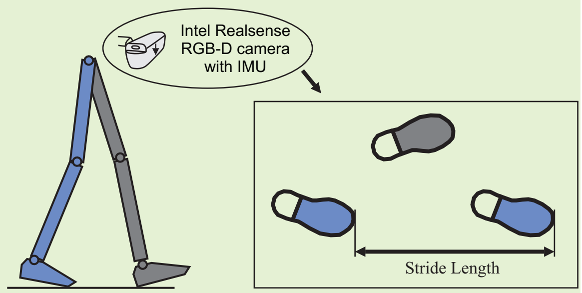
Visual inertial odometry based gait analysis using waist-attached RGB-D camera and inertial sensors

Duc Cong Dang, and Young Soo Suh, *Member, IEEE*

***Abstract*—In this paper, a visual inertial odometry algorithm is proposed to estimate walking stride length and reconstruct walking trajectory. Depth and color image data from a downward-looking waist-mounted RGB-D camera is fused with its internal Inertial Measurement Unit (IMU) data in an estimation algorithm to perform foot detection and position estimation. Floor plane and foot positions in stance phases are calculated and used as landmarks to construct measurement equations for updating in the filter. A smoothing problem is formulated as a linear optimization problem to improve filter result. Experiments are performed to evaluate the walking trajectory reconstruction and the overall root mean square errors (RMSE) of walking stride length estimation is about 3.8 centimeters.**



***Index Terms*—Visual inertial odometry, Gait analysis, RGB-D camera, Inertial sensor, Kalman filter.**

I. Introduction

Gait analysis is to describe human walking ability based on quantitative parameters such as cadence, stride length, and walking speed [1]. Even though considerable research has been done in gait analysis, this area is still gaining attraction largely with the advent of new technologies and techniques [2].

Recently, a variety of sensor-based systems are developed to perform gait analysis, where sensors are placed on the floor (force sensors), inside insole of shoes (pressure or sensors), or on the subject body (wearable systems like inertial sensors or EMG electrodes). Floor platform-based and pressure sensors provide an accurate measurement of force pattern and foot pressure distribution to detect step and gait phases [3]. A force sensing resistor or piezoelectric-based in-socket sensor system can be used to detect gait phases from signal responses [4], [5]. EMG is used to measure the muscle electrical activity during walking, and to derive gait phases using the amplitude of EMG signals [6]. There are limitations of space and cost for those non-wearable systems.

Inertial sensors provide a cost effective solution with high sampling rates and require low computing capability to detect steps and estimate gait parameters [7]–[9]. However, accuracy and robustness of inertial sensors-based systems need to improve due to accumulative errors.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2019R1I1A3A01057414).

Duc Cong Dang is with the University of Ulsan, 44607 Korea (e-mail: congdd.ac@gmail.com).

Young Soo Suh is with the Department of Electrical, Electronic and Computer Engineering, University of Ulsan, 44607 Korea (e-mail: yssuh@ulsan.ac.kr).

Vision based gait analysis can be separated into direct and indirect approaches. Direct method with optoelectronic systems is a gold standard for movement tracking based on markers placed on certain key points of the body with a sub-millimeter precision [10], [11]. However, those systems are expensive, difficult to set up, and cannot be used outside the clinic or laboratory. Indirect vision based approach extract features of the subjects (gender or human identity [12]) and gait parameters (step length and duration or ankle angles and distance [13], [14]) from image processing and machine learning technique through a sequence of images. Others use single or multiple fixed depth cameras to reconstruct 3-D shape and volume during walking which can be used to extract skeleton data and detect symmetrical gait troubles [15], [16].

Visual odometry (VO) is widely used in autonomous navigation systems, humanoid robots or aerial vehicles to track and reconstruct the motion of the camera in real-time using sequential images [17]–[19]. However, there is currently no research on applying visual inertial odometry to gait analysis. In this paper, we propose a visual-inertial approach to detect feet and estimate walking trajectory for gait parameter estimation, where a body-installed RGB-D camera are used for data acquisition. Relative attitude and relative position estimated from a visual odometry algorithm and detected stance foot position from depth image are used as measurement updates for a filtering and smoothing algorithm to compute two feet's trajectories. The proposed algorithm is designed based on the structure given in Fig. 1, where stance foot position estimation is the main objective.

The remainder of this study is organized as follows. Section II presents our system setup for data acquisition, the definition

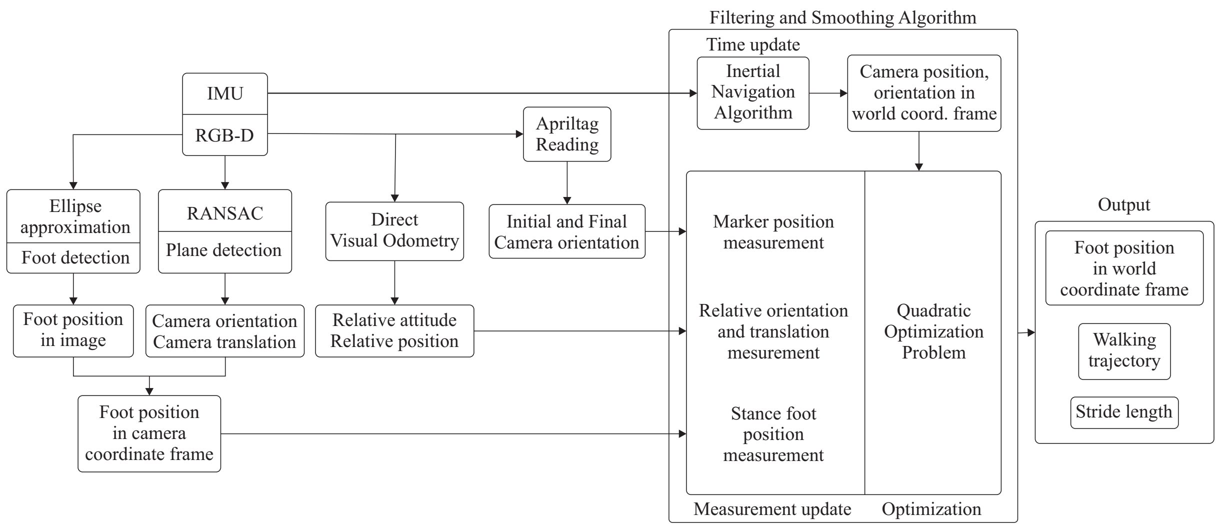


Fig. 1. Algorithm structure.

of the coordinate systems, and some notations. Section III describes the system equations and state definition for an inertial system along with extended states for the visual odometry information. Measurement equation from the visual odometry algorithm is derived for updating relative position and relative orientation. Section IV presents the floor plane parameter estimation and foot tracking algorithm. Measurement equations for the markers at the initial and final step, and the measurement equation for the stance foot which is used as a landmark, are derived in Section V. The filtering algorithm and the formulation of the smoothing algorithm are summarized in Section VI. Section VII presents the experiment results and the discussion of the results, and finally Section VIII summarizes the conclusions and potential future works.

II. System Overview

A. Hardware setup

The system setup is shown in Fig. 2. Images are acquired using an Intel Realsense D455 RGB-D camera with a resolution of 640x480 pixels and a frame rate of 30Hz. Internal IMU of the camera is configured to provide 200Hz accelerometer and 200Hz gyroscope data. Two infrared markers are mounted on top of both feet to provide ground truth data from an optical tracker system. The intrinsic parameters of the camera are pre-calibrated using a chessboard with Camera Calibration Toolbox of MATLAB. Images and inertial data timestamps are synchronized.

B. Notation

For a vector $a \in R^3$, $[a\times] \in R^{3\times 3}$ denotes the corresponding skew-symmetric matrix. For a quaternion $q \in R^4$, $C(q) \in \mathrm{SO}(3)$ denotes the corresponding rotation matrix. $q_{identity} = \begin{bmatrix} 1 & 0_{1\times 3} \end{bmatrix}'$ denotes the identity quaternion. Let $\hat{q}$ denote the estimated value of a quaternion $q \in R^4$. Assuming the estimation error is small, we use the following three vector error model $\bar{q}_e \in R^3$ [20]:

$$q = \hat{q} \otimes f_{quat}(\bar{q}_e) \tag{1}$$

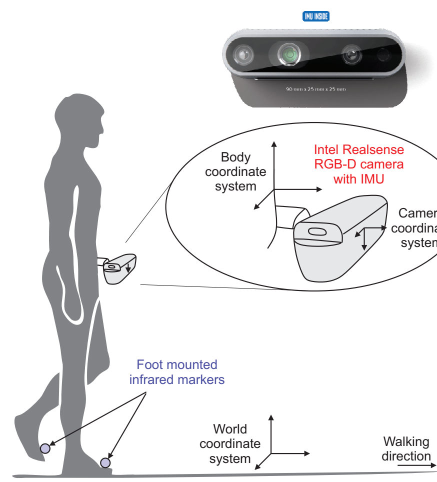


Fig. 2. System overview.

where $\otimes$ denotes the quaternion multiplication, and

$$f_{quat}(\bar{q}_e) = \begin{bmatrix} 1 \\ \bar{q}_e \end{bmatrix} \in \begin{bmatrix} R \\ R^3 \end{bmatrix}.$$

Representing (1) with the rotation matrix, we obtain

$$C(q) = (I - 2[\bar{q}_e\times])C(\hat{q}).$$

Let $f_{vector} : R^4 \to R^3$ be a function extracting the vector part of a quaternion. For a matrix $A \in R^{n\times m}$, $A(r_1 : r_2, c_1 : c_2) \in R^{(r_2-r_1+1)\times(c_2-c_1+1)}$ denotes a submatrix of A consisting of rows $\{r_1, \cdots, r_2\}$ and columns $\{c_1, \cdots, c_2\}$.

III. Visual Inertial Odometry Filtering

In this section, the state variables are defined to derive dynamic equations for a Kalman-based filter using IMU data. These states are extended to include visual odometry information. The visual odometry algorithm provides relative position and attitude of consecutive image frames as the measurement updating for the filter.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2022.3227950

A. System equation and state definition

Let $q \in R^4$, $r \in R^3$ and $v \in R^3$ be quaternion, position and velocity of IMU. The quaternion q represents the rotation from the world coordinate system to the body coordinate system. Let $y_a \in R^3$ and $y_g \in R^3$ be the accelerometer and gyroscope outputs:

$$\begin{array}{rcl} y_a & = & C(q)\tilde{g} + a_b + b_a + \eta_a \\ y_g & = & \omega + b_g + \eta_g \end{array} \quad (2)$$

where $\tilde{g} \in R^3$ is the local gravitation vector, $a_b \in R^3$ is the external acceleration and $\omega \in R^3$ is the angular velocity. $b_a \in R^3$ and $b_g \in R^3$ are the accelerometer and gyroscope bias. $\eta_a \in R^3$ and $\eta_g \in R^3$ are white Gaussian sensor noises, whose covariances are given by $r_a I_3$ and $r_g I_3$.

Usually, the sampling period T_{camera} of a camera is larger than the sampling period T_{imu} of an IMU. In this paper, we assume that the sampling period of a camera is an integer multiple of the sampling period of an IMU: that is, $T_{camera} = M_{ratio} T_{imu}$ for a positive integer M_{ratio}. There are two discrete time indices k (with the sampling period T_{camera}) and i (with the sampling period T_{imu}). The function $f_{discrete}(k)$ relates two discrete indices:

$$i = f_{discrete}(k) = (k-1)M_{ratio} + 1.$$

In the attitude and position filtering and smoothing, estimation error terms are usually estimated instead of direct estimation of attitude and position [21]. The estimation error terms $\bar{q}_{e,i} \in R^3$, $r_{e,i} \in R^3$ and $v_{e,i} \in R^3$ are defined by:

$$\begin{array}{rcl} q_i & = & \hat{q}_i \otimes f_{quat}(\bar{q}_{e,i}) \\ r_i & = & \hat{r}_i + r_{e,i} \\ v_i & = & \hat{v}_i + v_{e,i} \end{array} \quad (3)$$

where $\hat{q}_i \in R^4$, $\hat{r}_i \in R^3$ and $\hat{v}_i \in R^3$ represent estimated values of q_i, r_i and v_i, respectively.

The dynamic equation of estimation error terms is given by [20]

$$\begin{bmatrix} \bar{q}_{e,i+1} \\ r_{e,i+1} \\ v_{e,i+1} \\ b_{g,e,i+1} \\ b_{a,e,i+1} \end{bmatrix} = F_i \begin{bmatrix} \bar{q}_{e,i} \\ r_{e,i} \\ v_{e,i} \\ b_{g,e,i} \\ b_{a,e,i} \end{bmatrix} + \zeta_i \quad (4)$$

where $b_{g,e,i} \in R^3$ and $b_{a,e,i} \in R^3$ are error terms of sensor biases b_g and b_a. The covariance of noise ζ_i is given by [20]

$$Q_d = \mathrm{E}\{\zeta_i \zeta_i'\} = \mathrm{Diag}(0.25 r_g I_3, 0_{3\times3}, r_a I_3, 0_{9\times9}, Q_{b_g}, Q_{b_a}).$$

F_i is computed by

$$F_i = \begin{bmatrix} F_{11,i} & F_{12,i} \\ 0_{6\times9} & I_6 \end{bmatrix} = \exp(\begin{bmatrix} A_{11,i} & A_{12,i} \\ 0_{6\times9} & 0_{6\times6} \end{bmatrix} T_{imu}) \quad (5)$$

where

$$A_{11,i} = \begin{bmatrix} [-y_{g,i}\times] & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & I_3 \\ -2C'(\hat{q}_i)[y_{a,i}\times] & 0_{3\times3} & 0_{3\times3} \end{bmatrix} \in R^{9\times9}$$

$$A_{12} = \begin{bmatrix} -0.5 I_3 & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & -C'(\hat{q}_i) \end{bmatrix} \in R^{9\times6}.$$

Note that (4) is a discrete-time system equation with the sampling period T_{imu}. Since measurement information from the camera is available with the sampling T_{image}, (4) is transformed to a discrete-time system equation with the sampling period T_{image}.

Let $U_k \in R^9$ (notice that the image discrete index k is used instead of IMU discrete index i) be an error state vector defined by

$$U_k = \begin{bmatrix} \bar{q}_{e,f_{discrete}(k)} \\ r_{e,f_{discrete}(k)} \\ v_{e,f_{discrete}(k)} \end{bmatrix} \in R^{9\times1}. \quad (6)$$

Let $V_i \in R^{15}$ be a state vector with bias terms defined by

$$V_k = \begin{bmatrix} U_k \\ b_{g,e,f_{discrete}(k)} \\ b_{a,e,f_{discrete}(k)} \end{bmatrix} \in R^{15\times1}. \quad (7)$$

where $b_{g,e} \in R^3$ and $b_{a,e} \in R^3$ are error terms of sensor biases b_g and b_a.

By repeating (4) M_{ratio} times, the following equation is given by

$$V_{k+1} = \bar{F}_k V_k + \bar{\zeta}_k \quad (8)$$

where

$$\bar{F}_k = \prod_{i=f_{discrete}(k)}^{f_{discrete}(k)+M_{ratio}-1} F_i,$$

$$Q_{\bar{\zeta}} = \mathrm{E}\{\bar{\zeta}_k \bar{\zeta}_k'\} = \sum_{j=1}^{M_{ratio}} G_{k+j-1} Q_d G_{k+j-1}',$$

$$G_k = I_{15}, \; G_{k+j} = G_{k+j-1} \bar{F}_{f_{discrete}(k+1)-j},$$

$$1 \le j \le M_{ratio} - 1.$$

From the structure of F_i (see (5)), $\bar{F}_k$ and $\bar{\zeta}_k$ are partitioned as follows for later use:

$$\bar{F}_k = \begin{bmatrix} \bar{F}_{11,k} & \bar{F}_{12,k} \\ 0_{6\times9} & I_6 \end{bmatrix} \in \begin{bmatrix} R^{9\times9} & R^{9\times6} \\ R^{6\times9} & R^{6\times6} \end{bmatrix},$$

$$\bar{\zeta}_k = \begin{bmatrix} \bar{\zeta}_{k,1} \\ \bar{\zeta}_{k,2} \end{bmatrix} \in \begin{bmatrix} R^9 \\ R^6 \end{bmatrix},$$

$$Q_{\bar{\zeta}_k} = \begin{bmatrix} Q_{\bar{\zeta}_k,11} & Q_{\bar{\zeta}_k,12} \\ Q_{\bar{\zeta}_k,21} & Q_{\bar{\zeta}_k,22} \end{bmatrix} \in \begin{bmatrix} R^{9\times9} & R^{9\times6} \\ R^{6\times9} & R^{6\times6} \end{bmatrix}. \quad (9)$$

The visual odometry information at k-th discrete time is obtained using $(k-1)$-th and k-th images; thus the information imposes constraints on U_{k-1} and U_k. To cope with this fact, we construct an extended state X_k containing both U_{k-1} and U_k [22].

Let the extended state X_k be defined by

$$X_k = \begin{bmatrix} U_k \\ U_{k-1} \\ b_{g,e,k} \\ b_{a,e,k} \end{bmatrix} \in R^{24\times1}.$$

From (8), we have the following system equation for the Kalman filter.

$$X_{k+1} = \begin{bmatrix} \bar{F}_{11,k} & 0_{9\times 9} & \bar{F}_{12,k} \\ I_9 & 0_{9\times 9} & 0_{9\times 6} \\ 0_{6\times 9} & 0_{6\times 9} & I_6 \end{bmatrix} X_k + \bar{\eta}_k \tag{10}$$

where

$$\bar{\eta}_k = \begin{bmatrix} \bar{\zeta}_{k,1} \\ 0_{9\times 1} \\ \bar{\zeta}_{k,2} \end{bmatrix} \in \begin{bmatrix} R^9 \\ R^9 \\ R^6 \end{bmatrix}.$$

Let $Q_{\bar{\eta}_k}$ be the covariance of $\bar{\eta}$, which can be computed from $Q_{\bar{\zeta}}$.

B. Visual odometry

Visual odometry estimates the motion of a camera using sequential images [23]. Visual odometry can be divided into two methods: indirect and direct methods. Indirect methods extract and track point features in the environment. Direct methods use raw image data by comparing pixel intensities of consecutive images. In this paper, the direct method in [24] is used since this method is known to be robust in low-texture environments. Note that the camera in this paper is looking downward to floor, which generally gives low-texture images.

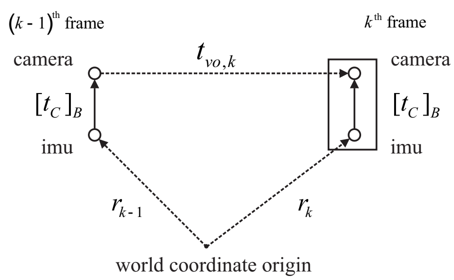


Fig. 3. Relative camera pose of two consecutive image frames

Let r_k and q_k be the position and attitude of the IMU represented in the world coordinate corresponding to k-th image frame (see Fig. 3). Let $[t_C]_B$ and q_B^C be the position and orientation of the camera in IMU (body) coordinate. Since $[t_C]_B$ and q_B^C are known constants, the position and attitude of the camera in world coordinate can be calculated as $r_k + C(q_k)'[t_C]_B$ and $q_k \otimes q_B^C$ [20].

The visual odometry algorithm provides the relative attitude $q_{vo,k}$ and relative position $t_{vo,k}$ of the camera with its previous frame, (represented in the camera coordinate at the k-th discrete time) that are $(k-1)$-th and k-th images.

The relative attitude $q_{vo,k}$ and relative position $t_{vo,k}$ satisfies the following equations:

$$q_k \otimes q_B^C = q_{k-1} \otimes q_B^C \otimes q_{vo,k}. \tag{11}$$

$$\begin{aligned} t_{vo,k} = C_B^C C(q_k)((r_{k-1} + C(q_{k-1})'[t_C]_B) \\ -(r_k + C(q_k)'[t_C]_B)). \end{aligned} \tag{12}$$

Let $\hat{q}_{vo} \in R^4$ and $\hat{t}_{vo} \in R^3$ be visual odometry estimate values and $q_{vo,e} \in R^4$ and $t_{vo,e} \in R^3$ be estimation error terms defined by

$$\begin{aligned} q_{vo,k} &= \hat{q}_{vo,k} \otimes f_{quat}(\bar{q}_{vo,e,k}) \\ t_{vo,k} &= \hat{t}_{vo,k} + t_{vo,e,k}. \end{aligned} \tag{13}$$

Inserting (3) and (13) into (11), we have

$$\begin{aligned} q_{identity} &= (q_B^C)^* \otimes q_k^* \otimes q_{k-1} \otimes q_B^C \otimes q_{vo,k} \\ &= (q_B^C)^* \otimes \hat{q}_k^* \otimes \hat{q}_{k-1} \otimes q_B^C \otimes \hat{q}_{vo,k} \\ &-(q_B^C)^* \otimes f_{quat}(\bar{q}_{e,k}) \otimes \hat{q}_k^* \otimes \hat{q}_{k-1} \otimes q_B^C \otimes \hat{q}_{vo,k} \\ &+(q_B^C)^* \otimes \hat{q}_k^* \otimes \hat{q}_{k-1} \otimes f_{quat}(\bar{q}_{e,k-1}) \otimes q_B^C \otimes \hat{q}_{vo,k} \\ &+(q_B^C)^* \otimes \hat{q}_k^* \otimes \hat{q}_{k-1} \otimes q_B^C \otimes \hat{q}_{vo,k} \otimes f_{quat}(\bar{q}_{vo,e,k}). \end{aligned} \tag{14}$$

In deriving the above equation, second-order error terms are ignored.

Let $z_{vo,q,k} \in R^3$ (the vector part of a quaternion) be defined by

$$z_{vo,q,k} = f_{vector}\left((q_B^C)^* \otimes \hat{q}_k^* \otimes \hat{q}_{k-1} \otimes q_B^C \otimes \hat{q}_{vo,k}\right). \tag{15}$$

Let $a \in R^4$ and $b \in R^4$ be quaternions and $\bar{c} \in R^3$ be a vector. It is straightforward to verify that

$$a \otimes \begin{bmatrix} 0 \\ \bar{c} \end{bmatrix} \otimes b = \begin{bmatrix} \star \\ L(a,b)\bar{c} \end{bmatrix} \tag{16}$$

where '$\star$' term is irrelevant and

$$L(a,b) = -(\bar{b}\bar{a}') + (a_0 b_0)I + b_0[\bar{a}\times] - a_0[\bar{b}\times] - [\bar{b}\times][\bar{a}\times].$$

Combining (14), (15) and (16), we obtain

$$\begin{aligned} z_{vo,q,k} &= L((q_B^C)^*, \hat{q}_k^* \otimes \hat{q}_{k-1} \otimes q_B^C \otimes \hat{q}_{vo,k})\bar{q}_{e,k} \\ &-L((q_B^C)^* \otimes \hat{q}_k^* \otimes \hat{q}_{k-1}, q_B^C \otimes \hat{q}_{vo,k})\bar{q}_{e,k-1} \\ &-L((q_B^C)^* \otimes \hat{q}_k^* \otimes \hat{q}_{k-1} \otimes q_B^C \otimes \hat{q}_{vo,k}, q_{identity})\bar{q}_{vo,e,k}. \end{aligned} \tag{17}$$

Let $z_{vo,t,k}$ be defined by

$$z_{vo,t,k} = \hat{t}_{vo,k} - \hat{t}_k \tag{18}$$

where

$$\hat{t}_k = C_B^C C(\hat{q}_k)(\hat{r}_{k-1} - \hat{r}_k) + C_B^C(C(\hat{q}_k)C(\hat{q}_{k-1})' - I)[t_C]_B.$$

Combining (14) and (18), we obtain

$$\begin{aligned} z_{vo,t,k} &\approx C_B^C C(q_k)(r_{e,k-1} - r_{e,k}) \\ &-2C_B^C[\bar{q}_{e,k}\times]C(\hat{q}_k)(\hat{r}_{k-1} - \hat{r}_k) \\ &-2C_B^C[\bar{q}_{e,k}\times]C(\hat{q}_k)C(\hat{q}_{k-1})'[t_C]_B \\ &-2C_B^C C(\hat{q}_k)C(\hat{q}_{k-1})'[\bar{q}_{e,k-1}\times]'[t_C]_B - t_{vo,e,k} \\ &= C_B^C C(q_k)(r_{e,k-1} - r_{e,k}) \\ &+2C_B^C[(C(\hat{q}_k)(\hat{r}_{k-1} - \hat{r}_k))\times]\bar{q}_{e,k} \\ &+2C_B^C[(C(\hat{q}_k)C(\hat{q}_{k-1})'[t_C]_B)\times]\bar{q}_{e,k} \\ &-2C_B^C C(\hat{q}_k)C(\hat{q}_{k-1})'[[t_C]_B\times]q_{e,k-1} - t_{vo,e,k}. \end{aligned} \tag{19}$$

Let $z_{vo,k}$ be defined by

$$z_{vo,k} = \begin{bmatrix} z_{vo,q,k} \\ z_{vo,t,k} \end{bmatrix} \in \begin{bmatrix} R^3 \\ R^3 \end{bmatrix}. \tag{20}$$

The measurement equation from the visual odometry algorithm is given by

$$z_{vo,k} = H_{vo,k}X_k + \eta_{vo,k} \tag{21}$$

where $H_{vo,k} \in R^{6\times 24}$ can be derived from (17) and (19). Let $R_{vo,k} \in R^{6\times 6}$ denote the covariance of $\eta_{vo,k}$ which is given by

$$R_{vo,k} = J_{vo,k}\begin{bmatrix} r_{vo,q}I_3 & 0_{3\times 3} \\ 0_{3\times 3} & r_{vo,t}I_3 \end{bmatrix}.$$

where $J_{vo,k} \in R$ is the sum of squares of brightness residual of each pixel [24].

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2022.3227950

IV. Floor and Foot Detection

As in Fig. 4, walking cycle consists of stance period (when a foot is on the floor) and swing period (when a foot is swinging). A stance period is detected (Section IV-B) and is used as a landmark (Section V-B) since a foot does not move during stance period even if the other body parts are moving.

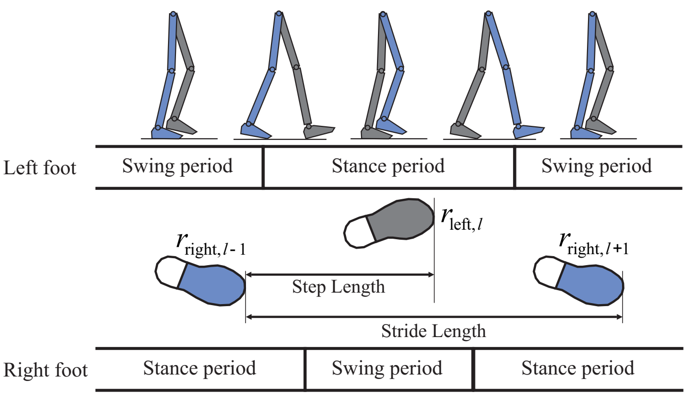


Fig. 4. Stance and swing period

Let $r_{left,l} \in R^3$ be the position (expressed in the world coordinate system) of the l-th left stance period foot. Similarly, $r_{right,l} \in R^3$ is defined for the right foot. The final goal of this paper is to estimate foot positions $r_{left,l}$ and $r_{right,l}$ and the walking stride length is computed from the foot positions (see Fig. 4).

In this paper, the stance period foot is detected from depth images. The foot could have been detected from RGB images. However, it is not easy to detect foot if the shoe and floor have similar color or lighting condition is poor. Since depth image does not depend on color and lighting conditions, more robust foot detection is possible.

A. Floor plane detection

From depth image, floor plane is detected. The floor plane helps to detect a stance period foot.

Let $[p(u,v)]_C \in R^3$ be a point on the floor represented in the camera coordinate system corresponding to (u,v) pixel, which is obtained from the depth camera. The floor plane equation is given by

$$[n_{floor}]_C[p]_C + [h_{floor}]_C = 0, \quad [h_{floor}]_C > 0 \tag{22}$$

where $[n_{floor}]_C \in R^3$ is the unit normal vector of the floor plane and $[h_{floor}]_C \in R > 0$ is the distance to the plane from the camera.

Plane parameters are estimated using RANSAC algorithm [25].

B. Foot detection and ellipse approximation

A stance period foot instep is slightly above the floor plane. Based on this observation, foot candidate points are selected.

Firstly, the height matrix $F_{height}(u,v)$ (the height from the floor) is constructed as follows:

$$F_{height}(u,v) = [\hat{h}_{floor}]_C - \hat{n}'_{floor}[p(u,v)]_C. \tag{23}$$

Then the foot candidate points are selected using the following condition:

$$h_{foot,min} \leq F_{height}(u,v) \leq h_{foot,max}. \tag{24}$$

For the pixels (u,v) satisfying (24), connected (8-direction connectivity) pixels are grouped [26]. If the number of points in a connected group is less than $N_{foot,min}$, then the group is discarded. If the number of remaining groups is greater than 2, only the two largest groups are selected as possible foot candidates.

For each candidate group of points, the foot edge is detected. Once foot edge is detected, foot is approximated by an ellipse with known major and minor axes length assuming that foot size is known.

Using a simple tracking algorithm, left and right feet are separated and stance period foot is detected. If a left stance period foot is detected at the k-th image, the position of a foot represented in the camera coordinate system is denoted by $y_{left,k} \in R^3$. Similarly $y_{right,k}$ is defined for the right foot. The detected stance feet belonging to the same stance period are grouped. For the l-th stance period step, let $S_{left,l}$ be a set of discrete indices with $y_{left,k}$ is available. An example is given in Fig. 5, where $S_{left,l}$ is given by

$$S_{left,l} = \{k, k+1, k+2, k+3\}. \tag{25}$$

Note that $y_{left,k}$ and $y_{left,k+1}$ are different since the camera is moving as the body is moving. However, if $y_{left,i}$ and $y_{left,i+1}$ are transformed to the world coordinate system, they should be the same since a stance period foot is not moving. Similarly, a set $S_{right,l}$ is defined for the right foot. Let $s_{index}(k)$ be a function representing the step number given image with discrete index k. For the example in (25), s_{index} is given by

$$l = s_{index}(k) = s_{index}(k+1) = s_{index}(k+2) = s_{index}(k+3).$$

Let S_{foot} be the union of all $S_{left,l}$ and $S_{right,l}$: that is, $k \in S_{foot}$ means that at least one stance foot is observed in the k-th image.

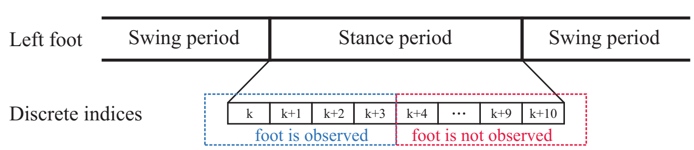


Fig. 5. Example of stance foot set $S_{left,l}$

Let $[y_{foot}]_C \in R^3$ be the stance foot (either left or right foot) position represented in the camera coordinate. The measured foot position is denoted by $\hat{y}_{foot}$ and the following measurement model is used:

$$y_{foot} = \hat{y}_{foot} + \eta_{foot} \tag{26}$$

where η_{foot} is a zero mean white Gaussian measurement noise with covariance $r_{foot}I_3 \in R^{3\times 3}$.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2022.3227950

V. Measurement Equations

In this section, in addition to measurement equation (21) from visual odometry algorithm, we propose two additional measurement updating as follows. April tags at known positions provide measurement updates in initial and final stances. Detected stance feet in Section IV are used as landmarks to reduce estimated foot position errors.

A. Measurement equation for markers on the floor

Two A3 size papers containing 5×7 Apriltag markers [27] are placed both in the starting and final walking points as in Fig. 6. Two papers are placed along the world coordinate y axis and the distance between the left upper Apriltags of two papers is denoted by L, which is measured with a tape measure.

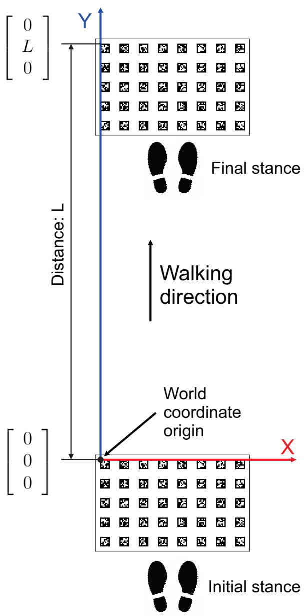


Fig. 6. Markers in the starting and final points

Let $[r_{tag,j}]_C \in R^3$ be four corners of Apriltags and $r_{img,j} \in R^2$ be the corresponding measured image coordinates. Then the two points are related as follows:

$$r_{img,j} = f_{img}([r_{tag,j}]_C) + \eta_{img,j} \tag{27}$$

where

$$[r_{tag,j}]_C = C'(q_C^B)(C(q)([r_{tag,j}]_W - r) - [P_c]_B). \tag{28}$$

Function $f_{img} : R^3 \to R^2$ is defined by

$$f_{img}(\begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix}) = \begin{bmatrix} \frac{x_j}{z_j} \\ \frac{y_j}{z_j} \end{bmatrix}.$$

An image measurement noise $\eta_{img,j} \in R^2$ is assumed to be independent Gaussian noise whose covariance is $\epsilon_1 I_2 \in R^{2\times 2}$. Note that $[r_{tag,j}]_W$ is known since the tag location on the A3 paper is known.

Attitude and position of the camera can be estimated using the homography estimation method in [28]. Let $\hat{q}_{tag,k} \in R^4$ and $\hat{r}_{tag,k} \in R^3$ denote estimated attitude and position and let $\bar{q}_{tag,e,k} \in R^3$ and $r_{tag,e,k} \in R^3$ denote the corresponding estimation errors, where the following is satisfied:

$$\begin{aligned} q_k &= \hat{q}_{tag,k} \otimes f_{quat}(\bar{q}_{tag,k,e}) \\ r_k &= \hat{r}_{tag,k} + r_{tag,k,e}. \end{aligned} \tag{29}$$

The estimation error covariance $R_{tag,k} \in R^{6\times 6}$ is given by

$$\begin{aligned} R_{tag,k} &= \mathrm{E}\{\begin{bmatrix} \bar{q}_{tag,k,e} \\ r_{tag,k,e} \end{bmatrix} \begin{bmatrix} \bar{q}_{tag,k,e} \\ r_{tag,k,e} \end{bmatrix}'\} \\ &= \epsilon_1 \left(\sum_{j=1}^{N_{tag,k}} \bar{H}'_{tag,j,k} \bar{H}_{tag,j,k} \right)^{-} \end{aligned} \tag{30}$$

where $N_{tag,k}$ is the number of observed tag corners in the k-th image.

$\bar{H}_{tag,j,k} \in R^{2\times 6}$ is given by

$$\begin{aligned} \bar{H}_{tag,j,k} &= \left. \frac{\partial f_{img}}{\partial r_{tag,j}} \frac{\partial r_{tag,j}}{\partial X_k} \right|_{r_{tag,j} = \hat{r}_{tag,j}} \\ &= \frac{1}{\hat{z}_i^2} \begin{bmatrix} \hat{z}_j & 0 & -\hat{x}_j \\ 0 & \hat{z}_j & -\hat{y}_j \end{bmatrix} J_{tag,j} \end{aligned}$$

where

$$[\hat{P}_{tag,j}]_C = C'(q_C^B)(C(\hat{q})([r_{tag,j}]_W - \hat{r}) - [t_C]_B) = \begin{bmatrix} \hat{x}_j \\ \hat{y}_j \\ \hat{z}_j \end{bmatrix} \tag{31}$$

$$\begin{aligned} J_{tag,j} = \big[& 2C'(q_C^B)[(C(\hat{q})([r_{tag,j}]_W - \hat{r}))\times], \\ & C'(q_C^B)C(\hat{q}) \big] \in R^{3\times 6}. \end{aligned} \tag{32}$$

$\hat{q}_{tag,1}$ and $\hat{r}_{tag,1}$ are used as initial attitude and position. After that only $\hat{r}_{tag,k}$ is used in the following measurement equation:

$$z_{tag,k} = \hat{r}_{tag,k} - \hat{r}_k = H_{tag,k} X_k + \eta_{tag,k} \tag{33}$$

where

$$H_{tag,k} = \begin{bmatrix} 0_{3\times 3} & I_3 & 0_{3\times 18} \end{bmatrix} \in R^{3\times 24}. \tag{34}$$

Recall that there are two A3 papers as in Fig. 6 which are in the starting and final points. When tags are observed in the starting point, $R_{tag,k}(4{:}6,4{:}6)$ is used as a measurement noise covariance. When tags are observed in the final point, the measurement noise covariance is increased to compensate uncertainties in L. Thus, the following measurement covariance is used:

$$R_{tag,k}(4{:}6,4{:}6) + \begin{bmatrix} \epsilon_2 I_2 & 0_{2\times 1} \\ 0_{1\times 2} & 0 \end{bmatrix}.$$

B. Measurement equation for the stance foot

The stance foot detected in Section IV is used as a landmark. The detected stance foot position is given by

$$\begin{aligned} [y_{foot}]_C &= (C_C^B)'(C(q)(\begin{bmatrix} \bar{r}_{foot} \\ 0 \end{bmatrix} - r) - [t_C]_B) \\ &= f_{foot}(q, r, \bar{r}_{foot}) \end{aligned} \tag{35}$$

where $\bar{r}_{foot} \in R^2$ is xy world coordinates for the stance foot position. The z axis value is 0 since a stance foot is on the ground.

As in SLAM algorithms [29], a new stance foot position in the world coordinate system is estimated whenever a new stance foot is observed. Let $\hat{\bar{r}}_{foot} \in R^2$ be the estimated foot position, which can be computed from (35) once $[\hat{y}_{foot}]_C$ is given. Let $\bar{r}_{foot,e,m} \in R^2$ be the m-th stance foot position estimation error:

$$\bar{r}_{foot,e,m} = \bar{r}_{foot,m} - \hat{\bar{r}}_{foot,m}.$$

The state X_k is extended to include $\bar{r}_{foot,e,m}$ term. Let the extended state $\bar{X}_{m-1,k}$ with $m-1$ observed stance feet be defined by

$$\bar{X}_{m-1,k} = \begin{bmatrix} X_k \\ \bar{r}_{foot,e,1,k} \\ \vdots \\ \bar{r}_{foot,e,m-1,k} \end{bmatrix} \in \begin{bmatrix} R^{24} \\ R^2 \\ \vdots \\ R^2 \end{bmatrix}. \quad (36)$$

Let the estimation error covariance of $\hat{\bar{X}}_{m-1,k}$ be denoted by $\bar{P}_{m-1,k}$:

$$\begin{aligned} \bar{P}_{m-1,k} &= \begin{bmatrix} P_{X,m-1,k} & P_{X,foot,m-1,k} \\ P_{foot,X,m-1,k} & P_{foot,m-1,k} \end{bmatrix} \\ &\in \begin{bmatrix} R^{24\times 24} & R^{24\times 2(m-1)} \\ R^{2(m-1)\times 24} & R^{2(m-1)\times 2(m-1)} \end{bmatrix}. \end{aligned} \quad (37)$$

When m-th stance foot is newly detected, the estimated state is extended as follows:

$$\hat{\bar{X}}_{m,k} = \begin{bmatrix} \hat{\bar{X}}_{m-1,k} \\ \hat{\bar{r}}_{foot,e,m,k} \end{bmatrix} \quad (38)$$

The state estimation error covariance $\bar{P}_{m,k}$ is given by

$$\bar{P}_{m,k} = \begin{bmatrix} \bar{P}_{m-1,k} & P'_{\star,m,k} \\ P_{\star,m,k} & P_{foot,m,k} \end{bmatrix} \quad (39)$$

where

$$P_{foot,m,k} = G_X P_{X,m-1,k} G'_X + r_{foot} G_y G'_y \quad (40)$$

$$P_{\star,m,k} = G_X \begin{bmatrix} P_{X,m-1,k} & P_{X,foot,m-1,k} \end{bmatrix} \quad (41)$$

and $G_X = \frac{\partial \bar{r}_{foot}}{\partial X}$ and $G_y = \frac{\partial \bar{r}_{foot}}{\partial v_{foot}}$ are computed from (44):

$$G_X = E_{12} \begin{bmatrix} 2C(\hat{q})'[([t_C]_B + C_C^B \hat{y}_{foot})\times] & I_3 & 0_{3\times 18} \end{bmatrix} \in R^{2\times 24} \quad (42)$$

$$G_y = E_{12} C(\hat{q})' C_C^B \in R^{2\times 3} \quad (43)$$

$$E_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

G_X is computed from the following equation (ignoring second-order error terms):

$$\begin{aligned} \bar{r}_{foot} &= E_{12}(\hat{r} + r_e + C(\hat{q})'(I + 2[q_e\times])([t_C]_B \\ &\quad + C_C^B(\hat{y}_{foot} + v_{foot}))) \\ &= E_{12}(\hat{r}_{foot} - 2C(\hat{q})'[([t_C]_B + C_C^B \hat{y}_{foot})\times]q_e \\ &\quad + r_e + C(\hat{q})' C_C^B v_{foot}). \end{aligned} \quad (44)$$

If the m-th detected foot $\hat{y}_{foot,m,k}$ has been observed before, it is used in the Kalman filter as a measurement. Let $z_{foot} \in R^3$ be defined by

$$z_{foot,m,k} = \hat{y}_{foot,k} - f_{foot}(\hat{q}_k, \hat{r}_k, \hat{\bar{r}}_{foot,m,k}). \quad (45)$$

From (35), we have

$$\begin{aligned} z_{foot,m,k} &= f_{foot}(q_k, r_k, \bar{r}_{foot,m,k}) - \eta_{foot,k} \\ &\quad - f_{foot}(\hat{q}_k, \hat{r}_k, \hat{\bar{r}}_{foot,m,k}) \\ &= 2(C_C^B)'[(C(\hat{q})(E'_{12}\hat{\bar{r}}_{foot,m,k} - \hat{r}))\times]\bar{q}_{e,k} \\ &\quad + (C_C^B)' C(\hat{q})(E'_{12}\bar{r}_{foot,e,m,k} - r_{e,k}) - \eta_{foot,k}. \end{aligned} \quad (46)$$

Inserting (46) into (45), we obtain

$$z_{foot,m,k} = \bar{H}_{foot,m,k} \bar{X}_{m,k} - \eta_{foot,k} \quad (47)$$

where

$$\bar{H}_{foot,m,k} = \begin{bmatrix} H_{foot1,m,k} & 0_{3\times 18+2(m-1)} & H_{foot2,m,k} \end{bmatrix} \quad (48)$$

$$\begin{aligned} H_{foot1,m,k} = \big[&2(C_C^B)'[(C(\hat{q})(\hat{r}_{foot,m,k} - \hat{r}))\times], \\ &-(C_C^B)' C(\hat{q}), \quad 0_{3\times 3} \big] \in R^{3\times 9}. \end{aligned}$$

$$H_{foot2,m,k} = (C_C^B)' C(\hat{q}) E'_{12} \in R^{3\times 2}.$$

When a state is extended, H_{tag} in Section V-A should also be extended to be compatible with the extended dimension by filling zeros.

VI. Filtering and Smoothing Algorithm

In this section, the results in Section III and V are combined. A Kalman filter is used to estimate q, r, v and r_{foot}, where $\hat{q}_{filter}$, $\hat{r}_{filter}$, $\hat{v}_{filter}$ and $\hat{r}_{filter,foot}$ are estimated values. Then based on the estimated value, a smoother algorithm is applied to obtain more accurate estimation, where $\hat{q}_{smoother}$, $\hat{r}_{smoother}$, $\hat{v}_{smoother}$ and $\hat{r}_{smoother,foot}$ are estimated values.

The filtering algorithm is fairly standard and is summarized in Algorithm 1.

```
initialization;
while k ≤ N_img do
    time update (10);
    measurement update using visual odometry (21);
    if foot is detected then
        if first observed then
            extend state (38);
        else
            measurement update using the foot (47);
        end
    end
end
```

Algorithm 1: Filtering algorithm

The smoothing problem is formulated as a linear optimization problem, where small errors in the filter estimated values are computed. For example, the true position $r_k \in R^3$ can be modeled as

$$r_k = \hat{r}_{filter,k} + r_{e,k}. \quad (49)$$

In the smoothing algorithm, $r_{e,k}$ is estimated, where $\hat{r}_{e,k}$ is the estimated value. Then the position is given by

$$\hat{r}_{smoother,k} = \hat{r}_{filter,k} + \hat{r}_{e,k}. \quad (50)$$

Let the state of the smoother $X_{smoother}$ be defined by

$$X_{smoother} = \begin{bmatrix} U_1 \\ \vdots \\ U_{N_{image}} \\ b_{g,e} \\ b_{a,e} \\ r_{foot,e,1} \\ \vdots \\ r_{foot,e,M_{foot}} \end{bmatrix} \in R^{9N_{image}+6+2M_{foot}}$$

where M_{foot} is the total number of detected stance periods (both left and right feet).

The smoothing algorithm can be formulated as a linear quadratic optimization problem:

$$\begin{aligned}
J(X) &= \textstyle\sum_{k=1}^{N_{image}-1} \text{quad}(\bar{\zeta}_{k,1}, Q_{\bar{\zeta}_k,11}) \\
&+ \textstyle\sum_{k=2}^{N_{image}} \text{quad}(z_{vo,k} - H_{vo,k}(1{:}6,1{:}18) \begin{bmatrix} U_k \\ U_{k-1} \end{bmatrix}, R_{vo,k}) \\
&+ \textstyle\sum^{k \in S_{tag}} \text{quad}(z_{tag,k} - H_{tag,k}(1{:}3,1{:}9)U_k, R_{tag,k}(1{:}3,1{:}3)) \\
&+ \textstyle\sum_{k \in S_{foot}} \text{quad}(z_{foot,s_{index}(k),k} - H_{foot1,s_{index}(k),k} X_k \\
&\qquad - H_{foot2,s_{index}(k),k}\, r_{foot,e,s_{index}(k),k}, R_{foot,k}) \\
&+ \text{quad}(\hat{b} + b_{g,e} - b_{g,init}, P_{b_{g,init}}) \\
&+ \text{quad}(\hat{b} + b_{a,e} - b_{a,init}, P_{b_{a,init}}) \\
&+ \text{quad}(U_{init} - U_1, P_{U,init}) \\
&+ \text{quad}(U_{final} - U_{N_{image}}, P_{U,final})
\end{aligned}$$

where quad is defined by

$$\text{quad}(a, B) = \frac{1}{2} a' B a.$$

The initial and final constraints U_{init} and U_{final} are given by

$$U_{init} = \begin{bmatrix} f_{vector}(\hat{q}_1 \otimes \hat{q}_{tag,1}) \\ \hat{r}_{tag,1} - \hat{r}_1 \\ 0_{3\times1} - \hat{v}_1 \end{bmatrix},$$

$$U_{final} = \begin{bmatrix} f_{vector}(\hat{q}_{f_{discrete}(N_{image})} \otimes \hat{q}_{tag,N_{image}}) \\ \hat{r}_{tag,N_{image}} - \hat{r}_{f_{discrete}(N_{image})} \\ 0_{3\times1} - \hat{v}_{f_{discrete}(N_{image})} \end{bmatrix}.$$

From (8), $\bar{\zeta}_{k,1}$ in the quadratic term quad$(\bar{\zeta}_{k,1}, Q_{\bar{\zeta}_k,11})$ is given by

$$\bar{\zeta}_{k,1} = U_{k+1} - \bar{F}_{11,k} U_k - \bar{F}_{12,k} \begin{bmatrix} b_{g,e} \\ b_{a,e} \end{bmatrix}.$$

Once the optimal X is computed, $\hat{q}_{smoother}$, $\hat{r}_{smoother}$, $\hat{v}_{smoother}$ and $\hat{r}_{smoother,foot}$ can be computed (see (50)).

VII. Experimental Setup and Results

A. Experiment

Experiments are performed to verify the proposed algorithm. Camera frame is attached to user's waist by velcro tapes and buckles clips, and two reflective markers (for the motion capture system) are mounted on top of both feet (see Fig. 7). A motion capture system is set up with 6 cameras from Optitrack, creating a tracking space of about 5m long in Fig. 8.

Five healthy persons are recruited to perform the data acquisition. Subjects information is given in Table I.

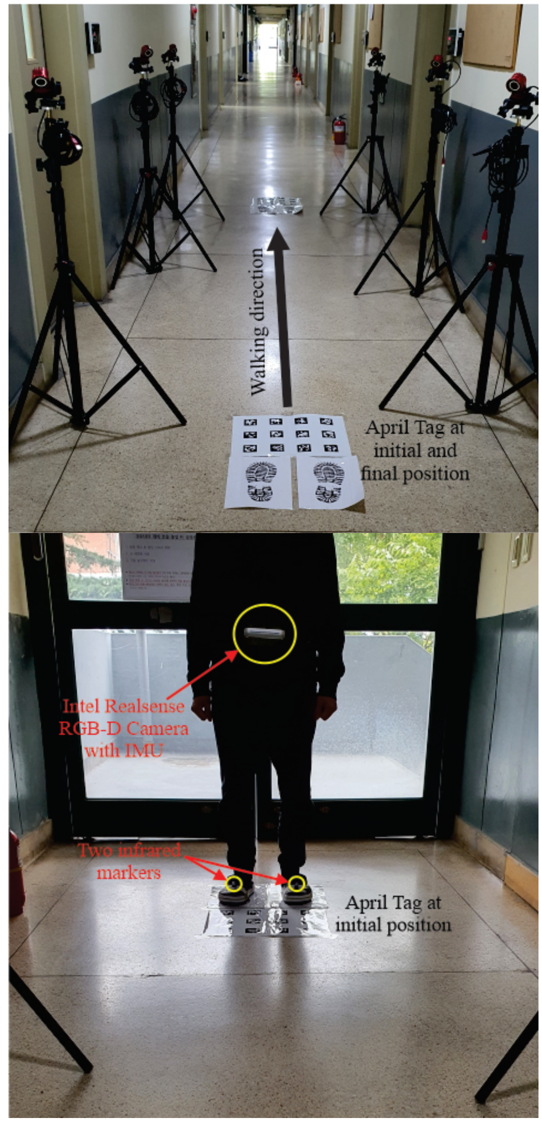


Fig. 7. Participant with equipped camera system doing the experiments.

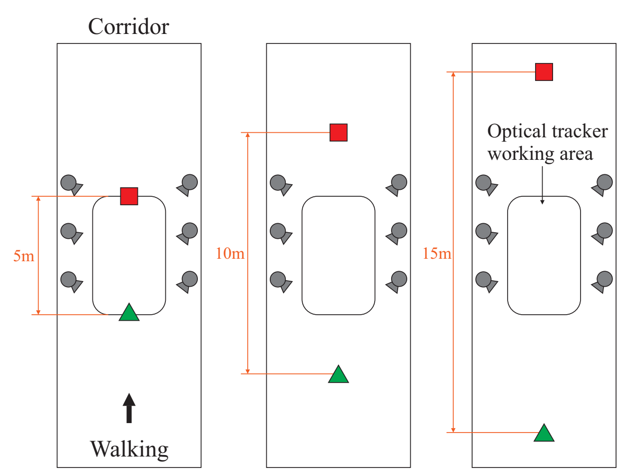


Fig. 8. Experiment setup, volunteer walks three straight paths from green triangle to red square through the working range of the optical tracker system.

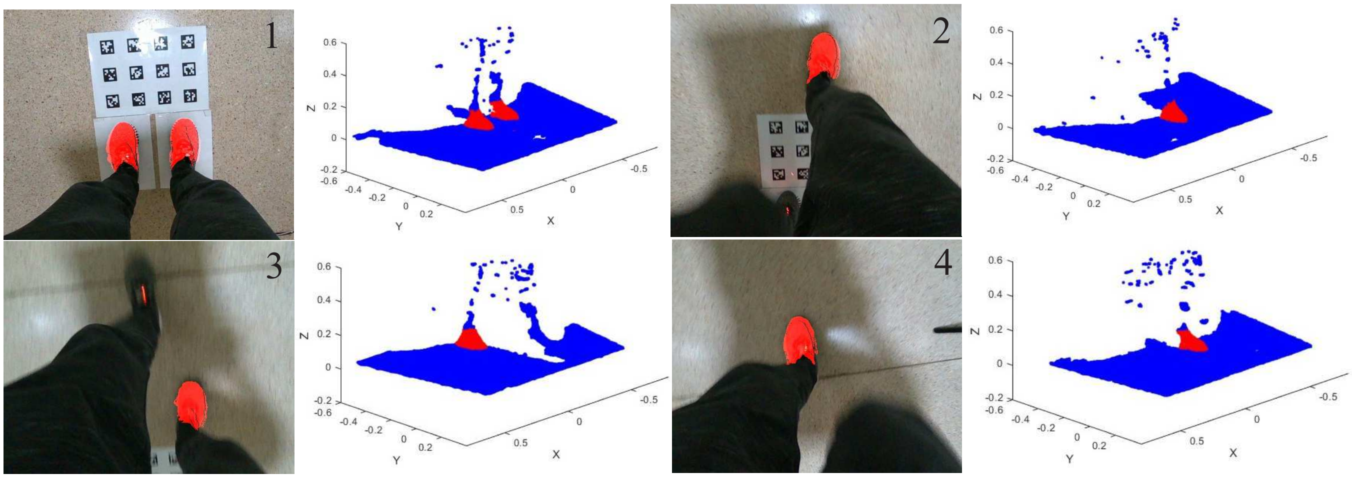


Fig. 9. Detected stance foot from point cloud data and masked in RGB image: 1) initial states; 2) first detected right foot; 3) already detected right foot, used as landmark; 4) first detected left foot.

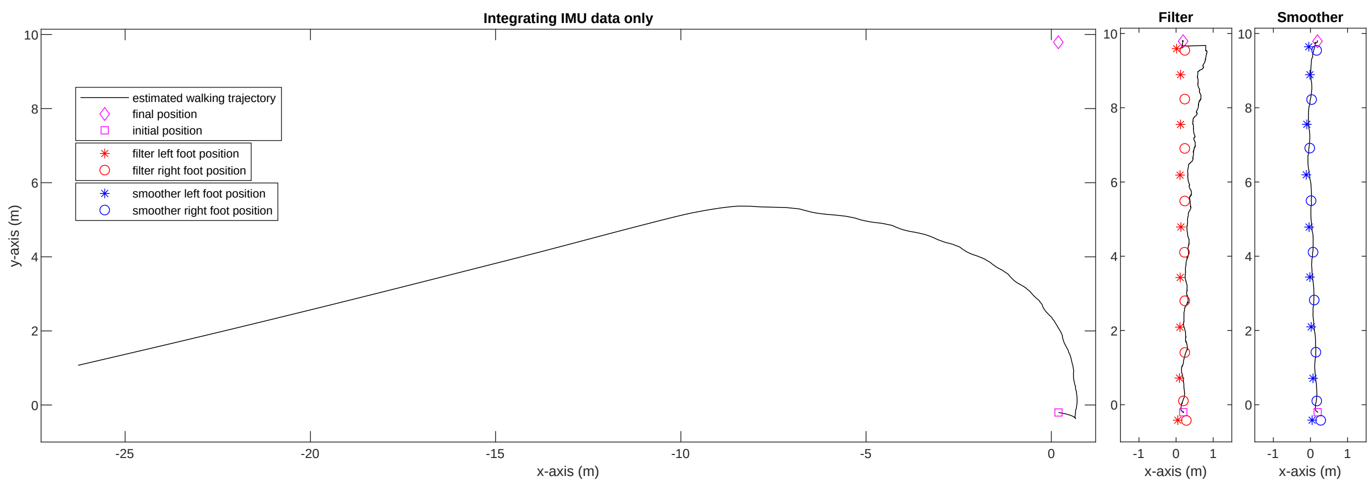


Fig. 10. Estimated walking trajectory using integrating internal IMU data only, proposed filter and smoother.

TABLE I
FIVE SUBJECTS INFORMATION

Volunteer	Age	Weight(kg)	Height(cm)
Range	26-35	55-75	161-185
Mean	30.2	62.2	170.2
Standard deviation	3.27	7.60	6.05

In the experiment, volunteers are asked to walk in a straight line inside a long corridor. Each person walks in separated paths with walking distance of 5, 10, and 15 meters, five times each. At the starting and ending point of each walking path, two identical printed Apriltag of A3 size are placed with the same orientation. All walking paths are designed so that the tracking range is at the middle of each path as in Fig. 8.

B. Results

Fig. 9 shows detected stance foot from point cloud depth data, where foot candidate points a selected using a simple threshold constraint of the height from the floor. Candidate points are grouped to defined foot edge. An ellipse approximation algorithm is used to fit the detected foot edge. Foot position in the camera coordinate is then calculated from ellipse position in the RGB images.

Fig. 10 shows a 10-meter-walking example of an estimated walking trajectory integrating the waist-mounted internal IMU data only without any measurement updating. It is clear that just integrating IMU data method leads to divergence.

From the proposed filter in Algorithm 1 and the smoothing algorithm, estimated walking trajectories are given in the right-hand side of Fig. 10. Left and right foot positions in world coordinate system are computed. The relative position, relative attitude from visual odometry and foot position as landmarks are used in the measurement updating equation. The estimated walking trajectory is improved and corrected in the smoothing algorithm. Note that the initial foot position is slightly negative since it is aligned with the origin of the calibrated optical tracker system.

To evaluate the accuracy, foot positions are compared with ground truth from optical tracker system in Fig. 11. In each trial, three strides in 5-meter working space of optical tracking system are selected to provide ground truth data. In total,

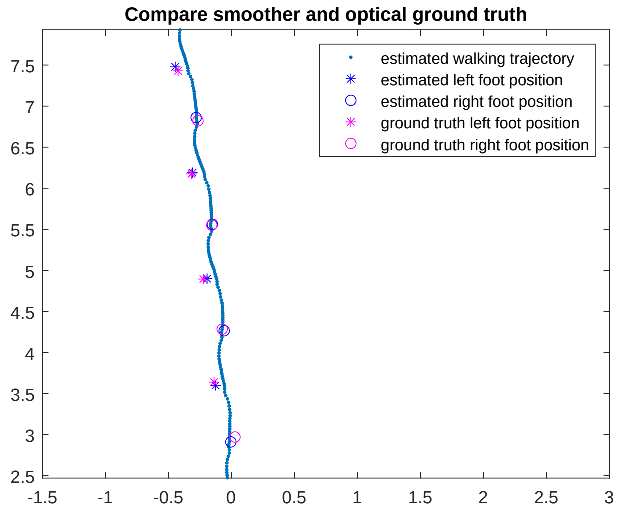


Fig. 11. Compare estimated foot position with ground truth.

there are 5 (person) x 3 (path) x 5 (trial) x 3 (stride) x 2 (left,right) = 450 strides data. Since the difference in walking paths might cause different estimated results, the number of selected strides for calculating stride length is the same for the whole experiment, and average stride length error over the walking paths for each volunteer is computed and given in Table II. The overall root mean square errors (RMSE) is about 3.8 centimeters.

TABLE II
ESTIMATED STRIDE LENGTH ERROR (UNIT: METER).

	Left			Right		
ID	**Max. error**	**RMSE**	**Std.**	**Max. error**	**RMSE**	**Std.**
1	0.043	0.033	0.017	0.039	0.033	0.014
2	0.049	0.027	0.019	0.037	0.026	0.012
3	0.075	0.067	0.031	0.056	0.047	0.029
4	0.050	0.030	0.025	0.073	0.059	0.024
5	0.057	0.036	0.024	0.046	0.020	0.017

VIII. Conclusion

The paper proposed a filter and smoother algorithm to fuse IMU data with RGB-D camera data to estimate stance foot positions and reconstruct walking trajectory. With the IMU-Camera system attached to user waist, walking range is unrestricted. Therefore, longer walking range can be achieved in comparison with pressure mat or optical tracker system. Visual odometry algorithm provides incremental attitude and position of the camera alongside with position of the detected stance foot as a landmark in the measurement update equations of the filter. However, the proposed foot detection algorithm is restricted to only stance foot due to obscuration in the swing phases. In future works, we intend to improve the foot detection algorithm to be able to detect foot in every gait phases, and integrate it with the model of human body to reconstruct shape and pose, which can be used in gait analysis research.

References

[1] M. W. Whittle, "Clinical gait analysis: A review," *Human Movement Science*, vol. 15, no. 3, pp. 369–387, 1996.

[2] C. Prakash, R. Kumar, and N. Mittal, "Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges," *Artif Intell Rev*, no. 49, pp. 1–40, 2018.

[3] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. Hausdorff, "Effect of gait speed on gait rhythmicity in parkinson's disease: variability of stride time and swing time respond differently," *J Neuroeng Rehabil*, vol. 2, no. 23, 2005.

[4] F. Jasni, N. A. Hamzaid, T. Y. Al-Nusairi, N. H. Mohd Yusof, H. N. Shasmin, and S.-C. Ng, "Feasibility of a gait phase identification tool for transfemoral amputees using piezoelectric- based in-socket sensory system," *IEEE Sensors Journal*, vol. 19, no. 15, pp. 6437–6444, 2019.

[5] J. S. Park, C. M. Lee, S.-M. Koo, and C. H. Kim, "Gait phase detection using force sensing resistors," *IEEE Sensors Journal*, vol. 20, no. 12, pp. 6516–6523, 2020.

[6] D. Sutherland, "The evolution of clinical gait analysis: Part ii kinematics," *Gait Posture*, vol. 16, no. 2, pp. 159–179, 2002.

[7] A. Mannini and A. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors (Basel)*, vol. 10, no. 2, pp. 1154–1175, 2010.

[8] R. Caldas, M. Mundt, W. Potthast, F. Buarque de Lima Neto, and B. Markert, "A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms," *Gait Posture*, vol. 57, pp. 204–210, 2017.

[9] P. Federica, S. Luca, P. Giuseppe, B. Luigi, P. Robert, C. Cosimo, and B. Luigi, "Inertial sensors versus standard systems in gait analysis: a systematic review and meta-analysis," *European Journal Of Physical And Rehabilitation Medicine*, vol. 55, no. 2, pp. 268–280, 2019.

[10] M. Kadaba, H. Ramakrishnan, and M. Wootten, "Measurement of lower extremity kinematics during level walking," *J Orthop Res.*, vol. 8, no. 3, pp. 383–392, 1990.

[11] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[12] J. Lu, G. Wang, and P. Moulin, "Human identity and gender recognition from gait sequences with arbitrary walking directions," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, pp. 51–61, 2014.

[13] X. Gu, F. Deligianni, B. Lo, W. Chen, and G. Yang, "Markerless gait analysis based on a single rgb camera," in *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Nevada, USA, 2018, pp. 42–45.

[14] A. Rocha, H. Choupina, M. Vilas-Boas, J. Fernandes, and J. Cunha, "System for automatic gait analysis based on a single rgb-d camera," *PLoS One*, vol. 13, no. 8, 2018.

[15] A. Dubois and F. Charpillet, "A gait analysis method based on a depth camera for fall prevention," in *Annu Int Conf IEEE Eng Med Biol Soc*, 2014, pp. 4515–4518.

[16] E. Auvinet, J. Meunier, and F. Multon, "Multiple depth cameras calibration and body volume reconstruction for gait analysis," in *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, 2012, pp. 478–483.

[17] C. Li, L. Yu, and S. Fei, "Real-time 3d motion tracking and reconstruction system using camera and imu sensors," *IEEE Sensors Journal*, vol. 19, no. 15, pp. 6460–6466, 2019.

[18] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, 2004.

[19] K. Yousif, A. Bab-Hadiashar, and R. Hoseinnezhad, "An overview to visual odometry and visual slam: Applications to mobile robotics." *Intell Ind Syst*, no. 1, pp. 289–311, 2015.

[20] F. L. Markley and J. L. Crassidis, *Fundamentals of Spacecraft Attitude Determination and Control*. New York: Springer, 2014.

[21] C. Creamer, "Spacecraft attitude determination using gyros and quaternion measurements," *The Journal of the Astronautical Sciences*, vol. 44, no. 3, pp. 357–371, 1996.

[22] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3565–3572.

[23] G. Huang, "Visual-inertial navigation: A concise review," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 9572–9582.

[24] C. Kerl, J. Sturm, and D. Cremers, "Robust odometry estimation for rgb-d cameras," in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 3748–3754.
[25] L. Yang, Y. Li, X. Li, Z. Meng, and H. Luo, "Efficient plane extraction using normal estimation and ransac from 3d point cloud," *Computer Standards & Interfaces*, vol. 82, p. 103608, 2022.
[26] I. Selesnick and C. Burrus, "Generalized digital butterworth filter design," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1688–1694, 1998.
[27] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 3400–3407.
[28] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
[29] J. Sola, "Simultaneous localization and mapping with the extended kalman filter," 2014, last accessed 2021-10-29. [Online]. Available: http://www.iri.upc.edu/people/jsola/JoanSola/objectes/curs_SLAM/SLAM2D/SLAM%20course.pdf

Duc Cong Dang was born in An Thi Town, Hung Yen, Vietnam in 1992. He received the B.S. degrees in automatic control from Hanoi University of Science and Technology, Hanoi, in 2015 and the M.S. degree in electrical engineering from University of Ulsan, South Korea, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering at University of Ulsan, South Korea. His research interests include pedestrian navigation, gait analysis, sensor fusion, wireless communication, and automatic control.

Young Soo Suh (Member, IEEE) received the B.S. and M.S. degrees from the Department of Control and Instrumentation Engineering, Seoul National University, Seoul, South Korea, in 1990 and 1992, and the Ph.D. degree from the Department of Mathematical Engineering and Information Physics, The University of Tokyo, Tokyo, Japan, in 1997. He is currently a Professor with the Department of Electrical Engineering, University of Ulsan, Ulsan, South Korea. His current research interests include networked control systems and attitude estimation and control and personal navigation systems.