Whole-Body Pose Estimation in Physical Rider–Bicycle Interactions With a Monocular Camera and Wearable Gyroscopes

Pose estimation of human–machine interactions such as bicycling plays an important role to understand and study human motor skills. In this paper, we report the development of a human whole-body pose estimation scheme with application to rider–bicycle interactions. The pose estimation scheme is built on the fusion of measurements of a monocular camera on the bicycle and a set of small wearable gyroscopes attached to the rider’s upper- and lower-limbs and the trunk. A single feature point is collocated with each wearable gyroscope and also on the body segment link where the gyroscope is not attached. An extended Kalman filter (EKF) is designed to fuse the visual-inertial measurements to obtain the drift-free whole-body poses. The pose estimation design also incorporates a set of constraints from human anatomy and the physical rider–bicycle interactions. The performance of the estimation design is validated through ten subject riding experiments. The results illustrate that the maximum errors for all joint angle estimations by the proposed scheme are within 3 degs. The pose estimation scheme can be further extended and used in other types of physical human–machine interactions.

[DOI: 10.1115/1.4035760]

Keywords: pose estimation, multibody kinematic optimization, human–machine interactions, sensor fusion, gyroscope

1 Introduction

Whole-body pose or gait information is commonly used to study human movement science and biomechanics. Accurate and reliable human whole-body pose information benefits both clinical disease diagnosis and rehabilitation treatment. Most studies of pose estimation are for human activities such as walking, running, and stance. For human–machine interactions, such as bicycling, driving, kayaking, and riding segways, pose or gait estimation plays an important role to understand and study human motor skills. However, it is not straightforward to estimate the whole-body poses in these activities due to complex, machine-coupled dynamic movements, and few studies have been reported for pose estimation in these interactions.

Optical or magnetic marker-based human motion capture systems are commercially available and commonly used. However, in addition to high cost, these motion capture systems are limited to laboratory usage and it is difficult to be used for personal activities at home or in outdoor environment. Wearable sensors, such as small-size gyroscopes, accelerometers, and inertial measurement units (IMU), have been used to attach on human body to obtain the motion, segment orientation, or position information [1,2]. It is well known that directly integrating the measurements of these sensors produces unacceptable results due to drifting noises. Integration of the measurements from these inertial sensors and other complementary sensors (e.g., global positioning system, magnetometers, and ultrasonic sensors) is commonly taken to enhance the accuracy of the pose or position estimation.

We propose a visual-inertial fusion framework to simultaneously estimate the poses of the human whole-body and the moving platform. Several attractive features motivate us to use bicycling as an application example to demonstrate the proposed pose estimation scheme. First, bicycling provides a good application example to demonstrate the proposed pose estimation scheme. First, bicycling provides a good application example to demonstrate the proposed pose estimation scheme.
example for the above-mentioned physical human–machine interactions. Sitting on the moving bicycle, the rider actively reacts to balance the unstable platform through whole-body movements. The multilimb and body movements on a moving bicycle provide a rich and challenging problem to identify and extract the human motion from the platform motion. Second, recent clinical studies show that bicycling helps diagnose or treat patients with peripheral motor or balancing disorders [3], Parkinson’s diseases [4], and mental retardation [5]. These bicycle-based rehabilitation and clinical applications need accurate pose estimations in outdoor environment.

One of major applications of visual-inertial fusion is for robotic localization and navigation [6–11]. The popular fusion methods used in these applications include the extended Kalman filter (EKF) or the unscented Kalman filter. Many existing developments focus on the estimation and calibration of the rigidly mounted camera-IMU pair for robot navigation and mapping applications [12–14]. Integration of visual and inertial measurements is also applied to human motion tracking and localization. For example, in Ref. [15], measurements from the IMU mounted on human body are fused with the camera located on the body or off-body fixtures. In Ref. [16], a real-time hybrid visual-inertial fusion scheme is presented to track only the three-dimensional articulated human arm motion for home-based rehabilitation. To estimate human walking, in Ref. [17] a monocular visual odometry is fused with IMU measurements and a human walking model. In Ref. [18], the use of monocular video with the multibody dynamic models is reported to estimate the lower-limb poses for real-time human tracking. The above-mentioned visual-inertial integrations consider the cases of inertial frame-fixed cameras or rigidly connected camera-inertial sensor pairs. For the inertial frame-fixed camera, the limitation is that the workspace is confined and it is not suitable for outdoor usage. Rigidly connected camera-inertial sensor pairs cannot be used as wearable sensors for the pose estimation of human activities because they are not of light-weight and small-size and may cause significant intrusive effects on human motions.

In the application such as the rider–bicycle interactions, the camera is proposed to be mounted on the moving platform and not rigidly fixed together with the inertial sensors. The monocular camera provides only relative distances between the body segments and the moving platform. The rationales of the bicycle-mounted camera paired with the human body-mounted gyroscope sensors come from several reasons. First, for whole-body human pose estimation in physical human–machine interactions such as bicycling, the attitudes of a large number of body segments need to be measured or estimated. It is desirable to separate the camera with the inertial sensors to obtain the relative position of multiple locations on the body segments simultaneously with that only one camera is needed. Second, without attaching a relatively large amount of cameras on human body segments, the small-size wearable sensors such as miniature gyroscopes and the point markers bring minimally intrusive disturbances for human movements. Finally, the setup procedure of the proposed system is simple and the cost of the system is low. The noncollocated camera-gyroscope approach is attractive for human-in-the-loop or human–machine interactions applications.

In Refs. [19–22], the measurements from the body IMU and force sensors are fused to estimate the bicycle rider trunk orientation. With the help of those force sensors, the absolute angles are indirectly captured and used to compensate for the estimation drifts of the trunk and the bicycle angles by using the inertial sensors. The limitation of such fusion is that it cannot be applicable for complex structures formed by limb segments. In Ref. [23], a visual-inertial integration scheme is presented to estimate the real-time upper-limb pose when riding a bicycle. This paper extends the approaches and the results in Ref. [23] in several aspects. We improve the visual-inertial fusion framework to estimate the rider’s whole-body segments orientations using a simpler system configuration than that in Ref. [23]. Second, the feature point used in this work is a single dot and much simpler than the rectangular square used in Ref. [23]. The feature point simplification brings challenges to the visual-inertial fusion. Finally, we conduct extensive multiple subject experiments to validate and demonstrate the design. This paper is an extension of the conference publications [23,24] with significantly additional analyses and experiments.

The rest of the paper is organized as follows: In Sec. 2, we describe the system configuration and setup. Section 3 presents the visual measurements and gyroscope models. Section 4 discusses the visual-inertial fusion design. The experimental results and discussions are presented in Sec. 5. Finally, we summarize the concluding remarks in Sec. 6.

2 Rider–Bicycle Systems and Configuration

2.1 Rider–Bicycle Systems. Figure 1(a) shows the instrumented bicycle and the wearable sensors. Two sets of the gyroscope units (both from Motion Sense Inc., Hangzhou, China) are used in the experiments. One tri-axial gyroscope unit (model 605) is mounted to the bicycle frame and the trunk, respectively, and four small-size tri-axial gyroscope units (model slimAHRS) are attached to the rider’s forearms and thighs, as shown in Fig. 1(b). A small-size battery and a wireless transmitter module are integrated and packed with wearable gyroscope units. A feature dot attached to the outside of the gyroscope box provides the position information. Using one single dot as the feature point reduces the image processing time and enables the real-time applications. Around the feature point, a rectangular square marker is used to provide the ground truth poses for outdoor experiments [20]. A total of seven feature points are attached to the rider’s body: four on the upper-limb (one on each forearm together with the gyroscope and the other one on the upper arm), one on the trunk, and the other two on the lower-limb (one on each thigh) (see Fig. 2).

A commercial mountain bicycle was modified and equipped with various sensors. A high-resolution monocular camera (model Manta G-145 from Allied Vision Technologies, 1392 × 1040 pixels, 16 fps) is mounted on an extended rod rigidly connected to the bicycle frame. The image acquisition and processing are conducted on the onboard computer through a Gigabit Ethernet connection. The wireless receivers of the bicycle and wearable gyroscopes are connected to a real-time embedded system (NI cRIO 9082-RT) mounted on the rear rack. Two optical encoders (Grayhill, Series 63R 256 lines) are mounted on the crank and the steering fork to measure their absolute angles, respectively. All sensor measurements are sampled at the frequency of 100 Hz except that the image sampling rate is 16 Hz. The ground truth of both the rider and the bicycle poses are obtained by the optical motion capturing systems (8 Bonita cameras from Vicon Inc., Oxford, UK) for indoor experiments. The data collections among the Vicon system, cRIO, and the camera computer are synchronized through wireless networks connections.

2.2 System Configuration. Figure 1(b) shows the configuration of the visual-inertial system. We define the inertial frame $N$ with the Z-axis downward, the camera frame $C$, the image frame $I$, and the feature frames $F_i, i = 1, ..., 7$. As shown in Fig. 1(b), $F_1$ and $F_2$ are for the features attached to left upper-limb, $F_3$ and $F_4$ for the right upper-limb, $F_5$ for the trunk, and finally $F_6$ and $F_7$ for the features on the left and right thighs, respectively.

We simultaneously estimate the joint angles of the upper- and lower-limbs, the trunk, and the bicycle roll angle. The upper-limb can be modeled as a three-joint link: the wrist (three degrees-of-the-freedom (3DOF)), the elbow (1DOF), and the shoulder (3DOF). During bicycling, the hands are assumed firmly attached to handlebar. Therefore, the limb-trunk system constitutes an...
articulated chain and the two ends of the chain are the handlebar and seat, respectively. As long as the trunk angles, the wrist angles, and the elbow angles are known, the shoulder angles are indeed uniquely determined. Therefore, in this study we neglect the estimations for shoulder angles and focus only on the estimations of the trunk angles, the wrist angles, and the elbow angles. The elbow is considered as a 1DOF joint [25] with joint angle $\gamma_{ilq}$, the wrist as a 3DOF joint with orientation by the Euler angles $\theta_{lw}, \phi_{lw}, \psi_{lw}$, and the elbow as a 3DOF link: the hip (thigh) joint is a 3DOF link whose orientation is captured by Euler angles $\theta_{lh}, \phi_{lh}, \psi_{lh}$, and the 1DOF knee joint with joint angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. Figure 2 shows the details of the limb pose configurations.

Although the Euler angle representation is used, we can calculate the pronation-supination, radial-ulnar, and flexion-extension in the anatomical conventions definition. Moreover, for each 3DOF joint (i.e., wrist, trunk, and hip joints), we introduce the quaternion representation to avoid the potential singularity issue encountered by Euler angles $\theta_{ilw}, \phi_{ilw}, \psi_{ilw}$, $i = l, r$, for the left and right limbs, respectively. Euler angles $\theta_{ilw}, \phi_{ilw}, \psi_{ilw}$, and $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively. For presentation conventions, we also define $\beta_{ilq}$, the 1DOF ankle joint with angle $\delta_{ilw}, i = l, r$, for the left and right limbs, respectively.

3 Vision Measurements and Gyroscope Kinematic Models

3.1 Vision-Based Measurements. Figure 1(a) shows the feature markers on the human body. We treat the center of the black dot as the location of the feature point. Five of these markers are collocated with the five gyroscopes (see Fig. 1(b)). The feature

mark is segmented from the image by the threshold. Similar to the approach in Ref. [23], the Ostu algorithm [26] is chosen to calculate the optimum separating threshold through minimizing the intraclass variance of two classes (i.e., foreground and background pixels) to distinguish markers from the environment. The Blob analysis method [27] is then used to extract image features and to generate the size and position of the target. A run-length encoding method is used to process the image before conducting the Blob analysis for quickly locating the features.

A pinhole model is considered for the monocular camera as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K I \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$$

(2)

where $(u, v)$ in $I$ is the image of point $(x, y, z)$ in $C$, $K_i$ is the camera intrinsic parameter matrix and obtained through the calibration process. We denote the transform matrix between $C$ and $F_i$ as $\mathbf{T}_i$, $i = 1, ..., 7$. For the $i$th feature point $(x_i, y_i, z_i)$, its coordinates $(x_i, y_i, z_i)$ in $C$ satisfy

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \mathbf{T}_i \begin{bmatrix} x_f \\ y_f \\ z_f \end{bmatrix}$$

(3)

where $\mathbf{T}_i$ is further decomposed as

$$\begin{bmatrix} x_f \\ y_f \\ z_f \end{bmatrix} = \begin{bmatrix} R & P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix}$$

(4)

where frame $B$ is attached to the bicycle. For the feature frames on the upper-limb, $B$ is chosen as the handlebar frame and for the feature frames on the trunk and the lower-limb, $B$ is attached on the bicycle seat.

Using Eqs. (2)–(4) and setting the feature point as the origin of the world frame, i.e., $x_f = y_f = z_f = 0$, we obtain
where $\mathbf{R}$ is a constant and related to the camera mounting parameters. We write both $\mathbf{R}$ and $\mathbf{p}$ as functions of attitudes of feature $\mathcal{F}_i$, i.e., $X$. From Eq. (5), we have

$$u_i = I_i(X)/I_{i3}(X), \quad v_i = I_{i2}(X)/I_{i3}(X), \quad i = 1, \ldots, 7$$

where $I_{ij}(X), \quad j = 1, 2, 3$, is the $j$th element of $I_i(X)$. The image positions of all seven feature points are concatenated together as

$$h(X) = [u_1 \quad v_1 \ldots \quad u_7 \quad v_7]^T$$

### 3.2 Gyroscope Kinematic Models

For the bicycle gyroscope, we have the following relationship for the measurements $\omega_i$ [20]

$$\omega_i = R^T_i(\delta)R_i^T_i(\varphi_i) \begin{bmatrix} 0 \\ \varphi_i \\ 0 \end{bmatrix} + R^T_i(\varphi_i) \begin{bmatrix} \psi_i \\ 0 \\ 0 \end{bmatrix}$$

where $R_i(\delta)$ represents the three-dimensional rotational matrix around the $i$-axis with angle $\vartheta_i, \ i = x, y, z$, and $\delta$ is the angle between the bicycle gyroscope and the horizontal plane (see Fig. 1(b)). Similarly, the forearm gyroscope measurements $\omega_{f_i}, \ i = l, r$, for the left and right arms are written as

$$\omega_{f_i} = \begin{bmatrix} \hat{\varphi}_{f_{i0}} \\ 0 \\ + R^T_i(\phi_{f_{i0}})R^T_i(\phi_{f_{i0}}) \begin{bmatrix} 0 \\ \phi_{f_{i0}} \\ 0 \end{bmatrix} + R^T_i(\phi_{f_{i0}}) \begin{bmatrix} \psi_{f_{i0}} \\ 0 \\ 0 \end{bmatrix} \\ \psi_{f_{i0}} \end{bmatrix}$$

where $\beta$ and $\gamma$ are the bicycle caster and steering angles, respectively. The trunk gyroscope measurements $\omega_b$ are written as

$$\omega_b = R^T_i(\theta_i)R^T_i(\varphi_b) \begin{bmatrix} 0 \\ \varphi_b \\ 0 \end{bmatrix} + R^T_i(\varphi_b) \begin{bmatrix} \psi \\ 0 \\ 0 \end{bmatrix}$$

and the lower-limb gyroscope measurements $\omega_h, \ i = l, r,$

$$\omega_h = R^T_i(\varphi_h)R^T_i(\varphi_h) \begin{bmatrix} 0 \\ \varphi_h \\ 0 \end{bmatrix} + R^T_i(\varphi_h) \begin{bmatrix} \psi_h \\ 0 \\ 0 \end{bmatrix}$$

The kinematic equations in Eqs. (8)–(10) give the relationship between the attitude rates and the gyroscope measurements through the Euler angle representation. It is desirable to write these kinematic relationships by using the quaternion representation. This is achieved by using the transformation between the Euler angles and quaternion representations. For example, the Euler angles of the wrist are written as

$$\begin{bmatrix} \varphi_{iw} \\ \phi_{iw} \\ \theta_{iw} \end{bmatrix} = \begin{bmatrix} 2^{(\sin \theta_{iw} \sin \phi_{iw} + \sin \theta_{iw} \cos \phi_{iw})} \\ 2^{(\cos \theta_{iw} \sin \phi_{iw} + \cos \theta_{iw} \cos \phi_{iw})} \\ 2^{(\sin \phi_{iw})} \end{bmatrix} \arctan \frac{2^{(\sin \theta_{iw} \sin \phi_{iw} + \sin \theta_{iw} \cos \phi_{iw})}}{1 - 2^{(\cos \theta_{iw} \sin \phi_{iw} + \cos \theta_{iw} \cos \phi_{iw})}}$$

Taking a time derivative of Eq. (11), we obtain

$$\begin{bmatrix} \hat{\varphi}_{iw} \\ \hat{\phi}_{iw} \\ \hat{\theta}_{iw} \end{bmatrix} = f_{iw}(x_{iw}, \dot{x}_{iw}), \quad i = l, r$$

Due to the length of the functions in Eq. (12), we here omit to write them explicitly. Plugging Eqs. (11) and (12) into Eq. (8), we obtain the kinematic relationship between $x_{iw}, \ y_{iw}, \ \text{and} \ \omega_{iw}$. Similar relationships are obtained for $x_b$ and $\varphi_b$ with the gyroscope measurements $\omega_b$ and $\omega_h$. These kinematic relationships will be used in the EKF design in Sec. 4.2.

### 4 Rider–Bicycle Interactions Constraints and Extended Kalman Filter-Based Pose Estimation

In this section, we first present the constraints imposed by the physical rider–bicycle interactions and then discuss the EKF-based whole-body pose estimation design.

#### 4.1 Human Anatomy and Rider–Bicycle Interaction Constraints

##### 4.1.1 Lower-Limb Pedaling Motion Constraint

Let us first consider the lower-limb constraints due to pedaling [28]. Figure 2(b) illustrates the lower-limb pose during pedaling. The hip joint, the knee joint, and the ankle joint are denoted as $A, B,$ and $C$, respectively. We also denote the foot–pedal contact point as $D$. Without causing confusion, we use $A, B, C,$ and $D$ to indicate the corresponding frames attached to hip, knee, ankle, and pedal joints. It is then straightforward to obtain that the homogeneous transformation from $D$ to $A$ is

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{R}_i \\ 0 \\ \rho_i \end{bmatrix}$$

for $i = l, r$, where $\mathbf{T}_i = R^T_i(\varphi_i)R^T_i(\varphi_i) \begin{bmatrix} \varphi_i \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{T}_i = R^T_i(\varphi_i)R^T_i(\varphi_i) \begin{bmatrix} \psi_i \\ 0 \\ 0 \end{bmatrix}$.
and \(l_{ibr}, l_{ibu}, \) and \(l_{if}\) are the lengths of segment links of the lower-limb. From Eq. (13), we obtain the position vector \(\mathbf{r}_{CBi}\) of point \(D\)

\[
\mathbf{r}_{CD} = \mathbf{r}_{CBi} + \mathbf{r}_{BI} + \mathbf{r}_{BC} + \mathbf{r}_{CB} + \mathbf{r}_{Bi}, \quad i = l, r
\]

in the hip frame as a function of the quaternion coordinates of the lower-limb \(\mathbf{x}_l\) (through a transformation similar to Eq. (11)).

On the other hand, when pedaling, the feet are always on the pedal and the position of \(D\) can be obtained through the crank angle \(\phi_c(t)\) (measured by the crank encoder). Thus, we obtain that

\[
\mathbf{r}_{BI} = \mathbf{r}_O + \mathbf{r}_{OD} (\phi_c(t)), \quad i = l, r
\]

where constant vector \(\mathbf{r}_O\) represents the location of pedal center \(O\) in the hip frame, and \(\mathbf{r}_{OD}(\phi_c(t)) = \left[ r_x \cos \phi_c(t) \ 0 \ r_y \ sin \phi_c(t) \right]^T\) (positive for \(i = l\) and negative for \(i = r\)) with \(r_x\) is the crank radius. Combining Eqs. (14) and (15), we have

\[
c_i^C (\mathbf{q}^T ; \gamma_i^b; \delta^a) := \mathbf{R}^b \mathbf{R}^e \mathbf{r}_{BI} + \mathbf{R}^b \mathbf{r}_{BC}; = d_i(t)
\]

where \(d_i(t) := \mathbf{r}_O - \mathbf{r}_{BI} + \mathbf{r}_{OD}(\phi_c(t))\) is a function of crank angle \(\phi_c(t)\).

### Upper-Limb Anatomical Constraints

The second set of constraints is the location of the rider neck \(E\) as shown in Fig. 1(b). We obtain the position vector of \(E\), denoted as \(\mathbf{r}_{BE}(\mathbf{x}_{nw}; \mathbf{x}_e)\) (only the state variables in the function arguments are written explicitly), through the linkage of the handlebar and upper-limb from the bicycle seat. On the other hand, we calculate the position vector of \(E\) in the trunk frame as

\[
\mathbf{r}_{BE}(\mathbf{x}_{nw}; \mathbf{x}_e) = \mathbf{R}(\mathbf{x}_e) \mathbf{p}_E
\]

where \(\mathbf{R}(\mathbf{x}_e)\) is the rotation matrix with quaternion \(\mathbf{x}_e\) and constant vector \(\mathbf{p}_E = [0, 0, 0]^T\) is the relative position vector of \(E\) in trunk frame (see Fig. 1(b)). Therefore, we write down the second set of constraints as

\[
c_2^C (\mathbf{x}_{nw}; \mathbf{x}_e; \mathbf{x}_h) := \mathbf{r}_{BE}(\mathbf{x}_{nw}; \mathbf{x}_e) = \mathbf{R}(\mathbf{x}_e) \mathbf{p}_E
\]

We obtain the third constraint from human anatomy that the shoulder length is constant, denoted as \(l_s\). We calculate the two shoulder points from the left and right elbows and thus, we express this constraint as

\[
c_3^C = c_3(\mathbf{x}_e) = l_s
\]

### Quaternion Coordinate Constraint

The last set of constraints are the unity of the quaternion representation for the upper-limb, trunk, and the thigh joints, i.e.,

\[
c_4^C = c_4(\mathbf{x}_e) = 1
\]

In summary, a total of 15 constraints are obtained, i.e.,

\[
\mathbf{c}_1^C(\mathbf{q}^T ; \gamma_i^b; \delta^a) \in \mathbb{R}^6, \mathbf{c}_2^C(\mathbf{x}_{nw}; \mathbf{x}_e; \mathbf{x}_h) \in \mathbb{R}^3, \mathbf{c}_3^C \in \mathbb{R}, \text{ and } \mathbf{c}_4^C \in \mathbb{R}^3.
\]

### 4.2 Extended Kalman Filter-Based Pose Estimation

From Eq. (7), we solve for \(\dot{\phi}_b\) and obtain

\[
\dot{\phi}_b = g_b(\omega_b) = \left[ \cos \delta \ 0 \ \sin \delta \right] \omega_b
\]

as the first state equation. Similarly, from Eqs. (8), (11), and (12) we solve for \(\dot{x}_{nw}\) and \(\dot{x}_{nw}\) and obtain (the closed-form expressions in \(X\) are omitted here due to the length of these functions)

\[
\dot{x}_{nw} = g_{nw}(\mathbf{x}, \omega_{nw}, \omega_b), \quad \dot{x}_{nw} = g_{nw}(\mathbf{x}, \omega_{nw}, \omega_b)
\]

as the second set of the state equations. From Eq. (9), we solve for \(\dot{x}_{h}\) (through transformations (11) and (12)) and obtain

\[
\dot{x}_{h} = g_h(\mathbf{x}, \omega_{h}, \omega_b)
\]

as the third set of the state equations.

Since only one gyroscope is attached to each lower-limb, unlike the wrist, trunk, or thigh joints, the knee angle rate \((\dot{\gamma}_k)\) and the ankle angle \((\dot{\delta}_a)\) cannot be written as similar to those kinematic relationships in Eqs. (8), (9), or (10). To overcome such a measurement deficiency and to include \(\dot{\gamma}_k\) and \(\dot{\delta}_a\) into the EKF state equation, we take the time derivative of Eq. (16) and obtain

\[
\frac{\partial c_i^C}{\partial \mathbf{q}} \mathbf{q}_d + \frac{\partial c_i^C}{\partial \gamma_i^b} \dot{\gamma}_i^b + \frac{\partial c_i^C}{\partial \delta_i^a} \dot{\delta}_i^a = \dot{d}_i(t)
\]

Equation (23) has three equations and combining with Eq. (10), we solve for \(\dot{x}_{il} = [\dot{\mathbf{q}}^T \ \dot{\gamma}_i^b \ \dot{\delta}_i^a]^T\) and obtain

\[
\dot{x}_{il} = g_{il}(\mathbf{x}, \omega_{il}, \omega_b), \quad \dot{x}_{il} = g_{il}(\mathbf{x}, \omega_{il}, \omega_b)
\]

We form the state vector \(\mathbf{x}(\mathbf{t}), \Omega(t) = \left[ \mathbf{q}^T \ \dot{q}^T \ \mathbf{q}_d^T \ \dot{\gamma}_i^b \ \dot{\delta}_i^a \right]^T \in \mathbb{R}^{29}\), where measurements \(\Omega(t) = \left[ \omega_{nw}^T \ \omega_{nw}^T \ \omega_{nw}^T \ \omega_{nw}^T \ \omega_{nw}^T \ \omega_{nw}^T \ \mathbf{l}^T \ \mathbf{d}^T \right]^T \in \mathbb{R}^{12}\). We rewrite the EKF state equation in discrete-time as

\[
\mathbf{X}(\mathbf{k}) = \mathbf{X}((\mathbf{k}-1) + \Delta T \ (\mathbf{X}((\mathbf{k}-1), \Omega((\mathbf{k}-1)))) + \mathbf{w}(\mathbf{k})
\]

at the 4th sampling time and \(\Delta T = 0.01\) s is the data sampling period, where \(\mathbf{w}(\mathbf{k}) \sim \mathcal{N}(0, \sigma_w)\) is the mean-zero white noise with variance matrix \(\sigma_w\). The EKF outputs use the vision-based measurements in Eq. (6) and the constraints discussed in Sec. 4.1, namely,

\[
\mathbf{Y}(\mathbf{k}) = \begin{bmatrix} 
\mathbf{h}(\mathbf{k}) \\
\mathbf{c}_1^C(\mathbf{k}) \\
\mathbf{c}_2^C(\mathbf{k}) \\
\mathbf{c}_3^C(\mathbf{k}) \\
\mathbf{c}_4^C(\mathbf{k}) \\
\end{bmatrix} + \mathbf{n}(\mathbf{k}) = \mathbf{h}_f(\mathbf{k}) + \mathbf{n}_f(\mathbf{k})
\]

where \(\mathbf{Y}(\mathbf{k}) \in \mathbb{R}^{29}\) and \(\mathbf{n}(\mathbf{k}) \sim \mathcal{N}(0, \sigma_f)\) is the white noise vector with variance matrix \(\sigma_f\). We obtain the Jacobian matrices \(\mathbf{F}(\mathbf{k}) = \mathbf{I} + \mathbf{H}(\mathbf{k}) \mathbf{P}(\mathbf{k}) \mathbf{H}(\mathbf{k})^T\) for state dynamics and \(\mathbf{H}(\mathbf{k}) = \partial \mathbf{h}_f / \partial \mathbf{X}(\mathbf{k})\) for the outputs. The EKF design is calculated as \([29]\)
The observability condition is checked for the EKF design (using matrices F and H) and the rank condition is satisfied.

Remark 1. The EKF design contains the bicycle roll kinematics (20) and uses measurements from the bicycle gyroscope and the wearable body gyroscopes. None of these sensors gives absolute bicycle roll angle information. As shown in Sec. 5, it is interesting that the EKF-based sensing fusion helps maintain a bounded estimation error for the bicycle roll angle while the directly integrated strapdown gyroscope measurements result in a diverging estimation error.

Remark 2. In the EKF design, we do not consider any noise models for the gyroscope measurements. As demonstrated in Ref. [20], the inclusion of the first-order colored noise model could help further reduce the estimation errors. However, the improvements of the estimation performance with the noise model are marginal and therefore, we do not present the noise model here.

5 Experiments

5.1 Experimental Results. We conduct the human riding experiments in both an indoor laboratory as shown in Fig. 1(a) and an outdoor setting. Due to the space constraint, the subjects ride the bicycle for a circular trajectory of a radius of 2 m and a moving speed around 1–1.5 m/s in an indoor environment.

For outdoor riding experiments, the subjects ride for a large circle with a radius around 7 m and a moving speed around 2.5–3 m/s.

The physical parameters for the bicycle are as follows: β = 16.7 deg, δ = 10 deg, and \( r_c = 0.17 \) m. We recruited ten healthy and experienced bicycle riders (eight male and two female with age: 27 ± 3 years, height: 176 ± 4 cm, and weight: 70 ± 7 kg) to conduct the experiments. The duration for each riding experiment was around 2 mins. Before conducting experiments, we measured all subjects’ biomechanic parameters and the locations of the gyroscopes and feature markers on their body segments for pose calculations. All the subjects signed their informed consent using a protocol approved by the Institutional Review Board (IRB) at Rutgers University. In the following, we first describe the indoor experimental results and then present the outdoor experiments.

Figure 3 shows the experimental results of the pose estimation for one subject. We present the Euler angle representation for a better visualization. For comparison purposes, we also plot the pose estimates by directly integrating strapdown gyroscope measurements. Figures 3(a) and 3(b) show the pose estimates for the right upper-limb, namely the wrist angles and the elbow angle. Fig. 3(c) shows the trunk pose estimation, and Figs. 3(d) and 3(e) show the right lower-limb’s pose estimation, including the thigh’s Euler angles, the knee angle, and the ankle angle. Finally, Fig. 3(f) shows the estimates of the bicycle roll angle. The results shown in Fig. 3 clearly demonstrate that the visual-inertial fusion results closely follow the ground truth of the rider and the bicycle poses. The fusion scheme certainly outperforms the direct integration of strapdown gyroscope measurements, which drift over time and the errors increase dramatically after 25 s. The corresponding estimation error comparisons are shown in Fig. 4. The maximum errors for all joint angle estimations by the visual-inertial fusion are within 2–3 deg. Compared with other fusion-based pose estimation designs for regular human activities, the proposed approach for human–machine interactions shows the comparable accuracy level. For example, 2.4–3.2 deg mean root-mean-square (RMS) orientation estimation errors for different segments are...
Fig. 4 Pose estimation errors from one subject experiment. The Vicon system measurements are considered as the ground truth: (a) right wrist Euler angle estimation errors, (b) right elbow angle error $\Delta \theta_{re}$, (c) trunk Euler angle estimation error, (d) right thigh Euler angle estimation errors, (e) right leg ankle estimation error $\Delta \theta_{ra}$ and the knee angle estimation error $\Delta \theta_{rk}$, and (f) the bicycle roll angle estimation error $\Delta \phi_b$.

Fig. 5 The mean and standard deviation (SD) of the pose estimation errors of the body segments from ten subject indoor experiments. The mean values of the EKF design are plotted by the solid lines and the one SD values are plotted as the dash lines. (a) Left-wrist Euler angle estimation statistics, (b) left-thigh Euler angle estimation statistics, (c) right-wrist Euler angle estimation statistics, (d) right-thigh Euler angle estimation statistics, and (e) trunk Euler angle estimation statistics.
demonstrated by using the fusion of the inertial and the magnetic sensors in Ref. [30]. A 2.8 deg mean RMS error is shown by fusing the accelerometers and the gyroscopes in Ref. [31].

To further demonstrate the performance of the EKF-based design for indoor testing, we compute the statistics of the pose estimation errors for all ten subjects. Figure 5 shows the calculated statistics of the estimation errors of all limbs and trunk joint angles over time. We plot the estimation error statistics obtained from both the visual-inertial fusion and the gyroscope integration schemes. It is clearly observed that for all subject runs, the estimation errors by the EKF fusion are near zero. Table 1 further lists the mean and one SD of the RMS errors for all subjects. The third row in Table 1 shows the EKF estimation errors for the upper-limb pose without using constraints $c_2^u$ and $c_3^u$. It is interesting to observe that incorporating $c_2^u$ and $c_3^u$ into the EKF slightly improves the estimation performance. The results shown in Table 1 and Fig. 5 confirm the consistently superior performance of the EKF-based pose estimation than those by the direct integration of gyroscope measurements.

In Fig. 4(b), no integration results are plotted because only one gyroscope is attached to each forearm and no gyroscope measurements for the upper arm. For the lower-limb, although only one gyroscope is attached to each thigh, we still obtain the knee and ankle angles accurately through the gyroscope measurements and constraint (16). Clearly, these constraints enable the design to use few number of wearable gyroscopes. The pose estimation in

<table>
<thead>
<tr>
<th></th>
<th>$\phi_{uw}$</th>
<th>$\phi_{wr}$</th>
<th>$\theta_{uw}$</th>
<th>$\theta_{wr}$</th>
<th>$\phi_u$</th>
<th>$\phi_r$</th>
<th>$\theta_u$</th>
<th>$\theta_r$</th>
<th>$\gamma_u$</th>
<th>$\gamma_r$</th>
<th>$\gamma_u$</th>
<th>$\gamma_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EKF</td>
<td>3.36 ± 0.7</td>
<td>1.57 ± 0.5</td>
<td>1.30 ± 0.1</td>
<td>0.77 ± 0.1</td>
<td>1.74 ± 0.1</td>
<td>1.93 ± 0.1</td>
<td>1.78 ± 0.2</td>
<td>1.82 ± 0.1</td>
<td>1.84 ± 0.4</td>
<td>1.79 ± 0.6</td>
<td>1.53 ± 0.7</td>
<td></td>
</tr>
<tr>
<td>EKF no $c_2^u$, $c_3^u$</td>
<td>4.03 ± 1.0</td>
<td>1.69 ± 0.4</td>
<td>1.51 ± 0.2</td>
<td>1.01 ± 0.1</td>
<td>2.03 ± 0.2</td>
<td>1.92 ± 0.1</td>
<td>2.11 ± 0.5</td>
<td>2.01 ± 0.1</td>
<td>2.14 ± 0.5</td>
<td>1.89 ± 0.6</td>
<td>1.73 ± 0.8</td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>10.97 ± 2.3</td>
<td>9.20 ± 2.3</td>
<td>6.02 ± 0.6</td>
<td>5.28 ± 0.3</td>
<td>7.78 ± 1.3</td>
<td>4.97 ± 0.7</td>
<td>8.21 ± 1.6</td>
<td>4.51 ± 0.8</td>
<td>4.72 ± 0.6</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>EKF</td>
<td>2.42 ± 1.0</td>
<td>0.98 ± 0.3</td>
<td>0.99 ± 0.3</td>
<td>0.43 ± 0.1</td>
<td>1.37 ± 0.3</td>
<td>0.82 ± 0.1</td>
<td>1.43 ± 0.9</td>
<td>1.67 ± 0.8</td>
<td>1.36 ± 0.1</td>
<td>1.57 ± 0.2</td>
<td>1.33 ± 0.7</td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>13.30 ± 3.3</td>
<td>5.32 ± 0.8</td>
<td>3.43 ± 1.2</td>
<td>5.74 ± 1.1</td>
<td>3.99 ± 1.2</td>
<td>4.81 ± 1.3</td>
<td>5.43 ± 2.1</td>
<td>4.97 ± 1.7</td>
<td>8.37 ± 1.2</td>
<td>7.52 ± 2.1</td>
<td>5.75 ± 2.8</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6 Pose estimation and comparison results with strapdown gyroscope integration and the ground truth from camera system from one subject’s outdoor riding experiment. (a) Right wrist Euler angle estimation, (b) trunk Euler angle estimation, (c) right thigh Euler angle estimation, and (d) bicycle roll angle $\phi_b$ estimation.
outdoor experiments shows the consistent results similar to the above-discussed indoor-experiment results. Figure 6 shows the rider pose and bicycle roll angle estimation. The rider pose estimates shown in Figs. 6(a)–6(c) demonstrate that the EKF-based scheme produces the accurate estimation results and the performance is comparable to the results in the indoor experiments. The prediction results of the bicycle roll angle shown in Fig. 6(d) also demonstrate the superior performance than the IMU integration.

Figure 7 shows the time trajectories of the mean and one standard deviation of the pose estimates of major body segments and the trunk of the ten bicycle riders in outdoor experiments. From the plots, we clearly observe that within around 50 s riding experiments, the estimates of all postangles by the EKF-based fusion scheme are all around zero, and the one SD values are mostly within a range of 2 deg. Therefore, the proposed visual-inertial fusion scheme can be used to estimate the human body motion in bicycle riding in outdoor environment. On the other hand, from the plots in Fig. 7, we also notice that the SDs for most joint angles of human body segments slowly grow over time though the growths are much smaller than those by directly integrating of the gyroscope signals (see comparison results in Fig. 4). The slow.

![Fig. 7](image1)

Fig. 7 The mean and SD of the pose estimation errors of the body segments from ten subject outdoor experiments. The mean values of the EKF design are plotted by the solid lines and the one SD values are plotted as the dash lines: (a) left-wrist Euler angle estimation statistics, (b) left-thigh Euler angle estimation statistics, (c) right-wrist Euler angle estimation statistics, (d) right-thigh Euler angle estimation statistics, and (e) trunk Euler angle estimation statistics.

![Fig. 8](image2)

Fig. 8 The estimation errors for the right wrist pose angles under various loss percentages of the camera images

<table>
<thead>
<tr>
<th>LoV (%)</th>
<th>Right arm</th>
<th>Left arm</th>
<th>Trunk</th>
<th>Right leg</th>
<th>Left leg</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.01 ± 0.52</td>
<td>1.20 ± 0.31</td>
<td>1.32 ± 0.63</td>
<td>1.63 ± 0.63</td>
<td>1.06 ± 0.34</td>
</tr>
<tr>
<td>20</td>
<td>2.38 ± 0.94</td>
<td>1.39 ± 0.58</td>
<td>1.55 ± 0.83</td>
<td>1.92 ± 0.84</td>
<td>1.34 ± 0.56</td>
</tr>
<tr>
<td>40</td>
<td>4.04 ± 1.64</td>
<td>2.40 ± 1.05</td>
<td>2.08 ± 1.12</td>
<td>3.44 ± 1.50</td>
<td>2.60 ± 1.17</td>
</tr>
<tr>
<td>60</td>
<td>4.94 ± 2.00</td>
<td>3.55 ± 1.54</td>
<td>3.39 ± 2.05</td>
<td>4.40 ± 1.93</td>
<td>3.48 ± 1.56</td>
</tr>
<tr>
<td>80</td>
<td>7.40 ± 2.96</td>
<td>4.48 ± 1.94</td>
<td>4.42 ± 2.35</td>
<td>5.70 ± 2.47</td>
<td>4.38 ± 1.97</td>
</tr>
<tr>
<td>100</td>
<td>8.92 ± 3.57</td>
<td>5.76 ± 2.48</td>
<td>5.50 ± 2.78</td>
<td>7.60 ± 2.69</td>
<td>5.43 ± 2.41</td>
</tr>
</tbody>
</table>

Table 2 Statistics of the RMS errors (in deg) with one-SD under loss of vision (LoV) frames
growth of the estimation errors is primarily due to the vision estimation errors.

5.2 Discussions. The main contribution of the work is the development of a low-cost, wearable-visual-inertial design for whole-body human pose estimation in physical human–machine interactions with a moving and dynamic platform. Comparing with the commercially available inertial-based whole-body sensor systems (e.g., Ref. [2]), our approach has several attractive features. The integration of the gyroscopes with visual measurements as well as the physical constraints is robust and reliable while the accuracy of the use of magnetometers in the sensor systems [2] is potentially vulnerable to highly dynamic motion and environmental disturbances, such as existence of metallic objects nearby. Additionally, we focus on using the possible least measurement information while maintaining the consistent estimation performance in all joint angles. Finally, the cost of the wearable sensors (only gyroscopes and one camera) in our approach is much less than most commercial whole-body sensor systems.

One potential limitation of using visual measurements in practice is to obtain the high-quality, reliable images, especially for rapid human movement. Image occlusion, out of field of view of the camera, or blurring images reduces the vision-based pose estimation accuracy. Figure 8 shows the estimation errors of the EKF fusion results for the right wrist angles under a random loss of camera images for all ten subjects. We calculated the EKF estimation errors by randomly dropping the image frames following a given loss percentage. With increasing percentages of the lost camera images, the EKF fusion performance deteriorates. Table 2 lists the statistics of the RMS of the pose estimation for each limb and the trunk among all ten subjects. Clearly, these results confirm that the performance of the visual-inertial fusion scheme is robust to a certain amount of the vision frame loss. For example, with a 20% image loss, the average estimation error of each limb is still less than 2.5 deg.

Besides the riding bicycle example, the proposed method could also be applicable for accurate pose estimations in other applications of human exercises or human–machine interactions, such as car driving, kayaking, or flying an airplane. The proposed pose estimation could also be potentially used for monitoring human activities of elderly and disabled patients (on or not on moving platforms) in a nonlaboratory environment. The system uses the miniaturized wearable gyroscopes and collocated point markers, and two cameras [2] easily attached to human body segments without significantly interfering the human activities in daily personal life. The setup procedure of the proposed system is simple since both the gyroscope bias and noise levels and the camera calibration can be conveniently obtained before the usage of the system.

6 Conclusion
We presented a visual-inertial fusion scheme for human whole-body pose estimation in human–machine interactions with applications to bicycle riding. The fusion scheme was built on the measurements of a calibrated monocular camera and a set of gyroscopes attached to rider’s body segments. With two feature points on each segment of the upper-limb and one feature point on each thigh of the lower-limb, the EKF-based fusion scheme estimated the whole-body poses when the subjects ride a bicycle. The EKF design also incorporated the constraints from human anatomy and the physical rider–bicycle interactions. The superior performance of the pose estimation design was extensively tested by multiple subjects for both indoor and outdoor experiments. The results showed that the maximum errors for all joint angle estimations by the proposed scheme were within 3 deg and this performance was similar to those of wearable IMUs for regular human activities. The proposed scheme provided a low-cost, high-accuracy, and wearable design for whole-body human pose estimation in human–machine interactions and can be potentially used in other types of human–machine interactions for outdoor or personal activities.

Acknowledgment
This work was supported in part by the U.S. National Science Foundation under Award Nos. CMMI-0954966 and CMMI-1334389, the Shanghai Eastern Scholarship Program through Shanghai University (J. Yi), the National Natural Science Foundation of China under Award No. 61403307 (Y. Zhang), and fellowships from Chinese Scholarship Council (K. Yu and X. Lu). The authors also thank K. Chen, M. Trkov, P. Wang, and other members at the Robotics, Automation, and Mechatronics (RAM) Lab at Rutgers University for their helpful discussions and experimental help.

References

071005-10 / Vol. 139, JULY 2017 Transactions of the ASME


