2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

Real-Time Detection of Food Consumption Activities Using Wearable Wireless Sensors*

Gregory Johnson, Yan Wang, and Rajesh Rajamani

Abstract— **This paper addresses research challenges associated with development of a wearable sensor system for detecting the food consumption activities of a subject. The objective is to automatically detect the occurrence of food consumption whenever it occurs, in order to use this activity detection to record a representative camera image of the food and count the number of bites of food consumed. The wearable system consists of two elastic bands – one each on the upper arm and wrist – instrumented with wireless inertial and magnetic sensors. Two major technical challenges include i) singularity issues with Euler angle estimation due to arm rotations that can exceed 90 degrees, and ii) the need to differentiate between eating and non-eating activities involving close hand-mouth proximity. The singularity challenge is addressed by using a direction cosine matrix estimation technique that utilizes a linear Kalman Filter. The differentiation between eating and non-eating activities is done using a support-vector-machine (SVM) based machine learning algorithm. Experimental results using wearable prototype bands show that both the DCM estimation and machine learning components work reliably and have the potential to be useful for home-based automated food intake detection.**

I. Introduction

The rapid rise in obesity over the past few decades comes from a frequent mismatch between calorie intake and physical activity. Physical activity can be monitored with wearable devices such as pedometers. However, no corresponding solutions are available for automated monitoring of food intake.

This paper takes initial steps towards developing a wearable food intake monitoring (FIM) system that can automatically record food consumption for the community of patients that need diet assessment. The FIM system will consist of a wrist band and an upper arm band, both of which are instrumented with sensors. The two bands will communicate wirelessly with a computer in the home to enable back up of recorded data to the cloud. The objective of the FIM system is to detect food consumption by the subject whenever it occurs, in order to potentially record the time, number of bites, and a reconstructed camera image of the food during each consumption event.

In this paper, we present the preliminary results of a prototype wearable FIM system consisting of a wrist band and an upper arm band, both of which are instrumented with inertial motion sensors (IMU). The two bands communicate wirelessly with a host computer and stream motion data. By monitoring hand position relative to the head, the system detects when the user is placing his/her hand near the mouth. By further utilizing machine learning tools, the activity is classified as eating or not eating, and the number of bites can be counted.

The first step in detecting eating is tracking hand position in 3D space, which requires knowledge of the spatial orientation of each FIM sensor. The 3D orientation of each sensor can be parameterized in a number of ways, including Euler angles, quaternions, and the 3x3 rotation matrix, also referred to as the direction cosine matrix (DCM). Though intuitive, Euler angles suffer from singularity issues [1], and are not suitable for estimation algorithms involving large rotations, such as arm movement. Quaternions do not suffer from these singularity issues, but the kinematic model describing their time evolution is non-linear, thus attitude filtering using a quaternion parameterization is often accomplished with the Extended Kalman Filter (EKF), or Unscented Kalman Filter (UKF) [2] which do not guarantee global stability. Although it contains redundancies, the DCM does not suffer from singularities and the kinematic model is linear [1], thus allowing the use of a linear Kalman Filter [3], guaranteeing convergence. Once the orientation of each sensor is known, the real-time position of the hand can be computed.

However, hand position itself does not provide full information on what activity is taking place. Machine learning tools can be utilized to further classify the subset of activities in which the hand is near the mouth. In this paper, a support vector machine (SVM) classifies hand activity taking place near the mouth as eating or not-eating. By limiting classification only to hand activity near the mouth, training can be limited to a smaller subset of activities (eating, teeth brushing, chin scratching, nose picking, etc).

The remainder of this paper is organized as follows. In Section II, the FIM system prototype hardware is detailed. In Section III, the use of the DCM for estimating hand position is described. In Section IV, experimental validation showing hand position estimation and hand-mouth proximity detection is presented. In Section V, a SVM is utilized to classify activity as eating or not-eating.

II. Wearable Sensor System Prototype

The FIM sensor system consists of two wireless, battery-powered sensor assemblies, one each on the wrist and the upper arm. Each assembly contains a nine-axis Inertial

*Research partially supported by the Digital Technology Center, University of Minnesota.

G. Johnson is a Ph.D. student in Mechanical Engineering at the University of Minnesota, Minneapolis, MN 55455, USA. Author PIN: 130870 (e-mail: joh13594@umn.edu).

Y. Wang is a Post-Doctoral Scholar in Mechanical Engineering at the University of Minnesota, Minneapolis, MN 55455, USA. Author PIN: 85469 (e-mail: wangyan0731@outlook.com).

R. Rajamani is a Professor of Mechanical Engineering at the University of Minnesota, Minneapolis, MN 55455, USA. Author PIN: 5245 (e-mail: rajamani@umn.edu).

Measurement Unit (IMU, InvenSense MPU-9250), consisting of a three-axis accelerometer, magnetometer, and gyroscope. The prototype system is shown in Fig. 1. The data from the IMU is sampled over an I^2C bus by a Bluetooth LE System on Chip (Nordic Semiconductor nRF51), which uses a 32-Bit ARM Cortex M0 CPU. Each assembly is powered by a 500 mAH lithium polymer battery. The assembly is packaged in a 3D printed housing and attached to a wearable sleeve.

In the current prototype, raw motion data consisting of two sets of IMU data is sent to the host computer at a 50Hz rate. The battery allows the device to send a constant data stream to the host for 36 hours before requiring charging. The host records raw data, estimates calibration parameters, estimates the proximity of the hand to the mouth, and detects whether eating events are occurring.

III. Direction Cosine Matrix Estimation

Orientation in 3D can be parameterized in a number of different ways. The three most common parameterizations are Euler angles, quaternions, and the full 3x3 DCM [1]. Euler angles consist of a series of three successive rotations about a specified order of axes. For the particular order of yaw(ψ)--pitch(θ)--roll(ϕ), a rotation that transforms a vector with representation v_n in the inertial frame to its representation v_b in the body frame can be written as follows [1]:

$$v_b = \underbrace{R(\phi)R(\theta)R(\psi)}_{R_n^b} v_n. \quad (1)$$

The matrices $R(\phi)$, $R(\theta)$, and $R(\psi)$ are given by:

$$R(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c(\phi) & s(\phi) \\ 0 & -s(\phi) & c(\phi) \end{bmatrix}, \quad R(\theta) = \begin{bmatrix} c(\theta) & 0 & -s(\theta) \\ 0 & 1 & 0 \\ s(\theta) & 0 & c(\theta) \end{bmatrix}$$

$$R(\psi) = \begin{bmatrix} c(\psi) & s(\psi) & 0 \\ -s(\psi) & c(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where $s(\,) = \sin(\,)$ and $c(\,) = \cos(\,)$.

The full DCM is R_b^n, and when the multiplication of the Euler rotations is carried out the DCM can be written as:

$$R_n^b = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix},$$

$$v_1 = \begin{bmatrix} c(\psi)c(\theta) \\ c(\psi)s(\phi)s(\theta) - c(\phi)s(\psi) \\ s(\phi)s(\psi) + c(\phi)c(\psi)s(\theta) \end{bmatrix},$$

$$v_2 = \begin{bmatrix} c(\theta)s(\psi) \\ c(\phi)c(\psi) + s(\phi)s(\psi)s(\theta) \\ c(\phi)s(\psi)s(\theta) - c(\psi)s(\phi) \end{bmatrix}, \; v_3 = \begin{bmatrix} -s(\theta) \\ c(\theta)s(\phi) \\ c(\phi)c(\theta) \end{bmatrix} \quad (3)$$

The singularity issue of the Euler angle parameterization is apparent when $\theta = \pm 90°$ [1]. In the $\theta = 90°$ case, the DCM becomes:

$$R_n^b = \begin{bmatrix} 0 & 0 & -1 \\ \mathrm{s}(\psi - \phi) & \mathrm{c}(\psi - \phi) & 0 \\ \mathrm{c}(\psi - \phi) & -\mathrm{s}(\psi - \phi) & 0 \end{bmatrix}. \quad (4)$$

In other words, the ψ and ϕ angles are underdetermined, which will cause issues in estimation algorithms. A better approach is to directly estimate the DCM. The remainder of this section presents the DCM kinematic model and estimation algorithm.

A. DCM Process Model

The true kinematic differential equation for the DCM is [4]:

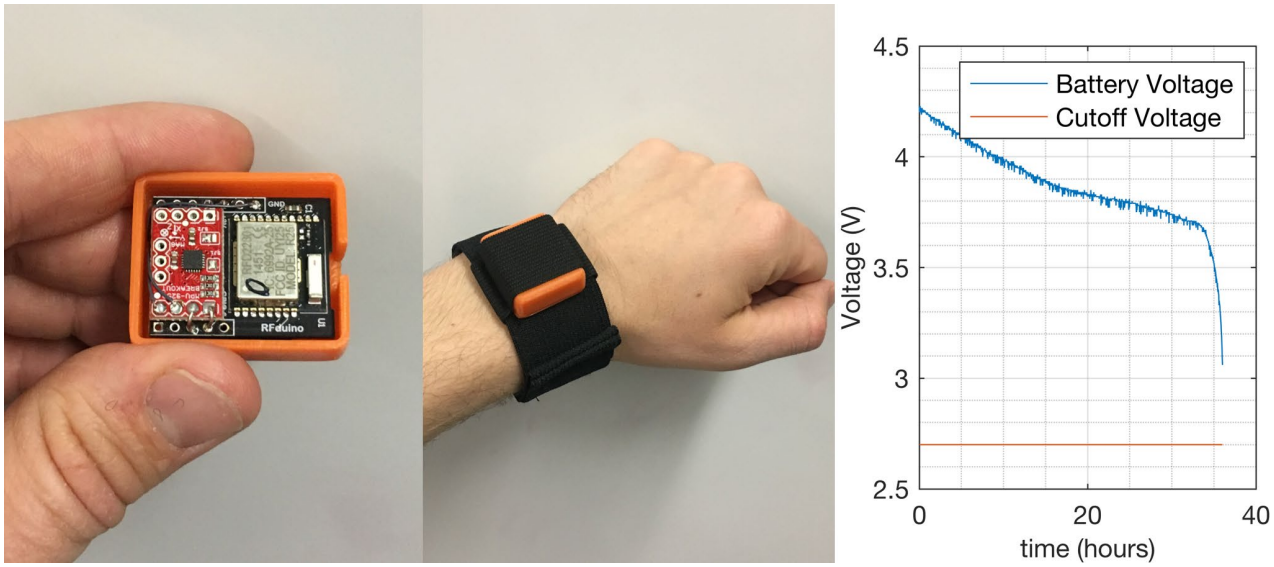

Fig. 1. Images show the prototype wrist sensor in the Food Monitoring System (FIM). Photo on the left shows the electronics housed in a 3D printed enclosure. Plot on the right shows battery voltage during constant streaming of data, demonstrating the system can easily operate for a full day of activity before charging.

$$\dot{R}_n^b = -S(\omega) R_n^b, \tag{5}$$

where R_b^n is a 3 x 3 matrix defined by equation (3) and $S(\omega)$ is a skew symmetric matrix of angular velocity in body coordinates defined as follows:

$$S(\omega) = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{6}$$

Likewise, the vectors v_1, v_2, and v_3 that make up the DCM follow the same kinematic differential equation:

$$\dot{v}_i = -S(\omega) v_i, \quad i = 1,2,3. \tag{7}$$

In reality, the true angular velocity is not known but is measured by the gyroscope, which can be modeled after removal of any bias as:

$$\omega = \omega_n + \varepsilon, \tag{8}$$

where $\varepsilon \sim N(0, \sigma_g^2 I_3)$ is a vector of zero mean Gaussian noise and ω_n is the true angular velocity. Assuming constant angular velocity between samples results in the discretization

$$v_i[k+1] = \exp\left(-S(\omega_k - \varepsilon_k)\Delta t\right) v_i[k]. \tag{9}$$

The matrix exponential is written in power series form as [5]

$$\exp\left(-S(\omega_k - \varepsilon_k)\Delta t\right) = I_3 - S(\omega_k)\Delta t + S(\varepsilon_k)\Delta t + h.o.t. \tag{10}$$

Neglecting higher order terms involving powers of Δt and $S(\varepsilon_k)$ greater than two gives the approximation:

$$\exp\left(-S(\omega_k - \varepsilon_k)\Delta t\right) \cong \exp\left(-S(\omega_k)\Delta t\right) + S(\varepsilon_k)\Delta t. \tag{11}$$

Finally, the discretized kinematic equation can be written as

$$v_i[k+1] = \exp\left(-S(\omega_k)\Delta t\right) v_i[k] + S\left(-v_i[k]\right)\Delta t \varepsilon_k. \tag{12}$$

To get the DCM process dynamics into a Kalman filter framework, it is necessary to stack v_1, v_2, and v_3 into the state vector $x = \begin{bmatrix} v_1^T & v_2^T & v_3^T \end{bmatrix}^T$, such that the process dynamics and noise can be written as [3]

$$x_{k+1} = F_k x_k + w_k$$

$$F_k = \begin{bmatrix} \exp\left(-S(\omega_k)\Delta t\right) & 0_3 & 0_3 \\ 0_3 & \exp\left(-S(\omega_k)\Delta t\right) & 0_3 \\ 0_3 & 0_3 & \exp\left(-S(\omega_k)\Delta t\right) \end{bmatrix}$$

$$w_k \sim N(0, Q_k), \quad Q_k = \begin{bmatrix} Q_{1,k} & 0_3 & 0_3 \\ 0_3 & Q_{2,k} & 0_3 \\ 0_3 & 0_3 & Q_{3,k} \end{bmatrix}$$

$$Q_{i,k} = (\Delta t)^2 S\left(-v_i[k]\right)\sigma_g^2 S\left(-v_i[k]\right)^T. \tag{13}$$

B. DCM Measurement Model

The low frequency or static component of the DCM can be fully estimated using the accelerometer and magnetometer measurements without using the gyroscopes. Typical measurement models for IMUs rely on knowledge of the local geomagnetic field vector, which varies by location on Earth. Here, the assumption is made only that the magnetic field is fixed relative to gravity, which allows the algorithm to work without prior knowledge of the magnetic field. First, define the inertial frame as Z aligned with gravity, X aligned with magnetic north, and Y by the right hand rule. By this convention, there is no magnetic field in the inertial Y direction [3].

Because the Z direction is down, the DCM component v_3 can be determined exclusively by the accelerometer with the assumption that motion is slow and the only measured acceleration is gravity.

$$v_3 = \left(\frac{1}{g}\right)\begin{bmatrix} a_x & a_y & a_z \end{bmatrix}^T, \quad g = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{14}$$

The magnetic field only has X and Z components, and thus the transformation of the magnetic field from body frame to inertial frame will follow:

$$\begin{bmatrix} B_X \\ 0 \\ B_Z \end{bmatrix} = \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}. \tag{15}$$

Because v_3 is found from the accelerometer, B_z is also available:

$$B_z = [v_3^T] \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = \left(\frac{1}{g}\right)\begin{bmatrix} a_x & a_y & a_z \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}. \tag{16}$$

The magnitude of the magnetic field is the same no matter the measurement frame, so B_x can be found by matching the magnitude of the magnetic field in inertial and body frames:

$$B_X^2 + B_Z^2 = b_x^2 + b_y^2 + b_z^2 \tag{17}$$

$$B_X = \sqrt{b_x^2 + b_y^2 + b_z^2 - B_Z^2}, \tag{18}$$

noting that B_X is positive by definition of the inertial frame. All information needed to find v_1 is now available by solving

$$\begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = B_x v_1 + B_z v_3 \tag{19}$$

TABLE I. SO(3) CONSTRAINED DCM KALMAN FILTER

Gain	$K_k = P_k^- H^T \left[HP_k^- H^T + R_k \right]^{-1}$
Correction	${}_{unc}\hat{x}_k^+ = \hat{x}_k^+ + K\left[y_k - H\hat{x}_k^- \right]$ $P_k^+ = \left[I - K_k H \right] P_k^-$
Orthonormal Constraint	${}_{unc}\hat{R}_n^b = \left[{}_{unc}\hat{x}_k^+(1:3), {}_{unc}\hat{x}_k^+(4:6), {}_{unc}\hat{x}_k^+(7:9) \right] = V\Sigma U^T$ $\hat{R}_n^b = V diag\left(I_2, detV detU \right) U^T$ ${}_c\hat{x}_k^+ = vec\left(\hat{R}_n^b \right)$
Propagation	$\hat{x}_{k+1}^- = F_k(\omega_k)\, {}_c\hat{x}_k^+$ $P_{k+1}^- = F_k(\omega_k) P_k^+ F_k^T(\omega_k)$

for v_1. Finally, the remaining vector v_2 can be found by enforcing the orthonormality constraint of rotation matrices:

$$v_2 = v_3 \times v_1 . \tag{20}$$

In the approach presented here, the elements of the DCM are treated as measurements by transforming the IMU measurements using (14), (19), and (20). Thus, the transformation of the sensor noise must also be considered. Let the sensor measurements be denoted as follows

$$y = y_n + \varepsilon = \left[a_x \quad a_y \quad a_z \quad b_x \quad b_y \quad b_z \right]^T + \varepsilon , \tag{21}$$

where $\varepsilon \sim N\left(0, P = diag\left(\sigma_a^2 I_3, \sigma_m^2 I_3\right)\right)$ and y_n are the deterministic noise free signals. Let $f(y)$ represent the transformations from the IMU measurements to the DCM elements given by (14), (19). Given sufficiently small noise, approximate the measurement transformation using a first order Taylor series approximation

$$f(y) = f(y_n + \varepsilon) \cong f(y_n) + \left[\frac{\partial f}{\partial y} \right]_{y=y_n} \varepsilon , \tag{22}$$

where $\left[\frac{\partial f}{\partial y} \right]$ is the Jacobian of $f(y)$. Thus, the transformation can be approximated using zero-mean Gaussian noise with covariance

$$R = \left[\frac{\partial f}{\partial y} \right]_{y=y_n} P \left[\frac{\partial f}{\partial y} \right]_{y=y_n}^T . \tag{23}$$

Since y_n is not available, in practice the sensor measurement $y_n + \varepsilon$ can be used to evaluate the Jacobian. After the transformation is applied, the measurement model is:

$$z_k = Hx_k , \quad H = I_9 \tag{24}$$

A detailed treatment of this derivation can be found in [3].

C. Kalman Filter Summary

Before summarizing the DCM Kalman filter, a final issue needs to be addressed concerning the Kalman filter correction step. While the propagation of the kinematic model preserves DCM orthonormality, the correction step does not. Thus, the estimate must be constrained. This is a common issue in attitude filtering using quaternions, and one approach is to introduce a normalization constraint step [6]. In the case of constraining the DCM estimate, the SVD can be used [1], in which the closest orthonormal matrix R_n^b to an unconstrained matrix ${}_{unc}R_n^b$ is found. This problem is referred to as Wahba's problem [1]:

$$R_n^b = \arg\max_{R_n^b \in SO(3)} tr\left(R_n^b \left[{}_{unc}R_n^b \right]^T \right) \tag{25}$$

There are several analytical solutions to (25), one of which uses the SVD of ${}_{unc}R_n^b$. If ${}_{unc}R_n^b$ has the SVD

$${}_{unc}R_n^b = V\Sigma U^T , \tag{26}$$

then the solution to (25) is [1]

$$R_n^b = V * diag(I_2, detV * detU) * U^T \tag{27}$$

Table I summarizes the SO(3) constrained Kalman filter developed using the process and measurement models described above.

IV. HAND AND MOUTH POSITION ESTIMATION

By using the constrained Kalman filter on two IMUs worn on the upper arm and wrist, the position of the hand, P_h, relative to the shoulder can be estimated using the upper and lower arm lengths, r_1 and r_2, respectively:

$$P_h = \left(R_n^{b1} \right)^T \begin{bmatrix} r1 \\ 0 \\ 0 \end{bmatrix} + \left(R_n^{b2} \right)^T \begin{bmatrix} r2 \\ 0 \\ 0 \end{bmatrix} . \tag{28}$$

With a third sensor on the body, the position of the mouth relative to the shoulder can also be estimated:

$$P_m = \left(R_n^{b1} \right)^T \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} , \tag{29}$$

where the mouth coordinates in body fixed coordinates relative to the shoulder are m_x, m_y, and m_z. Figure 2 shows the results of three experiments plotting the position of the hand relative to the body and the correct estimation of eating motion (A) compared to non-eating vertical plane motion (B) and non-eating horizontal motion (C). The Euler angles extracted from the DCM are also plotted in Fig. 2.

In reality, it is desirable to minimize the number of sensors needed for the FIM system. A necessary but not sufficient condition for eating is that the hand will be at a particular

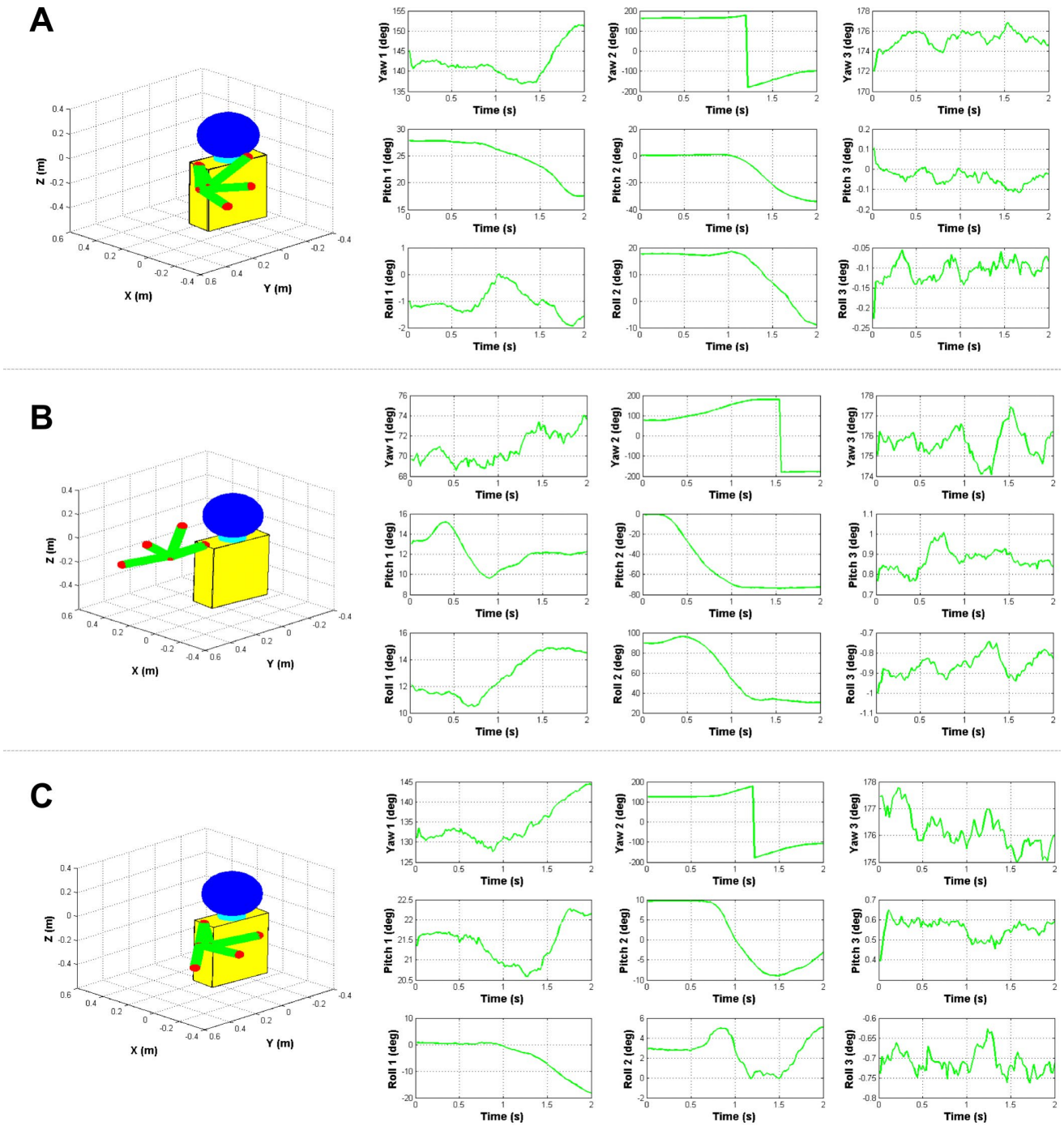

Fig. 2. Experimental data showing the correct reproduction of three different arm motions from recorded IMU data using the DCM estimation algorithm. Three IMUs were utilized: one on upper arm, one on lower arm, and one on the torso. Euler angles extracted from the DCM estimation are shown above each 3D plot. Experiment A demonstrates an eating motion. Experiment B demonstrates a waving motion. Experiment C demonstrates a horizontal sweep of the hand

radius away and height from the shoulder joint. These two parameters can be determined from *only the two sensors* on the arms. When the radius and height conditions are met, a machine learning algorithm can further classify the activity as eating or not eating.

V. Eating Detection Using Machine Learning

The Kalman filter described above can be used to find the position of the hand as shown in Fig. 2. The hand being near the mouth is a necessary condition for eating but does not guarantee that the activity is indeed eating. To reduce false-positive eating detection, machine learning can be used to improve FIM performance by classifying activities in which the hand is near the mouth. The advantage of this approach is that the machine learning algorithm only needs to be trained on activity that meets the radius and height criteria described previously. Without first estimating hand position, the algorithm needs to be trained against all possible hand activity. Thus, the classification of non-eating activity is a small subset

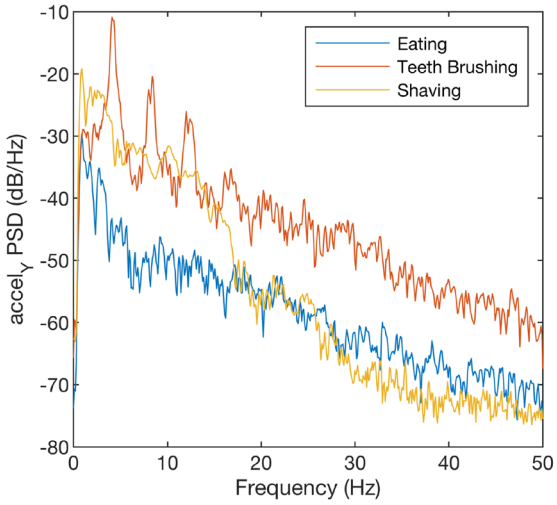

Fig. 3. Plot of acceleration PSD in direction of lower arm for three activities in which the hand is near the mouth: eating, teeth brushing, and shaving. Brushing has well defined peaks in the 5-15Hz range. Both brushing and shaving have high power at lower frequencies.

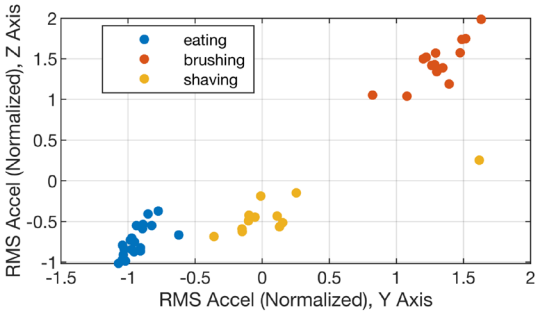

Fig. 4. Example of two features used for classification in the SVM, the RMS acceleration in y and z directions. In these experiments, the y direction is along the length of the arm and z is normal to the palm when the wrist is straight. Units are normalized by subtracting the mean and dividing by the standard deviation.

of the total daily arm activities, such as shaving, brushing teeth, drinking, scratching chin, nose picking, etc.

To test the effectiveness of using machine learning tools to classify hand activity near the mouth, experimental data was taken using a wrist-worn IMU. A total of 25 minutes of data was collected while eating, shaving, and brushing teeth. The labelled data was divided into 30 second long segments, and several features were extracted, including the mean and RMS value of acceleration, the values and frequencies of the first five peaks of the acceleration power spectral density (PSD), and the power in the 0.5-1.5 Hz, 1.5-5.0 Hz, 5.0-10.0 Hz, 10.0-15.0 Hz, and 15.0-20.0 Hz bands for each sensor axis. Figure 3 shows how the acceleration PSD in the direction of the lower arm varies for different activities. Teeth brushing has well defined peaks in the 5-15Hz range. Both teeth brushing and shaving are higher power than eating at lower frequencies. Figure 4 shows how the extracted features separate the three activities in this experiment. In particular, the RMS acceleration provides good feature separation.

The extracted features were used to train a Support Vector Machine (SVM) to classify the 30 second segments as eating or not-eating. The SVM fits the maximum margin hyperplane in the feature space by maximizing the distance between the hyperplane and the nearest data points [7]. The SVM was trained using all data and validated by 5-fold cross validation [8]. In k-fold cross validation, the samples are randomly divided into k folds. One fold is held out, and the remaining k-1 folds are used to train an SVM. This process is then repeated a total of k times, such that each group gets used once as the test group. The overall accuracy of the SVM trained using all of the data is estimated by taking the mean accuracy of the k SVMs. For the 25-minute motion dataset, the SVM achieved 98% classification accuracy, shown in Fig. 4.

VI. Conclusion

This paper has presented a novel prototype Food Intake Monitoring (FIM) system that consists of two IMUs worn on the upper and lower arms. By estimating the orientation of each IMU using the DCM, singularities are avoided, and the use of a linear estimation filter is possible. Once limb orientations are estimated, the radius and height of the hand relative to the shoulder are known, and activity that takes place in the region of the mouth can be classified as eating or non-eating. Preliminary results of classifying eating activity against shaving and tooth brushing suggest SVM is a good tool for eating detection. Future work will involve a more robust set of activities and a larger number of patients.

By detecting eating events throughout the day, the FIM system has the potential to improve dietary assessment because the data collected from it may be paired with the 24-hour dietary recall and food record methods to support more complete and accurate reporting of food intake.

References

[1] F. L. Markley and J. L. Crassidis, *Fundamentals of spacecraft attitude determination and control*. New York, NY: Springer, 2014.

[2] J. L. Crassidis, F. L. Markley, and Y. Cheng, "Survey of Nonlinear Attitude Estimation Methods," *J. Guid. Control. Dyn.*, vol. 30, no. 1, pp. 12–28, Jan. 2007.

[3] Y. Wang and R. Rajamani, "Direction cosine matrix estimation with an inertial measurement unit," *Mech. Syst. Signal Process.*, vol. 109, pp. 268–284, Sep. 2018.

[4] B. Wie, *Space Vehicle Dynamics and Control, Second Edition*. Reston ,VA: American Institute of Aeronautics and Astronautics, 2008.

[5] D. Choukroun, "Novel methods for attitude determination using vector observations," Technion - Israel Institute of Technology, 2003.

[6] M. Kok, J. D. Hol, and T. B. Schön, "Using Inertial Sensors for Position and Orientation Estimation," *Found. Trends® Signal Process.*, vol. 11, no. 1–2, pp. 1–153, Nov. 2017.

[7] K. P. Bennett and C. Campbell, "Support vector machines," *ACM SIGKDD Explor. Newsl.*, vol. 2, no. 2, pp. 1–13, Dec. 2000.

[8] M. Kuhn and K. Johnson, *Applied predictive modeling*.