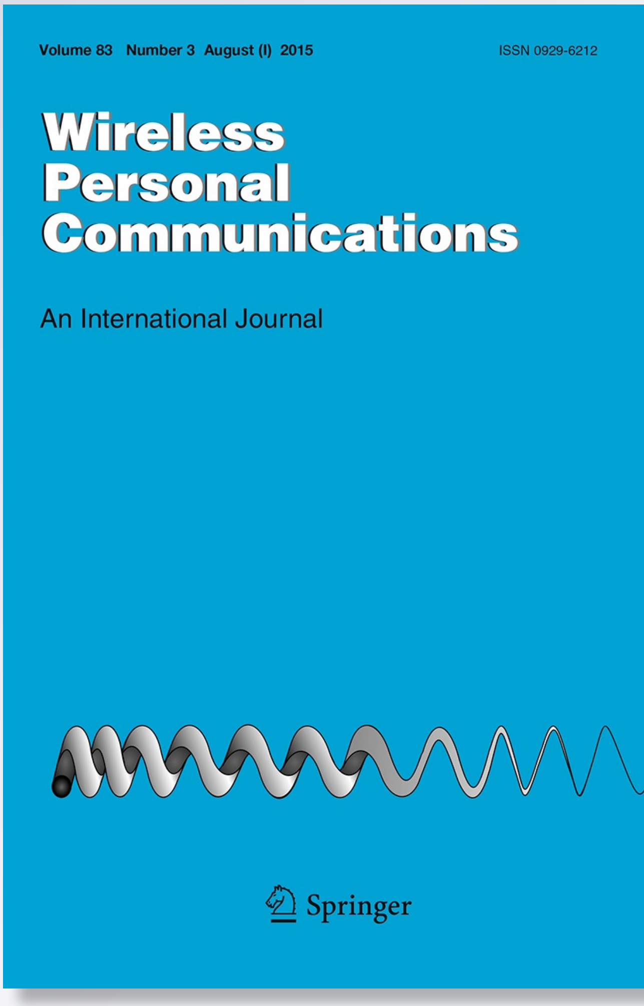
Single and Double Reference Points Based High Precision 3D Indoor Positioning with Camera and Orientation-Sensor on Smart Phone

Honggui Li

Wireless Personal Communications
An International Journal

ISSN 0929-6212
Volume 83
Number 3

Wireless Pers Commun (2015)
83:1995-2011
DOI 10.1007/s11277-015-2499-7



Springer

Wireless Pers Commun (2015) 83:1995–2011
DOI 10.1007/s11277-015-2499-7

Single and Double Reference Points Based High Precision 3D Indoor Positioning with Camera and Orientation-Sensor on Smart Phone

Honggui Li[1]

Published online: 10 March 2015

Abstract This paper proposes a high precision 3D indoor positioning method, which depends on one or two reference points in scene and camera and orientation-sensor on smart phone. The proposed method has the merits of implementing simplicity, high accuracy, friendly availability and low cost. The proposed method adopts only single or double reference points and camera imaging model for camera calibrating and 3D positioning. Classical camera calibration method uses a set of reference points, but the proposed camera calibration method is easier to implement. Traditional computer vision based positioning method, P3P, utilizes at least three points and has multiple solutions. The proposed positioning method is equal to P2P or P1P, but is faster and has unique solution. The proposed method achieves the high 3D positioning precision at the level less than 30 cm. The proposed method employs smart phone as hardware and software carrier, which is more usable than other methods. In a word, the proposed method is very suitable for low-cost and high precision 3D indoor positioning.

Keywords 3D indoor positioning · Orientation-sensor · P3P · Smart phone

1 Introduction

Smart phone has become the necessary device for human daily life, and its importance will be more and more tremendous. People use smart phone to explore the city, the country and all of the world, fetch the latest news, watch the current films, buy diverse commodities, accomplish stock transactions, and etc. The functions of smart phone have evolved more and more powerfully. It relies on the rapid improvement of hardware and software. Smart phone's hardware includes powerful processor and lots of useful sensors, such as accelerator sensor, gravity sensor, gyroscope sensor, light sensor, geomagnetic field sensor, proximity sensor, air pressure

✉ Honggui Li
hgli@yzu.edu.cn

1 Physics College of Science and Technology, Yangzhou University, Yangzhou 225000, Jiangsu, China

sensor, temperature sensor, and etc. In fact, camera itself on smart phone is also a kind of image sensor. Smart phone's software controls the work flow of the whole system and provides some software sensors, which are established on the hardware sensors. For example, smartphone operating system (SOS) provides software sensors, such as orientation sensor and rotation vector sensor. Orientation sensor is built on the hardware gravity sensor and the hardware geomagnetic field sensor, outputs the orientation of smart phone, and can be used for indoor positioning. Air pressure and temperature sensors can be utilized to decide the elevation, which is useful for indoor locating. The mean deviation of barometer, which is made up of air pressure and temperature sensors, can run up to 0.1 m with an uncertainty (95 %) of ± 0.85 m [33]. SOS application programming interface (API) can also tell us the focal length of camera on smart phone, which is valuable for indoor positioning.

Classical indoor positioning methods [1–3] include, radio signal based methods [4], computer vision based methods [5], sound and ultra-sound signals based methods, infrared signal based methods [6], ultra wide band (UWB) signal based methods [7], visible light communication (VLC) based methods [8], magnetic field based methods, and etc.

The theory of computer vision is complex and complete. With the swift development of hardware and software, it is possible to simplify the complicated theory of computer vision for realistic applications. For example, high performance embedded processor and various hardware sensors can be adopted to decrease the computation load of computer vision algorithms. Computer vision based indoor positioning methods use different visual features, such as point, corner, edge, line, marker, track, block, grid, color, range, ceiling, face, person, motion, optic flow, image sequences, images of different views, images of videos, images of stereo vision, omnidirectional images, panoramic images, and etc. Computer vision based indoor positioning systems usually utilize inertial measurement units (IMUs), which are various sensors including camera.

Gritton [9] beliefs that, micro-electro-mechanic system (MEMS) sensors will enhance next-gen content discovery, navigation and interactivity in consumer electronics (CE) devices. Atzori [10] makes use of the accelerometer and the compass of the smart phone to estimate the amount of steps and the direction, and employs speeded up robust features (SURF) algorithm to quickly and effectively detect key points in camera image. Wang [11] fuses the absolute positioning results from camera sensor networks and the relative positioning results from the smart phone-based dead reckoning (DR) method with reference points. Taylor [12] applies image feature points for smart phone based indoor guidance system. Ruotsalainen [13] utilizes digital compass and camera on smart phone, and vanishing points calculated from lines in consecutive images, for heading change detection. Werner [14] combines an image recognition system with a distance estimation algorithm, to gain a high-quality positioning service using the camera of a smart phone. Juang [15] calculates the relative position of the target object through image processing and distance computing algorithms. Ascher [16] uses two low-cost IMUs, an electronic compass and an altimeter on a smart phone, and an additional laser ranger for indoor navigation. Van Vinh [17] employs panoramic images and camera and compass on mobile phone for indoor navigation. Zhang [18] utilizes a modified Kalman filter based smart phone sensor data fusion method to achieve accurate orientation data, by detecting and minimizing the effect of magnetic field disturbance. Faragher [19] uses a distributed particle filter simultaneous localization and mapping (DPSLAM) method, to provide constraints on the drift of a simple hip-mounted IMU integrated into the smart phone and offer the core information on the movement of the user. Khalifa [20] describes the work in the progress of converting context to indoor positioning using built-in smart phone sensors, such as accelerometer and gyroscope. Kang [21] presents improved heading estimation solution for indoor positioning using smart phone equipped with sensors, such as accelerometer, magnetometer, and gyroscope. Pratama [22] focuses on displacement estimation by utilizing the accelerometer sensor integrated on a smart phone. Lukianto [23] uses inertial navigation systems (INSs) of smart phone, which are based on the integration of linear accelerations and angular velocities. Ausmeier [24] utilizes the accelerometer on a mobile phone to track users' walking for indoor navigation.

Classical computer vision based positioning method is P3P (the Perspective 3-Points Problem) [25, 26]. P3P has multiple solutions. In order to obtain sole solution, at least four non-coplanar points are needed to uniquely determine the 3D position of mobile objects [27]. Oementhon [28] develops an iterative optimizing algorithm, named POSIT, for pose estimation using a minimum of four reference points. Qingxuan [29] introduces a method of real-time and high-precision positioning by single camera based on P3P algorithm. D'Alfonso [30] uses the image provided by the camera and the gravity vector offered by the IMUs, to estimate the relative pose of the object through solving a modified P2P or P3P problem. It shows that the modified P2P problem always gives two solutions, while the modified P3P problem usually gives a single solution.

A new algorithm, which utilizes only one or two reference points, is proposed. The proposed method depends on the camera imaging model and the camera and orientation-sensor on smart phone.

Interest point detection is very important in the proposed method. Aanæs [31] investigates the performance of a number of existing well-established interest point detection methods, and proves that the fixed scale Harris corner detector performs overall best followed by the Hessian based detectors and the Difference of Gaussian (DoG). Gauglitz [32] proposes a comprehensive quantitative evaluation of detector-descriptor-based visual camera tracking, based on a carefully designed dataset of video sequences of planar textures with ground truth.

How to construct reference points in scene is another critical problem in the proposed method. The general choose is high power lighting LEDs, and it will also be adopted in the proposed method.

The rest part of this paper is arranged as following. Section 2 depicts the theory foundation of indoor positioning. Section 3 discusses the experimental results of indoor positioning. Section 4 is the conclusions and future work.

2 Theory

2.1 P3P

Traditional computer vision based positioning method is P3P. Figure 1 shows the diagram of P3P method. $\boldsymbol{O}_w$-x_w-y_w-z_w is the world coordinate system, and $\boldsymbol{O}_c$-x_c-y_c-z_c is the camera coordinate system.

The theory of P3P can be described as following formulas.

$$\begin{aligned}
\boldsymbol{P}_i &= k_i\boldsymbol{Q}_i, \quad i = 1,2,3 \\
\boldsymbol{P}_i &= [x_{pi} \quad y_{pi} \quad z_{pi}]^T, \boldsymbol{Q}_i = [x_{qi} \quad y_{qi} \quad -f]^T \\
d_{mn}^2 &= \|\boldsymbol{P}_m - \boldsymbol{P}_n\|_2^2 = \|k_m\boldsymbol{Q}_m - k_n\boldsymbol{Q}_n\|_2^2 \\
&= k_m^2(\boldsymbol{Q}_m \cdot \boldsymbol{Q}_m) - 2k_mk_n(\boldsymbol{Q}_m \cdot \boldsymbol{Q}_n) + k_n^2(\boldsymbol{Q}_n \cdot \boldsymbol{Q}_n) \\
&= k_m^2Q_{mm} - 2k_mk_nQ_{mn} + k_n^2Q_{nn} \quad m,n = 1,2,3, m \neq n
\end{aligned} \tag{1}$$

where $\boldsymbol{P}_i$ is the point in the camera coordinate system, and its coordinates are to be resolved, $\boldsymbol{Q}_i$ is the related point in the camera image plane, and its coordinates in camera coordinate system are known, k_i is the factor, which is to be resolved, f is the focal length of camera, which is known, d_{mn} is the distance between $\boldsymbol{P}_m$ and $\boldsymbol{P}_n$, and it is known, Q_{mn} is the dot product of $\boldsymbol{Q}_m$ and $\boldsymbol{Q}_n$.

If k_{i} is solved, $\boldsymbol{P}_i$ can be deduced from k_i and $\boldsymbol{Q}_i$, and the representation of $\boldsymbol{P}_i$ in the world coordinate system can also be obtained, according to the relationship between the world coordinate system and the camera coordinate system. Formula (1) can be resolved by iterated optimization method and has multiple solutions. Usually four non-coplanar points are used for obtaining the unique solution.

k_{i} can be resolved by iterating. The iterative increment Δ_i of k_i can be represented by following formulas.

$$\boldsymbol{K} = [\, k_1 \quad k_2 \quad k_3 \,]^T, \quad \boldsymbol{\Delta} = [\, \Delta_1 \quad \Delta_2 \quad \Delta_3 \,]^T$$

$$\begin{cases} f_{12}(\boldsymbol{K}) = k_1^2 Q_{11} - 2k_1k_2Q_{12} + k_2^2 Q_{22} - d_{12}^2 \\ f_{23}(\boldsymbol{K}) = k_2^2 Q_{22} - 2k_2k_3Q_{23} + k_3^2 Q_{33} - d_{23}^2 \\ f_{31}(\boldsymbol{K}) = k_3^2 Q_{33} - 2k_3k_1Q_{31} + k_1^2 Q_{11} - d_{31}^2 \end{cases}$$

$$\begin{cases} f_{12}(\boldsymbol{K}+\boldsymbol{\Delta}) \approx f_{12}(\boldsymbol{K}) + \begin{bmatrix} \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_3} \end{bmatrix} \boldsymbol{\Delta} = 0 \\ f_{23}(\boldsymbol{K}+\boldsymbol{\Delta}) \approx f_{23}(\boldsymbol{K}) + \begin{bmatrix} \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_3} \end{bmatrix} \boldsymbol{\Delta} = 0 \\ f_{31}(\boldsymbol{K}+\boldsymbol{\Delta}) \approx f_{31}(\boldsymbol{K}) + \begin{bmatrix} \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_3} \end{bmatrix} \boldsymbol{\Delta} = 0 \end{cases}$$

$$[f_{12}(\boldsymbol{K}) \quad f_{23}(\boldsymbol{K}) \quad f_{31}(\boldsymbol{K})]^T + \begin{bmatrix} \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_3} \\ \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_3} \\ \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_3} \end{bmatrix} \boldsymbol{\Delta} = 0 \tag{2}$$

$$\boldsymbol{F}(\boldsymbol{K}) + \boldsymbol{J}(\boldsymbol{K})\boldsymbol{\Delta} = 0$$

$$\begin{cases} \boldsymbol{F}(\boldsymbol{K}) = [f_{12}(\boldsymbol{K}) \quad f_{23}(\boldsymbol{K}) \quad f_{31}(\boldsymbol{K})]^T \\ \boldsymbol{J}(\boldsymbol{K}) = \begin{bmatrix} \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{12}(\boldsymbol{K})}{\partial k_3} \\ \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{23}(\boldsymbol{K})}{\partial k_3} \\ \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_1} & \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_2} & \dfrac{\partial f_{31}(\boldsymbol{K})}{\partial k_3} \end{bmatrix} \\ = \begin{bmatrix} 2k_1Q_{11} - 2k_2Q_{12} & 2k_2Q_{22} - 2k_1Q_{12} & 0 \\ 0 & 2k_2Q_{22} - 2k_3Q_{23} & 2k_3Q_{33} - 2k_2Q_{23} \\ 2k_1Q_{11} - 2k_3Q_{31} & 0 & 2k_3Q_{33} - 2k_1Q_{31} \end{bmatrix} \\ \boldsymbol{\Delta} = -\boldsymbol{J}^{-1}(\boldsymbol{K})\boldsymbol{F}(\boldsymbol{K}) \end{cases}$$

$\boldsymbol{K}$ can be found by following iterative formula.

$$\boldsymbol{K}^{l+1} = \boldsymbol{K}^l + \boldsymbol{\Delta}^l = \boldsymbol{K}^l - \boldsymbol{J}^{-1}\left(\boldsymbol{K}^l\right)\boldsymbol{F}\left(\boldsymbol{K}^l\right), \; \left|\boldsymbol{K}^{l+1} - \boldsymbol{K}^l\right| < \varepsilon \tag{3}$$

where l is the number of iteration, ε is a small constant, which is used for the terminate condition of iteration.

2.2 Camera Imaging Model

Figure 2 illustrates the traditional camera imaging model. The model focuses on the relationship between the world coordinate system and the camera coordinate system.

Classical camera imaging model can be described as following formulas. It can be assumed that camera image plane overlaps digital image coordinate plane, and both of them have the same central point. The skew of the axes of camera image plane can also be omitted.

$$
\begin{aligned}
&\boldsymbol{W} = [x_w \quad y_w \quad z_w \quad 1]^T, \boldsymbol{C} = [x_c \quad y_c \quad -f \quad 1]^T \\
&\boldsymbol{T} = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{R}_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&\boldsymbol{R}_y = \begin{bmatrix} \cos\beta & 0 & -\sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{R}_z = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 & 0 \\ -\sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&\boldsymbol{S} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{M} = \boldsymbol{S}\boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{T}\boldsymbol{W} = [x_m \quad y_m \quad z_m \quad 1]^T \\
&\boldsymbol{C} = \boldsymbol{S}P\left(\boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{T}\boldsymbol{W}\right) = P\left(\boldsymbol{S}\boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{T}\boldsymbol{W}\right) = P(\boldsymbol{M}) = \begin{bmatrix} -\frac{fx_m}{z_m} & -\frac{fy_m}{z_m} & -f & 1 \end{bmatrix}^T
\end{aligned}
\tag{4}
$$

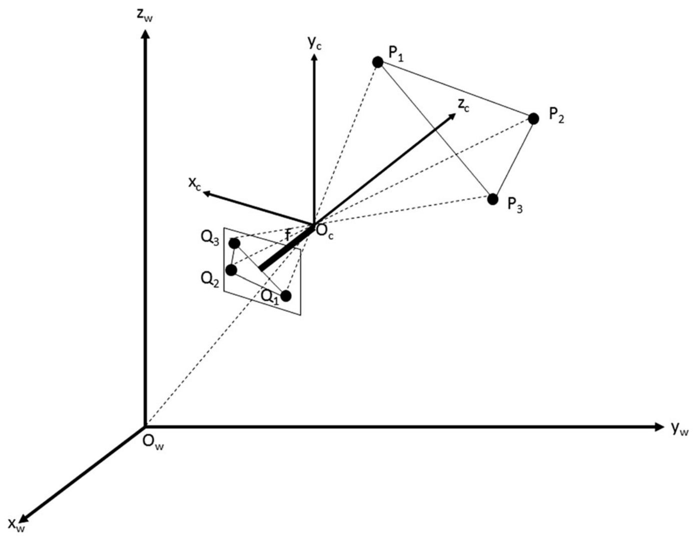


Fig. 1 The diagram of P3P method

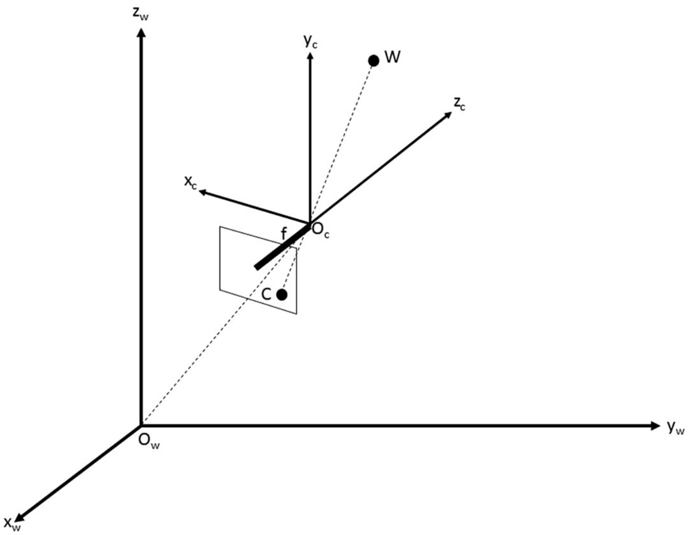


Fig. 2 Camera imaging model

where $\boldsymbol{W}$ is the homogeneous coordinates of one point in the world coordinate system, $\boldsymbol{C}$ is the homogeneous coordinates of the same point in the camera coordinate system, and only x_c and y_c are valuable for 2D image, $\boldsymbol{T}$ is the translation transformation matrix, which moves the original point $\boldsymbol{O}_w$ of the world coordinate system to the position of the original point $\boldsymbol{O}_c$ of the camera coordinate system, $\boldsymbol{R}_x$ is the rotation matrix which rotates the camera coordinate system around x_w axis with anticlockwise angle α, $\boldsymbol{R}_y$ is the rotation matrix which rotates the camera coordinate system around y_w axis with anticlockwise angle β, $\boldsymbol{R}_z$ is the rotation matrix which rotates the camera coordinate system around z_w axis with anticlockwise angle γ, α, β and γ can be measured by the orientation-sensor on the smart phone, P is the perspective transformation from the world coordinate system to image plane in camera coordinate system, f is the focal length of camera, and can be measure by SOS API of smart phone, $\boldsymbol{S}$ is the scale transformation matrix, which converts the coordinates from image plane to digital image coordinate plane, and $s_z = 1$ for 2D digital image.

For the convenience of computing, following formulas are given. The independent estimation of focal length f can be canceled for camera calibration. For the purpose of positioning based on the proposed method, only the inertial camera parameters, s_{xf} and s_{yf}, are needed to be calibrated.

$$\boldsymbol{S}_f = \boldsymbol{S}\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} s_x f & 0 & 0 & 0 \\ 0 & s_y f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} s_{xf} & 0 & 0 & 0 \\ 0 & s_{yf} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

$$\boldsymbol{M}_f = \boldsymbol{S}_f \boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \boldsymbol{T}\boldsymbol{W} = \begin{bmatrix} x_{mf} & y_{mf} & z_{mf} & 1 \end{bmatrix}^T$$

$$\boldsymbol{C} = \boldsymbol{S}_f P_f\left(\boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \boldsymbol{T}\boldsymbol{W}\right) = P_f\left(\boldsymbol{M}_f\right) = \begin{bmatrix} -\frac{x_{mf}}{z_{mf}} & -\frac{y_{mf}}{z_{mf}} & -f & 1 \end{bmatrix}^T$$

2.3 Camera Calibration with Single Reference Point

The two inertial camera parameters, s_{xf} and s_{yf}, in formulas (5) can be estimated with one reference point. One reference point $\boldsymbol{W}$, and related known information, such as $\boldsymbol{C}$, $\boldsymbol{T}$, $\boldsymbol{R}_x$, $\boldsymbol{R}_y$, and $\boldsymbol{R}_z$, can determine the two camera parameters. Following formulas depict the estimation method.

$$\begin{aligned}
&\boldsymbol{A} = \boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{TW} = \begin{bmatrix} x_a & y_a & z_a & 1 \end{bmatrix}^T \\
&\boldsymbol{C} = \begin{bmatrix} x_c & y_c & -f & 1 \end{bmatrix}^T = \boldsymbol{S}_f P_f\left(\boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{TW}\right) = \boldsymbol{S}_f P_f(A) \\
&\begin{cases} s_{xf} = -\dfrac{z_a x_c}{x_a} \\[2ex] s_{yf} = -\dfrac{z_a y_c}{y_a} \end{cases}
\end{aligned} \tag{6}$$

2.4 Camera Calibration with Double Reference Points

The two inertial camera parameters, s_{xf} and s_{yf}, in formulas (5) can be estimated with two reference points. Two reference points $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, and related known information, such as $\boldsymbol{C}_1$, $\boldsymbol{C}_2$, $\boldsymbol{T}$, $\boldsymbol{R}_x$, $\boldsymbol{R}_y$, and $\boldsymbol{R}_z$, can decide the two camera parameters. Following formulas describe the estimation method.

$$\begin{aligned}
&\boldsymbol{A}_1 = \boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{TW}_1 = \begin{bmatrix} x_{a1} & y_{a1} & z_{a1} & 1 \end{bmatrix}^T \\
&\boldsymbol{A}_2 = \boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{TW}_2 = \begin{bmatrix} x_{a2} & y_{a2} & z_{a2} & 1 \end{bmatrix}^T \\
&\boldsymbol{C}_1 = \begin{bmatrix} x_{c1} & y_{c1} & -f & 1 \end{bmatrix}^T = \boldsymbol{S}_f P_f\left(\boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{TW}_1\right) = \boldsymbol{S}_f P_f(\boldsymbol{A}_1) \\
&\boldsymbol{C}_2 = \begin{bmatrix} x_{c2} & y_{c2} & -f & 1 \end{bmatrix}^T = \boldsymbol{S}_f P_f\left(\boldsymbol{R}'_x\boldsymbol{R}'_y\boldsymbol{R}'_z\boldsymbol{TW}_2\right) = \boldsymbol{S}_f P_f(\boldsymbol{A}_2) \\
&\begin{cases} s_{xf} = -\dfrac{1}{2}\left(\dfrac{z_{a1} x_{c1}}{x_{a1}} + \dfrac{z_{a1} x_{c1}}{x_{a1}}\right) \\[2ex] s_{yf} = -\dfrac{1}{2}\left(\dfrac{z_{a2} y_{c2}}{y_{a2}} + \dfrac{z_{a2} y_{c2}}{y_{a2}}\right) \end{cases}
\end{aligned} \tag{7}$$

Traditional camera calibration method uses a set of reference points. The proposed method uses only one or two reference points, and it leads to simplicity of implementation.

2.5 3D Position Estimation with Single Reference Point

According to one known reference point $\boldsymbol{W}$, and related known information, such as $\boldsymbol{C}$, $\boldsymbol{S}_f$, P_f, $\boldsymbol{R}_x$, $\boldsymbol{R}_y$ and $\boldsymbol{R}_z$, the 3D position $\boldsymbol{L}$ of the original point $\boldsymbol{O}_c$ of camera coordinate system in the world coordinate system, can be calculated as following formulas.

$$
\begin{aligned}
&\boldsymbol{B} = \boldsymbol{S}_f \boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \\
&\boldsymbol{C} = P_f\left(\boldsymbol{S}_f \boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \boldsymbol{T}\boldsymbol{W}\right) = P_f(\boldsymbol{B}\boldsymbol{T}\boldsymbol{W}) \\
&\boldsymbol{D}\boldsymbol{L} = \boldsymbol{E} \\
&\boldsymbol{D} = \begin{bmatrix} b_{11} + x_c b_{31} & b_{12} + x_c b_{32} & b_{13} + x_c b_{33} \\ b_{21} + y_c b_{31} & b_{22} + y_c b_{32} & b_{23} + y_c b_{33} \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \end{bmatrix} \\
&\boldsymbol{L} = [-t_x \quad -t_y \quad -t_z]^T, \boldsymbol{E} = \begin{bmatrix} (\boldsymbol{B}(1,:) + x_c \boldsymbol{B}(3,:))\boldsymbol{W} \\ (\boldsymbol{B}(2,:) + y_c \boldsymbol{B}(3,:))\boldsymbol{W} \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \\
&\boldsymbol{L} = \begin{bmatrix} \dfrac{(d_{12}e_2 - d_{22}e_1) - (d_{12}d_{23} - d_{13}d_{22})t_z}{d_{12}d_{21} - d_{11}d_{22}} \\ \dfrac{1}{d_{12}}\left(e_1 - d_{11}\dfrac{(d_{12}e_2 - d_{22}e_1) - (d_{12}d_{23} - d_{13}d_{22})t_z}{d_{12}d_{21} - d_{11}d_{22}} - d_{13}t_z\right) \\ t_z \end{bmatrix}
\end{aligned}
\tag{8}
$$

where $\boldsymbol{L}$ is the location of smart phone, t_z can be measured by user cooperation or by other sensors on smart phone. If user takes a photo for positioning always in the same height, which is proportion to the height of user, t_z can be obtained easily. Actually, camera usually has a constant altitude in many indoor positioning applications. Otherwise air pressure sensor and temperature sensor on smart phone can be taken to gain the t_z accurately.

3D position estimation with single reference point is equal to P1P. It is described as following formulas. If k and $\boldsymbol{P}$ are resolved, $\boldsymbol{L}$ and $\boldsymbol{T}$ can be obtained on the basis of known $\boldsymbol{W}$. On the contrary, if $\boldsymbol{L}$ and $\boldsymbol{T}$ are solved, k and $\boldsymbol{P}$ can be gained from known $\boldsymbol{W}$ and $\boldsymbol{Q}$.

$$
\begin{aligned}
&\boldsymbol{P} = k\boldsymbol{Q} \\
&\boldsymbol{P} = \left[\boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z (\boldsymbol{W} - [\boldsymbol{L}' \quad 0]')\right](1:3) = \left[\boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \boldsymbol{T}\boldsymbol{W}\right](1:3) = [x_p \quad y_p \quad z_p]^T \\
&\boldsymbol{Q} = \boldsymbol{C}(1:3) = [x_q \quad y_q \quad -f]^T
\end{aligned}
\tag{9}
$$

2.6 3D Position Estimation with Double Reference Points

On the basis of two known reference points $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, and related known information, such as $\boldsymbol{C}_1$, $\boldsymbol{C}_2$, $\boldsymbol{S}_f$, P_f, $\boldsymbol{R}_x$, $\boldsymbol{R}_y$ and $\boldsymbol{R}_z$, the 3D position $\boldsymbol{L}$ of the original point $\boldsymbol{O}_c$ of camera coordinate system in the world coordinate system, can be computed as following formulas.

$$
\begin{aligned}
&\boldsymbol{B} = \boldsymbol{S}_f \boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \\
&\boldsymbol{C}_1 = P_f\left(\boldsymbol{S}_f \boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \boldsymbol{T} \boldsymbol{W}_1\right) = P_f(\boldsymbol{BTW}_1) \\
&\boldsymbol{C}_2 = P_f\left(\boldsymbol{S}_f \boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \boldsymbol{T} \boldsymbol{W}_2\right) = P_f(\boldsymbol{BTW}_2) \\
&\boldsymbol{DL} = \boldsymbol{E}, \boldsymbol{D} = \begin{bmatrix} b_{11} + x_{c1} b_{31} & b_{12} + x_{c1} b_{32} & b_{13} + x_{c1} b_{33} \\ b_{21} + y_{c1} b_{31} & b_{22} + y_{c1} b_{32} & b_{23} + y_{c1} b_{33} \\ b_{11} + x_{c2} b_{31} & b_{12} + x_{c2} b_{32} & b_{13} + x_{c2} b_{33} \\ b_{21} + y_{c2} b_{31} & b_{22} + x_{c2} b_{32} & b_{23} + y_{c2} b_{33} \end{bmatrix} \\
&\boldsymbol{L} = [-t_x \quad -t_y \quad -t_z]^T, \boldsymbol{E} = \begin{bmatrix} (\boldsymbol{B}(1,:) + x_{c1}\boldsymbol{B}(3,:))\boldsymbol{W}_1 \\ (\boldsymbol{B}(2,:) + y_{c1}\boldsymbol{B}(3,:))\boldsymbol{W}_1 \\ (\boldsymbol{B}(1,:) + x_{c2}\boldsymbol{B}(3,:))\boldsymbol{W}_2 \\ (\boldsymbol{B}(2,:) + y_{c2}\boldsymbol{B}(3,:))\boldsymbol{W}_2 \end{bmatrix} \\
&\boldsymbol{L} = \left(\boldsymbol{D}^T \boldsymbol{D}\right)^{-1} \boldsymbol{D}^T \boldsymbol{E}
\end{aligned}
\tag{10}
$$

where $\boldsymbol{L}$ is computed by least square.

3D position estimation with double reference points is equivalent to P2P. Following formulas describe the relationship. If k_i and $\boldsymbol{P}_i$ are solved, $\boldsymbol{L}$ and $\boldsymbol{T}$ can be gained on the basis of known $\boldsymbol{W}_i$. Otherwise, if $\boldsymbol{L}$ and $\boldsymbol{T}$ are resolved, k_i and $\boldsymbol{P}_i$ can be obtained on the basis of known $\boldsymbol{W}_i$ and $\boldsymbol{Q}_i$.

$$
\begin{aligned}
&\boldsymbol{P}_i = k_i \boldsymbol{Q}_i, \; i = 1,2,3 \\
&\boldsymbol{P}_i = \left[\boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \left(\boldsymbol{W}_i - [\boldsymbol{L}' \quad 0]'\right)\right](1:3) = \left[\boldsymbol{R}'_x \boldsymbol{R}'_y \boldsymbol{R}'_z \boldsymbol{T} \boldsymbol{W}_i\right](1:3) = [x_{pi} \quad y_{pi} \quad z_{pi}]^T \\
&\boldsymbol{Q}_i = \boldsymbol{C}_i(1:3) = [x_{qi} \quad y_{qi} \quad -f]^T
\end{aligned}
\tag{11}
$$

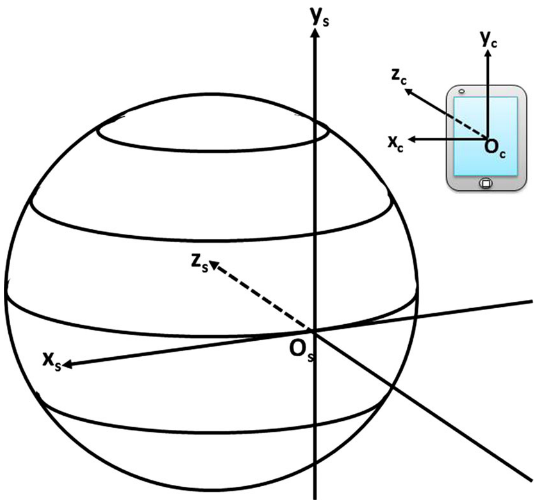
2.7 Sensor Coordinate System

Figure 3 illustrates the sensor coordinate system of smart phone, where the smart phone is in a default orientation. This sensor coordinate system influences the output of smart phone sensors. For example, the orientation sensor measures the degrees of rotation that a device makes around all three physical axes (x_s, y_s, z_s). x_s is defined as the vector product of y_s axis and z_s axis, and is tangential to the ground at the device's current location and roughly points west. y_s axis is tangential to the ground at the device's current location and points towards the magnetic North Pole. z_s axis points towards the center of the earth and is perpendicular to the ground.

2.8 Rotation Angle Computing

The rotation angle in formulas (4) can be calculated by following formula.

Fig. 3 Sensor coordinate system of smart phone



$$\boldsymbol{\theta}_r = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \boldsymbol{\theta}_c - \boldsymbol{\theta}_w = \begin{bmatrix} \theta_{cx} \\ \theta_{cy} \\ \theta_{cz} \end{bmatrix} - \begin{bmatrix} \theta_{wx} \\ \theta_{wy} \\ \theta_{wz} \end{bmatrix} = \begin{bmatrix} \theta_{cx} - \theta_{wx} \\ \theta_{cy} - \theta_{wy} \\ \theta_{cz} - \theta_{wz} \end{bmatrix} \tag{12}$$

where $\boldsymbol{\Theta}_c$ is the vector of rotation angle between related axes of the camera coordinate system and that of the sensor coordinate system, which can be obtained directly from the software orientation-sensor on smart phone, or indirectly from the hardware gravity sensor and the hardware geomagnetic field sensor, $\boldsymbol{\Theta}_w$ is the vector of rotation angle between related axes of the world coordinate system and that of the sensor coordinate system, which can be obtained from known information about world coordinate system.

Most buildings of the Northern Hemisphere are perpendicular to the ground and face South Pole. For the purpose of convenient, the axes of the world coordinate system are assumed to be parallel to the axes of sensor coordinate system. It is depicted as following formulas.

$$\boldsymbol{\theta}_r = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \boldsymbol{\theta}_c - \boldsymbol{\theta}_w = \boldsymbol{\theta}_c, \boldsymbol{\theta}_w = 0 \tag{13}$$

2.9 Reference Points Detection

Reference points can be high power lighting LEDs or special color LEDs. The methods of reference points detection in digital image plane, include classical interest point detection methods of computer vision and user cooperation (user interaction) based methods. In user cooperation based methods, user always takes a photo in which the single reference point is in the special position of camera image plane, such as the center of camera image plane or other predetermined position of camera image plane. Otherwise user can manually mark the single reference point on the screen of smart phone. In the case of double reference points, the middle point of the two reference points locates in particular position of camera image plane. Otherwise user also can manually mark the double reference points on the screen of smart phone. Figure 4 shows the examples of special reference-point positions in

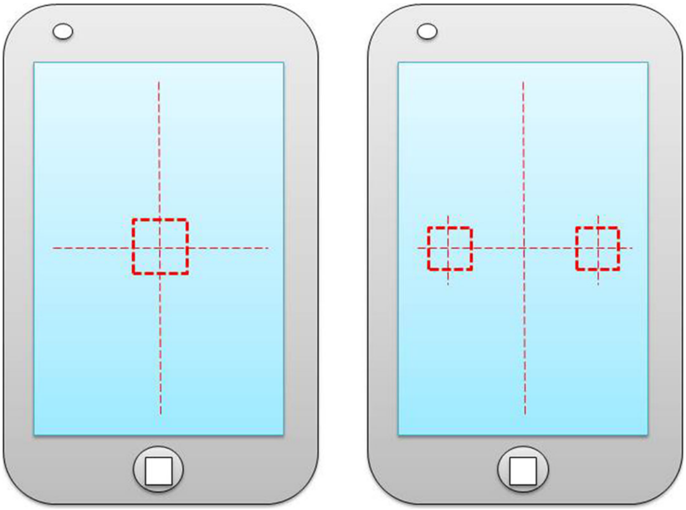


Fig. 4 The position of single (*left*) and double (*right*) reference points in camera plane

camera image plane for such two situations. Interest points detection method of computer vision can be used for reference points detecting.

2.10 Reference Points Deployment

For complex indoor positioning environments, reference points deployment is needed. For instance, the light from the reference point to the smart phone may be stopped by obstacle. For multi-floor environments, reference points must be placed in each floor. Each reference point must have distinct feature and position. LEDs with different color or shape, the line array of LEDs, or the plane array of LEDs can be used for reference points deployment.

3 Experiments

3.1 Experimental Environments

Figure 5 illustrates the experimental environments. One or two high power lighting LEDs, as reference points, are placed on the top of the experimental room. The world coordinate system is placed at the corner of the experimental room, and its axes are parallel to the related axes of the sensor coordinate system. The smart phone can be located at any point of the experimental room. The SOS can be any popular smart phone operating system, such as Android, iOS and Windows Mobile. The hardware of smart phone includes camera and orientation sensor, which is obtained from hardware sensors. An SOS application is made to take photos, capture the date from orientation sensor and locate the smart phone. Reference points detection depends on user cooperation.

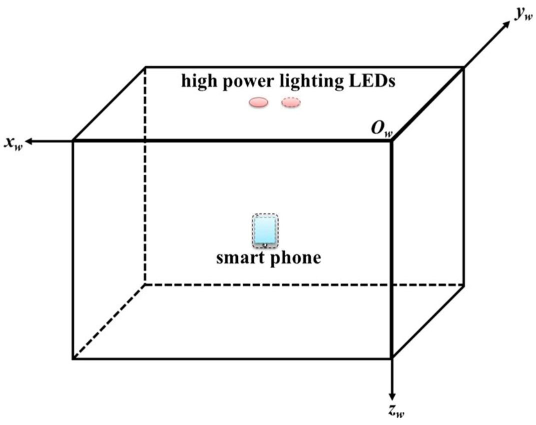


Fig. 5 The position of single (*left*) and double (*right*) reference points in world coordinate system

3.2 Camera Calibration Criterion

In order to obtain the optimal camera calibration results, the absolute error of reconstruction is defined as following formulas.

$$\begin{aligned} & e_r(\boldsymbol{S}_{fr}) = \|\boldsymbol{L}_s - \boldsymbol{L}_0\|_2 = \sqrt{(\boldsymbol{L}_s - \boldsymbol{L}_0)^T(\boldsymbol{L}_s - \boldsymbol{L}_0)} \\ & \boldsymbol{S}_{fr} = [\,s_{xf} \quad s_{yf}\,]^T, \boldsymbol{L}_s = [\,x_s \quad y_s \quad z_s\,]^T, \boldsymbol{L}_0 = [\,x_0 \quad y_0 \quad z_0\,]^T \end{aligned} \tag{14}$$

where $\boldsymbol{S}_{fr}$ is the candidate of camera parameter been estimated, $\boldsymbol{L}_s$ is the estimated position with camera parameter $\boldsymbol{S}_{fr}$, $\boldsymbol{L}_0$ is the true position.

Because the output of orientation sensor on smart phone has large random errors, it must be modified as following formulas.

$$\begin{aligned} & \boldsymbol{\theta}_m = \boldsymbol{\theta}_r + \boldsymbol{\delta} \\ & \boldsymbol{\delta} = [\,\delta_x \quad \delta_y \quad \delta_z\,]^T, \delta_x, \delta_y, \delta_z \in \left[-10^0, +10^0\right] \end{aligned} \tag{15}$$

where $\boldsymbol{\theta}_m$ is the modified vector of rotation angle, $\boldsymbol{\theta}_r$ is the original vector of rotation angle, $\boldsymbol{\delta}$ is the random error vector of orientation-sensor.

The optimal $\boldsymbol{S}_{fr}$ can be gained by finding the best experimental sample with least absolute error of reconstruction, and it is depicted as following formula. Exhaust searching from -10^0 to $+10^0$ with 1^0 interval is used to seek the suitable $\boldsymbol{\delta}$ for optimal $\boldsymbol{S}_{fr}$.

$$\min\ e_r\left(\boldsymbol{S}_{fr}(\boldsymbol{\delta})\right) \tag{16}$$

3.3 Camera Calibrating with Single Reference Point

The optimal value of δ_x, δ_y, δ_z, s_{xf} and s_{yf} for camera calibration with single reference point are showed in Table 1.

3.4 Camera Calibrating with Double Reference Points

The optimal value of δ_x, δ_y, δ_z, s_{xf} and s_{yf} for camera calibration with double reference points are showed in Table 2. Tables 1 and 2 indicate, one and two reference points based

Table 1 Experimental results of camera calibration with single reference point

Item	Value
δ_x	$+3^0$
δ_y	-2^0
δ_z	$+3^0$
s_{xf}	+2584.0
s_{yf}	−3345.8

methods have almost the same calibration results. Thus one reference point based calibration can take the place of two reference points based calibration in practice. One reference point based calibration is more convenient for implementation than two references points based calibration. Tables 1 and 2 show that, there is certain random deviation of the data from orientation-sensor on smart phone, at the level of several degrees. Tables 1 and 2 also prove that, the absolute value of s_{xf} is not equal to that of s_{yf}. This is because the horizontal scale property of camera is different to the vertical scale property. The negative sign of s_{yf} comes from the top–bottom exchange of digital image.

3.5 3D Positioning with Single Reference Point

t_z is determined by user cooperation. The absolute error of location estimation method is defined as following.

$$
\begin{aligned}
& e_i = \|\boldsymbol{L}_i - \boldsymbol{L}_{i0}\|_2 = \sqrt{(\boldsymbol{L}_i - \boldsymbol{L}_{i0})^T(\boldsymbol{L}_i - \boldsymbol{L}_{i0})} \\
& \boldsymbol{L}_i = [x_i \quad y_i \quad z_i]^T, \boldsymbol{L}_{i0} = [x_{i0} \quad y_{i0} \quad z_{i0}]^T \\
& E(e) = \frac{1}{K}\sum_{i=1}^{K} e_i, \mathrm{var}(e) = \frac{1}{K}\sum_{i=1}^{K}(e_i - E(e))^2
\end{aligned}
\tag{17}
$$

where $\boldsymbol{L}_i$ is the estimated position, $\boldsymbol{L}_{i0}$ is the true position, K is the number of location, E(e) is the average of absolute error, var(e) is the variance of absolute error.

The calibration results ($\boldsymbol{\delta}$ and $\boldsymbol{S}_{fr}$) in Table 1 are adopted for positioning. Table 3 shows the average and variance of absolute error for 3D positioning with single reference point. Figure 6 shows the experimental results of 3D positioning with single reference point. The red point means reference point, the green points mean true positions, and the blue points mean estimated positions.

3.6 3D Positioning with Double Reference Points

Least square method is used to compute the location of the smart phone. The calibration results ($\boldsymbol{\delta}$ and $\boldsymbol{S}_{fr}$) in Table 2 are adopted for positioning. Table 4 shows the average and

Table 2 Experimental results of camera calibration with double reference points

Item	Value
δ_x	$+3^0$
δ_y	-3^0
δ_z	$+3^0$
s_{xf}	+2583.8
s_{yf}	−3345.9

Table 3 Experimental results of 3D positioning with single reference point

Item	Value
E(e)	40.8802 (cm)
var(e)	79.2371 (cm^2)

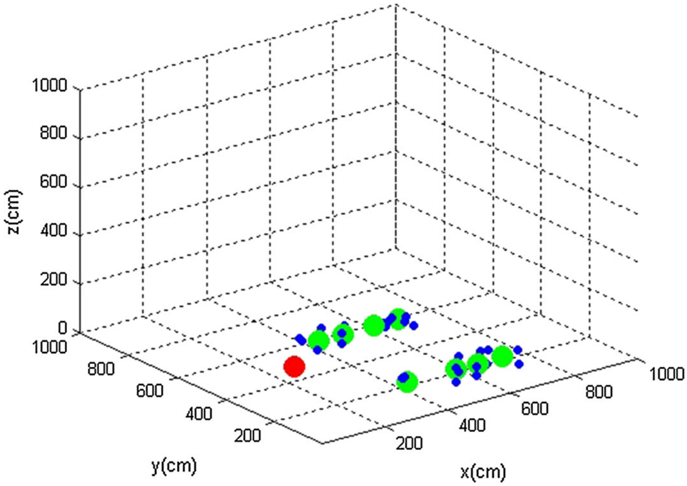


Fig. 6 The experimental results of 3D positioning with single reference point

variance of absolute error for 3D positioning with double reference points. Tables 3 and 4 indicate, 3D positioning with double reference points has better location accuracy than that of single reference point. For single reference point based method, t_z is estimated by user cooperation and has larger measurement error than that of double reference points based method. Figure 7 illustrates the experimental results of 3D positioning with double reference points. The red points represent reference points, the green points represent true positions, and the blue points represent estimated positions. Figures 6 and 7 show that, 3D positioning method with double reference points have higher positioning precision. But single reference point based positioning method needs less computing time and memory than that of double reference points.

Table 4 Experimental results of 3D positioning with double reference point

Item	Value
E(e)	22.8182 (cm)
var(e)	28.7403 (cm^2)

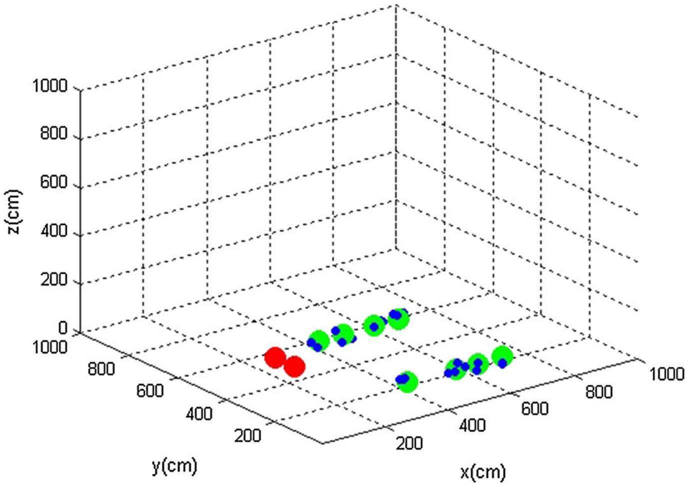


Fig. 7 The experimental results of 3D positioning with double reference points

3.7 Comparison Between P3P and the Proposed Method

Simulation experiments have also been done to evaluate the performance of the classical P3P algorithm, the proposed P2P algorithm, which means positioning with double reference points, and the proposed P1P algorithm, which represents positioning with single reference point.

Laptop computer for simulation experiments is DELL INSPIRON N5010, which includes 64-bit Intel quad-core processors. The host operating system is 64-bit Windows 8, and the management software of virtual machine is 32-bit VMWare 10.0.1, which is installed on Windows 8. The guest operating system is 32-bit Windows XP, and the software tool for application development is 32-bit Matlab 7.4.0, which is installed on Windows XP. Related Matlab applications are developed for simulation experiments.

Table 5 shows the experimental results, which include the average of total iterative number, the average of elapsed time, and the size of memory requirement. Table 5 illustrates, P1P and P2P are faster than P3P, but P3P has the least size of memory requirement. P3P is slower because it is an iterative algorithm. The memory requirement of P1P and P2P is acceptable in modern smart phone. In further, P3P must detect at least three reference points, but P1P and P2P only need to detect one and two reference points. One the contrary, more reference points means higher calibrating and positioning accuracy. So,

Table 5 Experimental results of comparison between P3P and the proposed method

Algorithm	Iterative number	Elapsed time (us)	Memory size (byte)
P3P	4	169.23	200
P2P	1	148.81	480
P1P	1	79.68	408

more reference points can be used in the phase of camera calibration. Table 5 also shows, P1P is better than P2P in elapsed time and memory size.

4 Conclusions

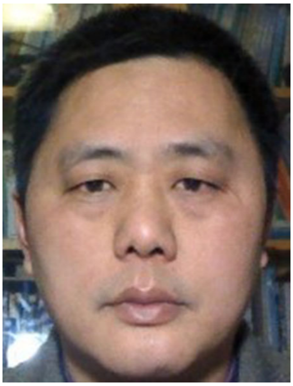
This paper discusses the high precision 3D indoor positioning method with camera and orientation-sensor on smart phone. Firstly, 3D camera imaging model, calibrating and positioning methods are studied. One or two reference points are used for camera calibrating and 3D positioning. The proposed camera calibrating method is easy to implement. The proposed positioning method is equal to P2P or P1P, but is faster and has unique solution. Double reference points based positioning method has higher precision than that of one reference point, but the latter needs less computing time and memory than that of the former. Secondly, simulation experiments of calibrating and positioning are implemented. Experimental results show, the proposed method has the advantages of simplicity, high-precision, usability and low–cost. So the proposed method is one of the ideal selections, for low-cost and high precision 3D indoor positioning.

The main error comes from the measurement precision of orientation-sensor on smart phone. Higher precision of orientation will result in higher positioning precision. Another disadvantage is that, reference points detection method depends on user cooperation, and it can be substituted by classical interest-point detection algorithms of computer vision. Finally, experimental environments are very ideal, complicated positioning environments should be researched. Such disadvantages will be resolved in our future work.

References

1. Harle, R. (2013). A survey of indoor inertial positioning systems for pedestrians. *IEEE Communications Surveys and Tutorials, 15*(3), 1281–1293.
2. Al Nuaimi, K. (2011). A survey of indoor positioning systems and algorithms, *In Proceedings of international conference on innovations in information technology* (pp. 185–190).
3. Mautz, R. (2011). Survey of optical indoor positioning systems, In *Proceedings of international conference on indoor positioning and indoor navigation* (pp. 1–7).
4. Atia, M. M. (2013). Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks. *IEEE Transactions on Mobile Computing, 12*(9), 1774–1787.
5. Kim, J. (2008). Vision-based location positioning using augmented reality for indoor navigation. *IEEE Transactions on Consumer Electronics, 54*(3), 954–962.
6. Lee, S. (2006). A pyroelectric infrared sensor-based indoor location-aware system for the smart home. *IEEE Transactions on Consumer Electronics, 52*(4), 1311–1317.
7. Suski, W. (2013). Using a map of measurement noise to improve UWB indoor position tracking. *IEEE Transactions on Instrumentation and Measurement, 62*(8), 2228–2236.
8. Wang, T. Q. (2013). Position accuracy of time-of-arrival based ranging using visible light with application in indoor localization systems. *Journal of Lightwave Technology, 31*(20), 3302–3308.
9. Gritton, C.W.K. (2012). MEMS sensors: Enabling next-gen content discovery, navigation and interactivity in CE devices, In *Proceedings of IEEE international symposium on consumer electronics* (pp. 1–8).
10. Atzori, L. (2012) Indoor navigation system using image and sensor data processing on a smartphone, In *Proceedings of international conference on optimization of electrical and electronic equipment* (pp. 1158–1163).
11. Wang, Q. (2012). Accurate indoor tracking using a mobile phone and non-overlapping camera sensor networks," In *Proceedings of IEEE international instrumentation and measurement technology conference* (pp. 2022–2027).
12. Taylor, B. (2012). Smart phone-based Indoor guidance system for the visually impaired, In *Proceedings of international conference on control automation robotics and vision* (pp. 871–876).

13. Ruotsalainen, L. (2011). Heading change detection for indoor navigation with a smartphone camera, In *Proceedings of international conference on indoor positioning and indoor navigation* (pp. 1–7).
14. Werner, M. (2011). Indoor positioning using smartphone camera, In *Proceedings of international conference on indoor positioning and indoor navigation* (pp. 1–6).
15. Juang, S.Y. (2012). Real-time indoor surveillance based on smartphone and mobile robot, In *Proceedings of IEEE international conference on industrial informatics* (pp. 475–480).
16. Ascher, C. (2010). Dual IMU Indoor Navigation with particle filter based map-matching on a smartphone, In *Proceedings of international conference on indoor positioning and indoor navigation*, (pp. 1–5).
17. Van Vinh, N. (2012). Self-positioning system for indoor navigation on mobile phones, In *Proceedings of IEEE international conference on consumer electronics* (pp. 114–115).
18. Zhang, R. (2013). Indoor localization using a smart phone, In *Proceedings of IEEE sensors applications symposium* (pp. 38–42).
19. Faragher, R.M. (2012). Opportunistic radio SLAM for indoor navigation using smartphone sensors, In *Proceedings of IEEE/ION position location and navigation symposium* (pp. 120–128).
20. Khalifa, S. (2013). Converting context to indoor position using built-in smartphone sensors, In *Proceedings of IEEE international conference on pervasive computing and communications workshops* (pp. 423–424).
21. Kang, W. (2012). Improved heading estimation for smartphone-based indoor positioning systems, In *Proceedings of IEEE international symposium on personal indoor and mobile radio communications* (pp. 2449–2453).
22. Pratama, A.R. (2012) Smartphone-based Pedestrian Dead Reckoning as an indoor positioning system, In *Proceedings of international conference on system engineering and technology* (pp. 1–6).
23. Lukianto, C. (2010). Pedestrian smartphone-based indoor navigation using ultra-portable sensory equipment, In *Proceedings of international conference on indoor positioning and indoor navigation* (pp. 1–5).
24. Ausmeier, B. (2012). Indoor navigation using a mobile phone, In *Proceedings of african conference on software engineering and applied computing* (pp. 109–115).
25. Haralick, R. M. (1994). Review and analysis of solutions of the three point perspective pose estimation problem. *International Journal of Computer Vision, 13*(3), 331–356.
26. Gao, X. (2003). Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(8), 930–943.
27. Ahmadinejad, F. (2013). A low-cost vision-based tracking system for position control of quadrotor robots, In *Proceedings of RSI/ISM international conference on robotics and mechatronics* (pp. 356–361).
28. Dementhon, D. F. (1995). Model-based object pose in 25 lines of code. *International Journal of Computer Vision, 15*(1), 123–141.
29. Qingxuan, J. (2006). The study of positioning with high-precision by single camera based on P3P algorithm, In *Proceedings of IEEE international conference on industrial informatics* (pp. 1385–1388).
30. D'Alfonso, L. (2013). P3P and P2P problems with known camera and object vertical directions, In *Proceedings of Mediterranean conference on control and automation*, (pp. 444–451).
31. Aanæs, H. (2012). Interesting Interest Points. *International Journal of Computer Vision, 97*(1), 18–35.
32. Gauglitz, S. (2011). Evaluation of interest point detectors and feature descriptors for visual tracking. *International Journal of Computer Vision, 94*(3), 335–360.
33. Keller, F. (2012). Calibration of smartphones for the use in indoor navigation, In *Proceedings of international conference on indoor positioning and indoor navigation* (pp. 1–8).



Honggui Li was born in Yangzhou, China, in 1971. He received B.S. degree in applied electronics from Yangzhou University in 1994, and received Ph.D. degree in mechatronic engineering from Nanjing University of Science and Technology in 1999. He is a senior member of Chinese Institute of Electronics. He works at Yangzhou University in China. His research interests include embedded computing, computer vision and machine learning.