2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Enhancing Prosthetic Safety and Environmental Adaptability: A Visual-Inertial Prosthesis Motion Estimation Approach on Uneven Terrains

Chuheng Chen, Xinxing Chen, Shucong Yin, Yuxuan Wang, Binxin Huang, Yuquan Leng and Chenglong Fu

***Abstract*—Environment awareness is crucial for enhancing walking safety and stability of amputee wearing powered prosthesis when crossing uneven terrains such as stairs and obstacles. However, existing environmental perception systems for prosthesis only provide terrain types and corresponding parameters, which fail to prevent potential collisions when crossing uneven terrains and may lead to falls and other severe consequences. In this paper, a visual-inertial motion estimation approach is proposed for prosthesis to perceive its movement and the changes of spatial relationship between the prosthesis and uneven terrain when traversing them. To achieve this, we estimate the knee motion by utilizing a depth camera to perceive the environment and align feature points extracted from uneven terrains. Subsequently, an error-state Kalman filter is incorporated to fuse the inertial data into visual estimations to obtain a more robust and accurate estimation, which is then utilized to derive the motion of the whole prosthesis for our prosthetic control scheme. Experiments conducted on our collected dataset and stair walking trials with powered prosthesis show that the proposed method can accurately track the motion of human leg and the prosthesis with the average root-mean-square error of toe trajectory less than 5 cm. The proposed method is expected to enable the environmental adaptive control for prosthesis, thereby enhancing amputee's safety and mobility in uneven terrains.**

I. Introduction

The emergence of powered prosthesis has contributed greatly to the restoration of mobility for millions of amputees [1]-[3]. However, significant challenges remain in achieving safe and stable locomotion with prosthesis on uneven terrains such as stairs and obstacles [4]. Inspired by humans' use of vision to guide their non-rhythmic locomotion in complex terrains [5], environmental information captured by vision sensors has been integrated in prosthetic control and shown potential in enhancing mobility across various terrains [6], [7].

Researchers have explored the use of different vision sensors in conjunction with classification algorithms to accomplish terrain classification (e.g., stair, ramp and level ground) and the estimation of corresponding parameters, such as stair height and width [8]-[11]. This information provides valuable environmental context that facilitates the seamless transition between different locomotion modes of powered prosthesis [12]. However, providing only terrain types and parameters cannot fully utilize the environmental information gathered from vision sensors. The inconsistency between amputees' real locomotion modes and the classification results when amputees adjust their strides before transition, as well as the limited awareness of prosthesis's movement and its spatial relationship with uneven elements of terrains, may lead to collisions and falls [13]. In addition, the estimation error of terrain parameters caused by the absence of complete terrain information (e.g., a complete step of stairs), and inconsistencies in the parameters between different steps may also reduce the foot clearance and pose risks for amputee's stair and obstacle negotiation [14]. Therefore, navigating complex uneven terrains still pose risks for amputees.

To enhance safe and predictive walking on uneven and unstructured terrains, our goal is to utilize environmental information to estimate the movement of the prosthesis within the environment and integrate it into our prosthetic control scheme. While visual odometry and SLAM are commonly used for motion estimation in mobile robots [15], [16], there is still a dearth of research focusing on motion estimation for prosthesis within the environment. Although radar has been used to estimate the distance from stairs and the motion of prosthesis, the motion estimation is only performed during the stance phase [17]. The swing phase with a more violent motion and higher risk of collisions is not taken into account. Besides, researchers have developed obstacle location and tracking system or exoskeleton to adaptively plan step length for obstacle crossing [18]. To obtain a more stable and accurate estimation, the system is mounted on user's waist. However, different from the exoskeleton [18] and biped robot [19], the prosthesis is not rigidly attached to the amputee's upper body, and its motion depends on the motion of amputee's residual limb, making it difficult to derive the prosthesis motion from the motion of user's upper body. As a result, the motion estimation systems mounted on waist or head for exoskeletons or mobile robots may not be suitable for prosthesis, and we choose to directly mount the vision sensor on the prosthesis.

This work was supported by the National Natural Science Foundation of China [Grant U1913205, 62103180]; Guangdong Innovative and Entrepreneurial Research Team Program [Grant 2016ZT06G587]; the Stable Support Plan Program of Shenzhen Natural Science Fund [Grant 20200925174640002]; the Science, Technology and Innovation Commission of Shenzhen Municipality [Grant SGLH20180619172011638, Grant ZDSYS20200811143601004, Grant JCYJ2023080709340701, Grant KYTDPT20181011104007 and Grant JCYJ20230807093407016]; high level of special funds (G03034K003) from Southern University of Science and Technology, Shenzhen, China; and Centers for Mechanical Engineering Research and Education at MIT and SUSTech. (*Corresponding author: Chenglong Fu.*)

The authors are all with Shenzhen Key Laboratory of Biomimetic Robotics and Intelligent Systems, Shenzhen, 518055, China, and also with Guangdong Provincial Key Laboratory of Human-Augmentation and Rehabilitation Robotics in Universities, Southern University of Science and Technology, Shenzhen, 518055, China (email: fucl@sustech.edu.cn and chenxx@sustech.edu.cn).

Nevertheless, mounting the sensors directly on prosthesis leads to challenges like limited camera field of view, reduced texture information, and rapid view angle changes. In such situations, feature-based methods like ORB-SLAM [20] may struggle to extract sufficient feature points from grayscale images, leading to tracking loss. Considering the significance of geometric features of the environment for walking with prosthesis, using the environmental point cloud captured by depth camera for motion estimation may be a better choice. However, the real-time integration with prosthetic control necessitates swift estimation calculations, making 3D point cloud-based methods impractical due to their computational demands [21]. Therefore, developing a lightweight motion estimation method suitable for motion characteristics of prosthesis and integrate it with control scheme is significant.

To address these limitations and enhance environmental adaptative walking of the prosthesis, a visual-inertial motion estimation approach is proposed and combined with our multi-objective optimization-based swing control scheme prototype in this paper. The proposed approach utilizes the environmental point cloud captured by a depth camera and its inertial information acquired with an inertial measurement unit (IMU) for motion estimation of the prosthesis in the sagittal plane. The sagittal plane's environmental structure significantly impacts users' gait and prosthetic control, making it the central focus of our study [22]. To that end, feature points are extracted from 2D environmental point cloud of stairs and obstacles and then used for an iterative closest point-based (ICP) alignment approach to obtain an initial camera motion estimation. Moreover, to improve the robustness and accuracy of estimation, a sagittal plane motion-oriented error-state Kalman filter (SP-ESKF) is incorporated to fuse the visual and inertial estimation of motion in the sagittal plane. The camera motion is combined with the prosthesis model to derive the movement of each joint and prosthetic foot, which is then utilized in our control scheme. The multi-objective optimization-based swing control utilizes time-serial estimated motion to calculate a multi-step-look-forward cost function to allow the prosthesis to plan desired knee and ankle angles without collision.

The primary contributions of the present paper include:

1) Developing a visual-inertial motion estimation method for prosthesis in uneven terrains to estimate its movement and position within the environment, improving environmental awareness and self-awareness of the prosthesis.

2) Proposing a sagittal plane motion-oriented error-state Kalman filter with a reduced parameter set, facilitating real-time environmental adaptive prosthesis control.

3) Integrating the prosthesis motion estimation into our optimization-based prosthetic control scheme, enhancing the environmental adaptability and walking safety for amputee.

The rest of this paper is described as follows. Section II introduces the proposed visual-inertial motion estimation approach for prosthesis. The experiment setup and results are presented in Section III and discussed in Section IV. Section V is the conclusion of this paper.

II. METHODS

The proposed visual-inertial prosthesis motion estimation approach is introduced in this section. An overview of this approach is presented in Fig. 1, which consists of three main components: visual estimation, inertial navigation and information fusion for motion estimation. First, the preprocessed environmental point cloud captured by a depth camera fixed on prosthetic knee is utilized to initially estimate the knee motion through frame-to-frame alignment with ICP algorithm. The visual estimation result is then fused with inertial estimation through SP-ESKF and the position of the whole prosthesis is derived. Finally, the integration of estimation results and the control scheme is briefly described.

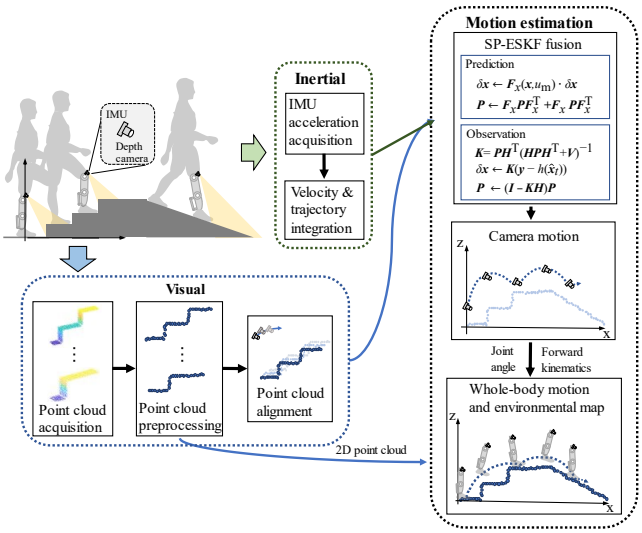


Fig. 1. The overview of the proposed visual-inertial prosthesis motion estimation approach. The visual estimation of camera movement using ICP algorithm and the inertial estimation are fused with SP-ESKF to obtain the final estimation result. The estimation results are utilized to derive the motion of the whole prosthesis with joint angle and forward kinematics.

A. Visual Estimation Based on ICP

1) Environmental Data Preprocessing

The 3D environmental point cloud in front of the user is acquired with a depth camera (CamBoard pico flexx2, pmdtechnologies, Germany) fixed on the prosthetic knee. The relative motion of the camera and the prosthesis in the environment is consistent due to their solid connection. Before alignment, preprocessing of the point cloud is required, which consists of two parts: sagittal plane projection and coordinate system conversion.

The point cloud directly captured by the depth camera contains 3D coordinate information of about 38k points per frame. The alignment of these points imposes a significant computational burden, presenting difficulties in its integration with real-time control of the prosthesis. As stated above, this study mainly focuses on the environmental information and the prosthesis motion in sagittal plane. To reduce the computational effort, the 3D point cloud is projected into sagittal plane for dimension reduction as:

$$\begin{cases} \boldsymbol{C}_{3D} = \{(x_i, y_i, z_i) \mid i = 1,...,n\} \\ \boldsymbol{J} = \{i \mid -0.05\text{m} < y_i < 0.05m\}, \\ \boldsymbol{C}_{2D} = \{(x_j, z_j) \mid j \in \boldsymbol{J}\} \end{cases} \quad (1)$$

where $\boldsymbol{C}_{3D}$ and $\boldsymbol{C}_{2D}$ are original 3D environmental point cloud and point cloud in the sagittal plane respectively. x_i, y_i, and z_i are the coordinate values of the point. $\boldsymbol{J}$ is the index set

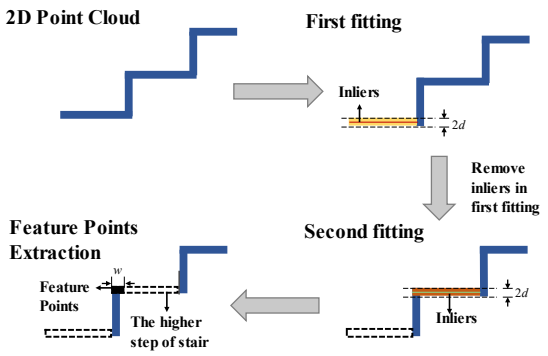


Fig. 2. The feature points extraction proscess. The yellow and brown lines are inliers fitted in the first and second fitting. The black short line incorporates the extracted feature points.

of points extracted from the 3D point cloud directly in front of the prosthesis. With sagittal plane projection, the original 3D point cloud is converted into 2D point cloud in the sagittal plane, and meanwhile, the camera motion can be considered as motion in the sagittal plane.

In addition, considering that our main concern is to utilize the position change of camera to derive the prosthesis motion, the orientation information (Euler angle) of camera is utilized to convert the coordinate system of point cloud from the camera system to the invariant ground system for camera rotation compensation. The camera orientation is acquired using an IMU (MTi 1-series, Xsens, Netherlands) fixed with the camera. Through the coordinate system conversion, the camera rotation is compensated, and its motion can be treated as translational motion. The conversion is described as:

$$\boldsymbol{C}_{\text{Gnd}} = \boldsymbol{R}_{\text{Cam}}^{\text{Gnd}} \boldsymbol{C}_{\text{Cam}}, \tag{2}$$

where $\boldsymbol{C}_{\text{Gnd}}$ and $\boldsymbol{C}_{\text{Cam}}$ denote the point cloud of the ground and camera coordinate systems respectively. $\boldsymbol{R}_{\text{Cam}}^{\text{Gnd}}$ describes the rotation matrix from the camera coordinate system to the ground coordinate system, which is calculated with the Euler angle of IMU. Besides, to reduce the impact of noise due to camera performance or ground material on the accuracy of alignment, the point cloud is finally smoothed through the k-Nearest Neighbor algorithm.

2) Feature Point Extraction

After pre-processing, the original 3D environmental point cloud is converted to 2D point cloud in the ground coordinate system, and the camera motion can be simplified to translational motion in sagittal plane. However, alignment of the entire 2D point cloud containing redundant points may lead to long alignment time and increased errors. Therefore, researchers have extracted feature points in the point cloud before alignment [23], [24]. Since the aim of this study is to help the predictive walking in uneven terrain such as stairs and obstacles, considering the characteristics of these terrains, the corner points are manually extracted using the random sample consensus (RANSAC) algorithm and are used for subsequent frame-to-frame alignment.

Before feature point extraction, the terrain in front of the user is recognized using the CNN-based environmental classification method proposed in our previous work [9]. If the terrain is recognized as stair or obstacle, the RANSAC algorithm is used to extract two median lines from the 2D point cloud of the terrain.

As shown in Fig. 2, a horizontal line with the most inliers is fitted first. After removing the inliers of the first line, another horizontal line with the most inliers is fitted. The median z coordinates of the two lines are compared and the feature points are extracted from the end of the line with larger z-coordinate (the upper surface of obstacle or the upper level of stair), where d and w are manually set thresholds of 0.02 m and 0.05 m, respectively.

3) Frame-to-frame Alignment with ICP

The movement of camera can be estimated by frame-to-frame alignment of the feature points extracted from 2D environmental point cloud. After extraction, if difference between the mean coordinate values of the feature points extracted at two consecutive moments t_i and t_{i-1} is within a preset threshold, the feature points of two frames are considered to be the same corner of the terrain. Then, the alignment is realized using ICP algorithm [25] and the displacement T between the point cloud in the user's sagittal plane is estimated. The ICP workflow is described as follows:

1) Given a source point cloud $\boldsymbol{C}_{\text{s}}$, a target point cloud $\boldsymbol{C}_{\text{t}}$ and an initial prediction T_0 of the transformation between $\boldsymbol{C}_{\text{s}}$ and $\boldsymbol{C}_{\text{t}}$, for each point in $\boldsymbol{C}_{\text{t}}$ ($C_i \in \boldsymbol{C}_{\text{t}}$), corresponding point that has the closest distance to it in the initially transformed source point cloud is searched:

$$C_i' = C_j, j = \arg\min_j \left\| C_i - (T + C_j) \right\|, C_j \in \boldsymbol{C}_s. \tag{3}$$

2) The displacement between source and target point clouds is estimated using corresponding point pairs to minimize the error, and then injected into the displacement estimated in the previous step:

$$\begin{cases} T^{n+1} = \triangle T + T^n \\ \triangle T = \arg\min_{\triangle T} \left\| \boldsymbol{C} - (\triangle T + \boldsymbol{C}') \right\| \end{cases}. \tag{4}$$

3) The above steps are iterated until convergence ($\Delta T \leq T_{\text{th}}$, T_{th} is taken as 10^{-6}) or the maximum number of iterations (taken as 20) is reached to obtain the final displacement.

B. IMU Measurement Model

The initial motion estimation of camera and prosthetic knee is achieved by frame-to-frame alignment. However, tracking errors caused by incorrect extraction of feature points or inability to extract feature points still exist. To solve this problem, the camera's inertial information measured by IMU can be used to estimate the motion state of camera.

IMU is capable of measuring three-axis acceleration of the camera, which is affected by Gaussian white noise and zero bias. Since the camera motion has been treated as translational motion through rotation compensation, the motion equations regarding angular velocity are not considered in this study. Therefore, the acceleration of IMU can be described by the following model:

$$\begin{cases} \boldsymbol{a}_m(t) = \boldsymbol{R}^{-1}\left(\boldsymbol{a}_w(t) - \boldsymbol{g}\right) + \boldsymbol{a}_b(t) + \boldsymbol{a}_n(t) \\ \boldsymbol{a}_n \sim N(0, \boldsymbol{\sigma}_a^2) \\ \dot{\boldsymbol{a}}_b = \boldsymbol{a}_{bn} \sim N(0, \boldsymbol{\sigma}_{a_b}^2) \end{cases}, \tag{5}$$

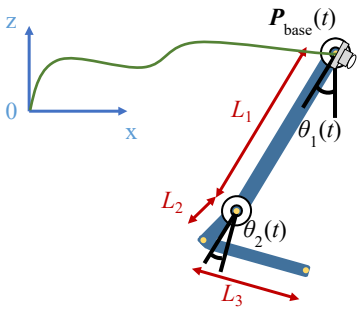


Fig. 3. The link model of the prosthesis. $\boldsymbol{P}_{\text{base}}$ denotes the estimated knee position. θ_1 and θ_2 are calculated with the joint angle measured by two encoders fixed on the knee and ankle of the prosthesis.

where $\boldsymbol{a}_m$ and $\boldsymbol{a}_w$ represent the measured acceleration and the acceleration in the world coordinate system; $\boldsymbol{a}_n$ and $\boldsymbol{a}_b$ denote the accelerometer noise and bias, with the noise following Gaussian distribution and the bias's derivatives following Gaussian distribution. $\boldsymbol{R}$ is the rotation matrix from the IMU coordinate system to the ground coordinate system in SO(2), which is calculated with IMU Euler angle. Besides, $\boldsymbol{a}_n$ and $\boldsymbol{a}_b$ are given by the IMU vendor, Xsens Technologies.

By integrating the acceleration, velocity and displacement of the IMU can be obtained. However, due to the noise and bias, direct integration can lead to error accumulation. To avoid the error increasing with time, it is also necessary to fuse the visual information to rectify the integration results.

C. Visual-inertial Fusion Based on SP-ESKF

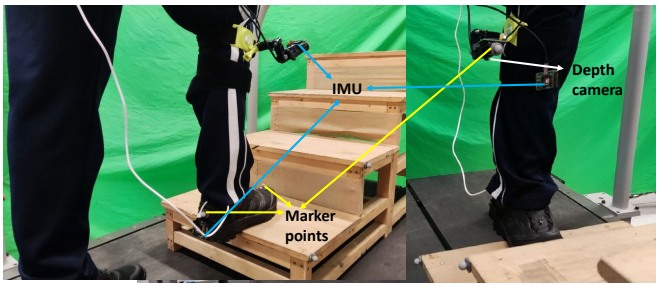
Error state Kalman filter combines the efficiency of Kalman filter with the flexibility of error state estimation and has been used in navigation and positioning systems [21]. Given our primary focus on motion estimation in the sagittal plane, and inspired by [26], we propose a sagittal plane motion-oriented error state Kalman filter with a reduced parameter set. The SP-ESKF combines the visual with the inertial estimation results to obtain a more robust and accurate knee motion estimation in the sagittal plane. The total system state of motion is described by the true state $\boldsymbol{x}_t = (\boldsymbol{p}_t, \boldsymbol{v}_t, \boldsymbol{a}_{bt}, \boldsymbol{g}_t)$, the nominal state $\boldsymbol{x} = (\boldsymbol{p}, \boldsymbol{v}, \boldsymbol{a}_b, \boldsymbol{g})$ and the error state $\delta\boldsymbol{x} = (\delta\boldsymbol{p}, \delta\boldsymbol{v}, \delta\boldsymbol{a}_b, \delta\boldsymbol{g})$.

The workflow can be divided into two steps: 1) updating the nominal state with IMU measurements, and updating the error state and covariance matrix according to the error state processing model; 2) calculating the Kalman gain K and rectifying the error state using the visual estimation results, and finally injecting the error state into the nominal state to obtain the true state of the system.

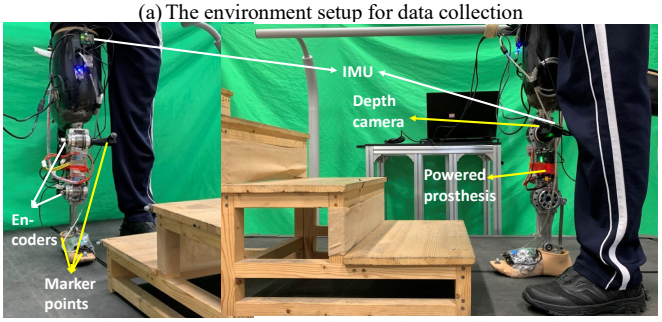
1) Updating Process of Nominal State and Error State

When IMU data is fed in, the nominal state is computed to predict idealized knee motion without noise. Meanwhile, the error state and covariance matrix will be updated. Since the system error is concentrated into the error state, the updating process of the nominal state can be described as:

$$\begin{cases} \boldsymbol{p} \leftarrow \boldsymbol{p} + \boldsymbol{v}\Delta t + \frac{1}{2}(\boldsymbol{R}(\boldsymbol{a}_m - \boldsymbol{a}_b) + \boldsymbol{g})\Delta t^2) \\ \boldsymbol{v} \leftarrow \boldsymbol{v} + (\boldsymbol{R}(\boldsymbol{a}_m - \boldsymbol{a}_b) + \boldsymbol{g})\Delta t \\ \boldsymbol{a}_b \leftarrow \boldsymbol{a}_b \\ \boldsymbol{g} \leftarrow \boldsymbol{g} \end{cases} . \quad (6)$$



(a) The environment setup for data collection



(b) The environment setup for prosthesis walking trails

Fig. 4. The experimental environment of our collected dataset and prosthesis walking. In our collected dataset of walking in different terrains, the depth camera and IMU are mounted on the shank of the subject. In prosthesis walking experiment, the depth camera and IMU are mounted on the knee of the prosthesis.

The updating process in matrix form of the error state is described as:

$$\begin{cases} \delta\boldsymbol{x} \leftarrow \boldsymbol{F}_x\delta\boldsymbol{x} \\ \boldsymbol{P} \leftarrow \boldsymbol{F}_x \cdot \boldsymbol{P} \cdot \boldsymbol{F}_x^{\mathrm{T}} + \boldsymbol{F}_i \cdot \boldsymbol{Q} \cdot \boldsymbol{F}_i^{\mathrm{T}} \\ \boldsymbol{F}_x = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F}_1 & 0 & 0 \\ 0 & \boldsymbol{I} & \boldsymbol{F}_2 & \boldsymbol{F}_1 \\ 0 & 0 & \boldsymbol{I} & 0 \\ 0 & 0 & 0 & \boldsymbol{I} \end{bmatrix} \in \mathbb{R}_{8\times8} \\ \boldsymbol{F}_i = \begin{bmatrix} 0 & 0 \\ \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \\ 0 & 0 \end{bmatrix} \in \mathbb{R}_{8\times4}, \delta\boldsymbol{x} = \begin{bmatrix} \delta\boldsymbol{p} \\ \delta\boldsymbol{v} \\ \delta\boldsymbol{a}_b \\ \delta\boldsymbol{g} \end{bmatrix} \in \mathbb{R}_{8\times1} \\ \boldsymbol{Q} = \begin{bmatrix} \sigma_a^2\Delta t^2\boldsymbol{I} & 0 \\ 0 & \sigma_{a_b}^2\Delta t\boldsymbol{I} \end{bmatrix} \in \mathbb{R}_{4\times4} \\ \boldsymbol{F}_1 = \boldsymbol{I}\Delta t, \boldsymbol{F}_2 = -\boldsymbol{R}\Delta t \end{cases} , \quad (7)$$

where $\boldsymbol{P}$ and $\boldsymbol{Q}$ denote the covariance matrix of the error state and the perturbation input; $\boldsymbol{F}_x$ is system transition matrix and $\boldsymbol{F}_i$ is perturbation transition matrix.

2) Error State Correction and Data Fusion

When the visual estimation is completed, the estimation result will serve as a measurement to correct the error state and achieve the fusion of visual and inertial estimation. The calculation of Kalman gain $\boldsymbol{K}$ and the correction of error states and the covariance matrix can be described as:

$$\begin{cases} \boldsymbol{K} = \boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{V})^{-1} \\ \delta\boldsymbol{x} \leftarrow \boldsymbol{K}(z - h(\boldsymbol{x}_t)) \\ \boldsymbol{P} \leftarrow (\boldsymbol{I}_{8\times 8} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{P} \\ \boldsymbol{H} = \dfrac{\partial h}{\partial \delta \boldsymbol{x}} = [\,\boldsymbol{I}_{2\times 2} \quad 0 \quad 0 \quad 0\,] \in \mathbb{R}_{2\times 8} \\ z = h(\boldsymbol{x}_t) + \boldsymbol{\epsilon}, \boldsymbol{v} \sim N(0, \boldsymbol{E}) \end{cases}, \quad (8)$$

where $\boldsymbol{P}$ represents Jacobian matrix of the observation $h(x)$ with respect to the error state and $\boldsymbol{E}$ represents the covariance matrix of the observation noise.

Then, the corrected error state is injected into the nominal state to obtain the true state of the system, and finally, the error state will be reset to zero and the covariance matrix is updated with the newest nominal state.

D. Prosthesis Motion Estimation and Control Scheme

After achieving the estimation of knee motion, the movement of the joint and toe of the prosthesis are derived with the prosthesis link model and forward kinematics.

As shown in Fig. 3, the base of the prosthesis is the knee joint and the end (prosthetic foot) is a bent linkage with a constant angle. By fixing another IMU to user's thigh, the angle between the thigh and the coronal plane can be measured and the clockwise rotation angle of the knee and ankle joints can be obtained with two encoders fixed on the two joint motors. Then, the positions of the ankle, heel and toe in the ground coordinate system can be calculated with forward kinematics. In addition, after obtaining the camera motion, the key frames are extracted based in the camera displacement, and the 2D point clouds of the key frames are added to the local map to construct the 2D point cloud environment map. In this study, the frames with displacements between consecutive frames within the range of 0.01 m to 0.05 m are taken as key frames.

After realizing the motion estimation, the estimation results should be integrated into prosthetic control to demonstrate the effect of the proposed method for prosthetic walking on uneven terrain. Different from exoskeletons and mobile robots, controlling the prosthesis to follow a pre-planned global trajectory to avoid the collision with uneven terrains may not be suitable since its motion depends on the amputee. Therefore, we propose a multi-objective optimization-based swing control scheme prototype for upstair walking of prosthesis and integrate the estimated motion into the scheme.

First, an IMU installed on the socket of the prosthesis and a six-dimensional force sensor at the ankle are utilized to measure the angle between the thigh and the coronal plane of the body, as well as the ground reaction force. These measurements determine the status of the human leg (flexion/extension) and prosthesis (stance/swing phase). In the swing phase, the motion estimation results are integrated into the employed multi-objective optimization function for obstacle collision avoidance. Other objectives such as smooth landing and motion fluidity are also considered, and the optimized state of prosthesis at time t is described as:

$$\begin{bmatrix} \dot{q}_{\mathrm{knee}}(t) \\ \dot{q}_{\mathrm{ankle}}(t) \end{bmatrix}^{*} = \underset{(q_{\mathrm{knee}}, q_{\mathrm{ankle}})}{\operatorname{argmin}} \sum_{i=1}^{n} f\left(\begin{bmatrix} \dot{q}_{\mathrm{knee}}(t+i) \\ \dot{q}_{\mathrm{ankle}}(t+i) \end{bmatrix} \right), \quad (9)$$

where $f(\cdot)$ is the optimization objective function.

This iterative process involves seeking the minimum value of the cumulative optimization objective function of the prosthesis motion state across sequential time steps to predictively derive the desired joint angles within the constraint range. Subsequently, the prosthesis is actuated using a position-based lower-level PD control to follow the desired joint angles.

III. Experiment and Results

A. Experimental Protocol

To evaluate our visual-inertial motion estimation method, an able-bodied subject was invited to perform 5 times of walking under three types of terrains (upstairs, downstairs and obstacle) to collect environmental point cloud and IMU dataset to estimate the motion of camera and the subject's foot.

The experiment environment is shown in Fig. 4(a). The height and width for stair and obstacle are 14.7 cm and 28 cm (for stair) and 14 cm and 13 cm (for obstacle), respectively. The depth camera and IMU were mounted together on subject's right shank, and another two IMUs were mounted on the subject's right shank and heel to obtain the ankle angle. The capture frame rate of the point cloud and the frequency of the IMUs were 30 frames per second and 100 Hz, respectively. Visual and IMU data are collected in two threads and time-synchronized by capturing and fusing the latest data from both threads. In addition, in the experiments, a motion capture system (Raptor-12HS, Motion Analysis Corporation, USA) was used to record the motion of the subject's lower limbs and the camera at a frequency of 120 Hz. The measurements of the motion capture system served as the ground truth of the camera and lower limb motion. The proposed method was evaluated by calculating the absolute trajectory error (ATE) between the motion trajectory estimated by the propose method and the trajectory measured by the motion capture system. Although our prosthesis control scheme utilizes prosthesis motion between frames rather than ATE, it can still demonstrate that the proposed method can achieve accurate motion estimation.

In addition, to evaluate the proposed method on powered prosthesis, as shown in Fig. 4(b), an able-bodied subject was invited to perform upstair walking with a powered prosthesis. The prosthesis was equipped with an L-shaped receiving chamber for the subject to wear and was controlled through the control scheme mentioned in Section II.D. To estimate the motion of prosthesis, the same depth camera and IMU as in previous experiments were fixed on its knee, and the prosthetic foot motion was calculated using the joint angle acquired by two encoders mounted on knee and ankle joint of the prosthesis. The motion capture system was also used to obtain the ground truth. Besides, in all experiments, the subject began walking in a position close to stair or obstacle and started performing the corresponding gait directly.

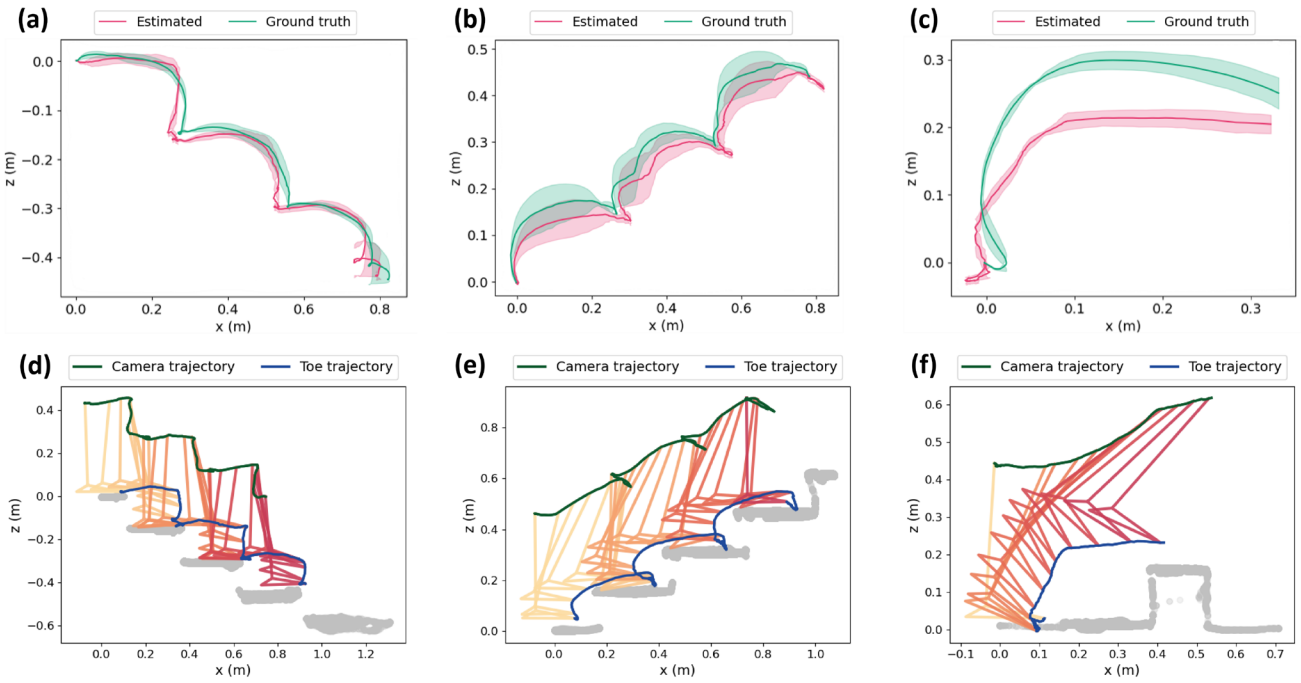


Fig. 5. The experiment result on the collected dataset. (a), (b) and (c) describe the estimated toe position of the subject in sagittal plane and the ground truth in downstair and upstair walking and obstacle crossing, respectively. The solid line and the shading in the graph represent the mean ± standard deviation of the toe positions, respectively. (d), (e) and (f) describe the estimated motion of the subject's leg in sagittal plane and the 2D point cloud map.

The experimental data processing was performed using Python 3.7 and Matlab 2020b on a computer equipped with an AMD Ryzen 7 4800H CPU (2.9 GHz), 16 GB of RAM, and an NVIDIA Geforce RTX 2060 GPU.

B. Estimation Result on the Collected Dataset

The proposed method is firstly evaluated using the dataset collected from walking of an able-bodied subject in different terrains. For each frame, the average time cost of point cloud preprocessing and visual estimation is 6.3 ms and 19.4 ms respectively. In addition, the time cost of IMU integration, data fusion, prosthesis motion calculation per frame are all less than 1 ms. Therefore, the total forward time is shorter than the acquisition time of camera (> 30 ms), indicating the promising real-time performance of the proposed method.

The estimated toe position of the subject in sagittal plane and the ground truth during walking in different terrains are shown in Fig. 5(a)-Fig. 5(c) and the motion of the subject's shank and foot and the constructed point cloud map during walking are shown in Fig. 5(e)-Fig. 5(f). Since collisions are more likely to occur between foot and uneven terrain during walking, the motion of foot is our main concern. The trajectories show that the proposed method tracks the toe motion well in all three uneven terrains. The average ATE in upstair and downstair walking are 2.96 cm ± 0.92 cm (upstair walking, x), 2.16 cm ± 0.46 cm (upstair walking, z), 2.79 cm ± 0.79 cm (downstair walking, x) and 1.63 cm ± 0.23 cm (downstair walking, z), respectively. In obstacle crossing, since the camera is not able to see the obstacle during the whole crossing, only the periods when the obstacle is visible are taken for ATE calculation, and the average ATE are 2.20 cm ± 0.14 cm (x) and 4.68 cm ± 0.47 cm (z), respectively, which are still under 5 cm.

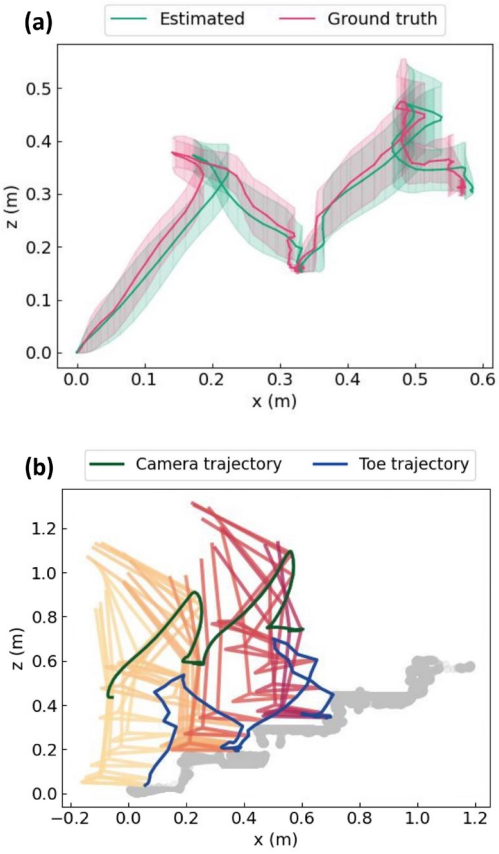


Fig. 6. The experiment result on data collected during prosthesis upstair walking. (a) describes the estimated toe position of the subject in sagittal plane and the ground truth. The solid line and the shading in the graph represent the mean ± standard deviation of the toe positions, respectively. (b) describes the estimated motion of the prosthesis in sagittal plane and the envirionmental map of 2D point cloud.

C. Estimation Result on the Prosthesis walking

Then, the proposed method is evaluated on data collected in upstair walking with the prosthesis. An able-bodied subject was invited to wear the prosthesis with an L-shaped receiving chamber and the prosthesis motion during walking was estimated using the proposed method. The estimated toe position of the prosthesis in sagittal plane and the ground truth are shown in Fig. 6(a), and the motion of the whole prosthesis during walking and the constructed map after smoothing is shown in Fig. 6(b). The average ATE of the toe in upper stair walking is 4.14 cm ± 0.85 cm (x) and 2.73 cm ± 0.71 cm (z).

In the prosthesis walking experiment, the quality of the point cloud is affected by a more drastic movement of the camera, which affects the accuracy of the motion estimation of prosthesis. However, despite a slight increase in estimation errors compared to normal upstair walking, the system can still track the prosthesis motion accurately. In addition, there were occasional frames where the camera could not observe the stairs. By combining the inertial information, the camera position can still be estimated, which improves the robustness of estimation.

IV. Discussion

In this study, a visual-inertial prosthesis motion estimation method is proposed to determine the movement of the powered prosthesis within the environment. Our aim is to enhance the safety and predictability of prosthesis mobility on uneven terrains like stairs and obstacles, thereby enhancing the prosthesis's environmental adaptability. The estimation results on dataset collected during walking of an able-bodied subject on different terrains and with prosthesis indicate that the ATE in all experiments is less than 5 cm, demonstrating the accuracy of the proposed method. Besides, the prosthetic upstair walking experiment shows that our method can be integrated with predictive control of the prosthesis and is expected to help amputees to walk on uneven terrain. More experiments on amputee walking with prosthesis on uneven terrains will be described in our future work.

In addition, in this study, since our main concern is the camera and prosthesis motion in the sagittal plane, the computation effort required for point cloud alignment is significantly reduced with environmental data preprocessing. According to [21], the time cost of directly performing ICP alignment for about 38k points in 3D space is about 150 ms, while the visual estimation time for each frame in this study is only 19.4 ms, which is only 12.9% of direct ICP alignment. Considering that the swing phase of a human gait cycle lasts about 800 ms [27], the processing speed of our method can meet the requirements for real-time prosthetic control. Besides, by incorporating a SP-ESKF which is designed towards sagittal plane movements, the inertial information in the sagittal plane is fused with the visual information to obtain a more robust motion estimation of the prosthesis.

Although the proposed method successfully realizes accurate motion estimation of the prosthesis within the environment, there are still some limitations. First, only several uneven terrains in real world are considered in this study. To enhance the environmental adaptative walking in various environments, the proposed method is required to generalize to more diverse terrains in future work. Besides, if the visual estimation fails for a rather long time, the estimation relying only on the inertial information will still present a large cumulative error. Therefore, in future work, other correction methods for the inertial information, such as zero velocity update (ZUPT) [28], can be incorporated with the proposed method.

V. Conclusion

The motion of powered prosthesis within the environment is important for navigating uneven terrains. In this paper, to enhance environmental adaptative walking of prothesis, a visual-inertial motion estimation method is proposed. The present approach utilizes the extracted feature points from the preprocessed 2D environmental point cloud for frame-to-frame alignment. Then, a sagittal plane motion-oriented error-state Kalman filter is developed to fuse the inertial and visual estimation results of the camera, which is utilized to derive the prosthesis motion and finally integrated into our prosthetic control scheme. This approach is evaluated on data collected in able-bodied subject's walking on different terrains and in upstair walking with powered prosthesis. The estimation result shows that the position of the prosthesis toe can be accurately estimated during walking (ATE: 4.14 cm ± 0.85 cm (x) and 2.73 cm ± 0.71 cm (z)). The proposed method is expected to improve the environmental awareness of the prosthesis, thereby enhancing the environmental adaptability and safety for walking on uneven terrains.

References

[1] A. Seker, A. Kara, S. Camur, M. Malkoc, M. M. Sonmez, and M. Mahirogullari, "Comparison of mortality rates and functional results after transtibial and transfemoral amputations due to diabetes in elderly pati-ents - A retrospective study", *Int. J. Surg.*, vol. 33, pp. 78-82, Sep. 2016.

[2] M. Tran, L. Gabert, S. Hood and T. Lenzi, "A lightweight robotic leg prosthesis replicating the biomechanics of the knee ankle and toe joint", *Sci. Robot.*, vol. 7, no. 72, pp. 1-18, Nov. 2022.

[3] T. Elery, S. Rezazadeh, C. Nesler and R. D. Gregg, " Design and validation of a powered knee–ankle prosthesis with high-torque, low impedance actuators", *IEEE Trans. Robot.*, vol. 36, no. 6, pp. 1649 1668, Dec. 2020.

[4] F. Gao, G. Liu, F. Liang and W.-H. Liao, "IMU-based locomotion mode identification for transtibial prostheses orthoses and exoskeletons", *IEEE Trans. Neural Syst. Trans. Neural Syst. Rehabil. Eng*., vol. 28, no. 6, pp. 1334-1343, Jun. 2020.

[5] J. S. Matthis, J. L. Yates and M. M. Hayhoe, "Gaze and the control of foot placement when walking in natural terrain", *Current Biol.*, vol. 28, no. 8, pp. 1224-1233, Apr. 2018.

[6] K. Zhang, H. Liu, Z. Fan, X. Chen, Y. Leng, C. W. De Silva and C. Fu, "Foot placement prediction for assistive walking by fusing sequential 3D gaze and environmental context", *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2509-2516, Apr. 2021.

[7] Y. Qian, Y. Wang, C. Chen, J. Xiong, Y. Leng, H. Yu and C. Fu, "Predictive locomotion mode recognition and accurate gait phase estimation for hip exoskeleton on various terrains", *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6439-6446, Jul. 2022.

[8] B. Laschowski, W. McNally, A. Wong and J. McPhee, "Comparative analysis of environment recognition systems for control of lower-limb exoskeletons and prostheses", *Proc. IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechtron. (BioRob),* pp. 105-110, Nov. 2020.

[9] K. Zhang, J. Luo, W. Xiao, W. Zhang, H. Liu, J. Zhu, Z. Lu, Y. Rong, C. W. De Silva and C. Fu, "A subvision system for enhancing the

environmental adaptability of the powered transfemoral prosthesis", *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3285-3297, Jun. 2021.

[10] K. Zhang, C. Xiong, W. Zhang, H. Liu, D. Lai, Y. Rong and C. Fu, "Environmental features recognition for lower limb prostheses toward predictive walking", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 465-476, Mar. 2019.

[11] S. Yin, X. Chen, T. Ma, Y. Wang. Y. Guo, Y. Leng and C. Fu, "Environmental recognition and multimode continuous-phase control for a powered transfemoral prosthesis", *Proc. IEEE Int. Conf. Dev. Learn. (ICDL)*, pp. 55-60, 2023.

[12] C. Chen, K. Zhang, Y. Leng, X. Chen and C. Fu, "Unsupervised sim-to-real adaptation for environmental recognition in assistive walking", *IEEE Trans. Neural Syst. Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1350-1360, May 2022.

[13] K. Zhang, J. Chen, J. Wang, Y. Leng, C. W. De Silva and C. Fu, "Gaussian-guided feature alignment for unsupervised cross-subject adaptation", *Pattern Recognit.*, vol. 12, 108332, Feb. 2022.

[14] T. K. Skervin, N. M. Thomas, A. J. Schofield, M. A. Hollands, "The next step in optimising the stair horizontal-vertical illusion: Does a perception-action link exist in older adults", *Exp. Gerontol.*, vol. 149, 111309, Jul. 2021.

[15] J. Lin and F. Zhang, "R3LIVE: A robust real-time RGB-colored LiDAR-inertial-visual tightly-coupled state estimation and mapping package", *Proc. Int. Conf. Robot. Autom. (ICRA)*, pp. 10672-10678, May 2022.

[16] S. Zhao, H. Zhang, P. Wang, L. Nogueira and S. Scherer, "Super odometry: Imu-centric lidar-visual-inertial estimator for challenging environments", *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pp. 8729-8736, 2021.

[17] B. Kleiner, N. Ziegenspeck, R. Stolyarov, H. Herr, U. Schneider and A. Verl, "A radar-based terrain mapping approach for stair detection towards enhanced prosthetic foot control", *Proc. 7th IEEE Int. Conf. Biomed. Robot. Biomechatronics (Biorob)*, pp. 105-110, Aug. 2018.

[18] M. Ramanathan et al., "Visual environment perception for obstacle detection and crossing of lower-limb exoskeletons", *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, pp. 12267-12274, Oct. 2022.

[19] J. Luo et al., "Robust bipedal locomotion based on a hierarchical control structure", *Robotica*, vol. 37, no. 10, pp. 1750-1767, Oct. 2019.

[20] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An open-source slam system for monocular stereo and RGB-D cameras", *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255-1262, Oct. 2017.

[21] S. Chen, C.-W. Chang and C.-Y. Wen, "Perception in the dark; development of a ToF visual inertial odometry system", *Sensors*, vol. 20, no. 5, pp. 1263, Feb. 2020.

[22] K. Zhang, W. Zhang, W. Xiao, H. Liu, C. W. De Silva and C. Fu, "Sequential decision fusion for environmental classification in assistive walking", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1780-1790, Sep. 2019.

[23] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3D point cloud map", *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* (*IROS*), pp. 4802-4809, Oct. 2018.

[24] X. Chen, Y. Wang, C. Chen, Y. Leng and C. Fu, "Towards environment perception for walking-aid robots: an improved staircase shape feature extraction method", *Intell. Robot.*, vol. 4, pp. 179-95. May 2024.

[25] F. Aghili and C.-Y. Su, "Robust relative navigation by integration of ICP and adaptive Kalman filter using laser scanner and IMU", *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 4, pp. 2015-2026, Aug. 2016.

[26] W. Xu, D. He, Y. Cai and F. Zhang, "Robots' state estimation and observability analysis based on statistical motion models", *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 5, pp. 2030-2045, Sep. 2022.

[27] X. Chen, K. Zhang, H. Liu, Y. Leng and C. Fu, "A probability distribution model-based approach for foot placement prediction in the early swing phase with a wearable IMU sensor", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2595-2604, 2021.

[28] H. Fourati, "Heterogeneous data fusion algorithm for pedestrian navigation via foot-mounted inertial measurement unit and complementary filter", *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 221-229, 2015.