

A Dynamic Partitioning Algorithm for Sip Detection using a Bottle-Attachable IMU Sensor

Henry Griffith¹, Yan Shi², Subir Biswas³

Department of Electrical and Computer Engineering
East Lansing, MI, USA

Abstract—Hydration tracking technologies are a promising tool for improving health outcomes across a variety of populations. As a non-wearable solution that is reconfigurable across containers, bottle-attachable inertial measurement unit (IMU) sensors offer numerous advantages versus alternative tracking approaches. This paper proposes a novel dynamic temporal partitioning and classification algorithm for spotting drinks within the streaming data generated by such sensors. By exploiting the distinguishing characteristics of the container's estimated inclination during drinking, the algorithm identifies candidate drink intervals for subsequent classification using a Threshold-Merge-Discard framework. The proposed approach is benchmarked against a slight variation of a previously introduced sliding window classifier for a series of experiments replicating the intended use case of the device. The new algorithm is shown to increase the true-positive detection rate by 23.7%, while reducing the number of required classification operations by more than an order of magnitude.

Keywords—Hydration management; online activity classification; dynamic time windowing; inertial measurement unit sensors

I. INTRODUCTION

Susceptibility to dehydration increases considerably with age due to a variety of factors [1]. Fluid consumption may be decreased due to reduced osmoreceptor sensitivity, dysphagia, cognitive impairment, as well as mobility restrictions. Reduced capacity of the kidneys to concentrate urine, polypharmacy, along with voluntary reductions in consumption due to incontinence further exacerbate the problem [2]. Estimates suggest that 20% to 30% of older adults are dehydrated, significantly increasing their risk of mortality, morbidity, and disability [3]. Deterioration of regulatory mechanisms may also result in hyponatremia, which has been associated with negative health outcomes such as falls [4]. Further complicating the issue, dehydration amongst the elderly is often misdiagnosed in clinical settings [5].

The large-scale ramifications of elderly dehydration are substantial, especially in developed countries with aging populations [6]. In the United States, Medicaid expenditures associated with hospital admissions for dehydration were estimated at \$5.5 billion in 2004 [7]. Recent evidence suggests that participants dehydrated at hospital admission have a six times greater chance of dying during their stay versus fully-hydrated individuals [8]. Dehydration is especially prevalent amongst residents of long-term care (LTC) facilities, thereby increasing the burden on limited caregiver resources [9].

Various interventions aimed at improving hydration amongst elderly individuals have been explored within inpatient settings. While most report positive outcomes [10], reliance upon manual monitoring and documentation of intake greatly impacts scalability and extension outside of a clinical environment. This latter limitation is especially concerning, as dehydration-related hospital readmissions are common [11], [12]. Moreover, visual estimates of consumption have been shown to overstate fluid intake [13].

Multiple technologies have been demonstrated for automated fluid consumption tracking. Approaches include containers with embedded sensing functionality (often denoted as smart-containers) [14], wearable [15], and video-based solutions [16]. Unfortunately, each class of approach is characterized by some limitation with respect to deployment amongst elderly individuals, especially in LTC facilities. Namely, economic constraints and market availability may limit the ability to procure a sufficient number and variety of augmented drinking containers. Furthermore, limited dexterity amongst the target population may prohibit the utilization of wearable solutions, while video approaches may be opposed on the basis of intrusiveness.

We have previously introduced an alternative solution for real-time hydration tracking using an attachable inertial measurement unit (IMU) sensor as shown in Fig. 1 [17]. This architecture offers similar advantages to traditional smart bottles with respect to privacy and usability, while providing reconfigurability across multiple drinking containers. Preliminary experiments demonstrated the ability of the sensor to identify drink events using a static sliding window (SSW) classifier. For a 30 second window duration with 50% overlap, an online classification accuracy of 99% was achieved for the previously gathered data set.

While promising as an initial proof-of-concept, the excessive sliding window duration considered in [17] inherently limits the ability to resolve drinks which are closely-separated in time, thereby limiting pragmatic viability without additional post-processing modifications. While such concerns may be partially addressed through reduction of window duration, SSW paradigms are still characterized by many known disadvantages, including inherent processing inefficiencies for sporadically occurring events, edge effects at event boundaries, and limited spotting precision for variable-duration events.

The research proposed herein addresses these deficiencies through the development and verification of a dynamic

temporal partitioning strategy. Namely, we perform pre-classification segmentation of the sensor data stream to identify candidate drink event intervals according to their unique inclination morphology. Candidate intervals are identified using a *Threshold-Merge-Discard (TMD)* algorithm. As the partitioning algorithm inherently discriminates against most confounding activities occurring during daily use (i.e.: transport, maintenance, etc.), the classification layer may be targeted for distinguishing drink events against actions with similar kinematics (i.e.: discharging of excess water, etc.). We verify our proposed algorithm using a newly collected data set intended to assess spotting performance for closely spaced drinks interleaved amongst typical daily living activities.

The primary contribution of this manuscript is the development and verification of the aforementioned two-stage temporal partitioning and classification algorithm. The algorithm is demonstrated to improve true-positive detection rate while dramatically reducing the number of required classifier operations versus an SSW classifier. Moreover, preliminary analysis suggests that localization error is also reduced. The effect of improved spotting localization on sip volume estimation will be explored in future work.

While the proposed bottle-attachable sensor offers a unique value proposition for the aging population as previously discussed, many core advantages (i.e.: reconfigurability across multiple containers, etc.) are broadly appealing. Moreover, the general strategies described in this manuscript for spotting sporadically occurring events of variable duration are of interest in a variety of online activity classification applications.

The remainder of the manuscript begins by providing a limited review of relevant work in the literature. Namely, we describe alternative hydration sensing architectures, as well as existing techniques for spotting variable-duration events within streaming sensor data. Next, details regarding experimental methods, including the employed hardware architecture, experimental script design, and analysis techniques are presented. The Results section provides benchmarking versus a slight variation of our previously considered SSW classifier. The manuscript concludes with a summary of our findings, along with a discussion of future research objectives.

II. RELATED WORK

A. Smart Hydration Tracking Solutions

Numerous hydration management technologies have been proposed in both the literature and commercial marketplace. While complete solutions are inherently complex cyber-physical systems, which must be cognizant of individual hydration needs, provide appropriate reminders, etc., this review focuses solely on the enabling sensing mechanisms for drink detection and volume estimation. As shown in the remainder of this section, many different sensing modalities have been considered for this application.

Solutions embedding sensing functionality within a drinking container are typically referred to as augmented or smart-containers. Various sensors have been proposed to enable sip detection and volume estimation in these products. Classic approaches perform direct measurement of the current

fluid amount using either pressure [18] or level sensors [19]. Alternatively, bottles with embedded devices for measuring the exiting flow rate have also been proposed [20]. More recent approaches place IMU sensors in either the structure or cap of the drinking vessel, and estimate sip volume as a function of bottle inclination. Suggestions for extending the utility of IMU-embedded bottles for alternative applications of benefit, such as activity tracking, have been documented [21]. With respect to our proposed approach, smart bottle solutions limit tracking to a single container. Moreover, by attaching a sensor to the exterior of the bottle, our approach offers a dry solution, thereby relieving potential durability concerns.

To address the restrictiveness imposed by augmented containers, various alternative techniques have been explored. For purposes of this review, these are organized as wearable, nearable, and contactless solutions. Amongst wearables, Amft and Tröster identified drinking events using a body sensor network consisting of IMUs placed on the upper limbs, an ear microphone, and an EMG and microphone combination configured in a throat collar [22]. Additional networks utilizing a variety of wearable inertial and acoustic sensors have also been demonstrated [23], [24]. While multi-sensor collection systems may be feasible for research applications, their complexity and restrictiveness limit practical viability versus our single sensor solution.

Subsequent work has alleviated the restrictiveness of multiple sensors, isolating functionality within a single wearable device. For example, Amft et al. used a single wrist-mounted IMU to spot drink events amongst daily living activities. This work also demonstrated the ability to discriminate between container types and fluid levels [25]. Most recently, Hamatani et al. [15] utilized the IMU sensors embedded within a Microsoft smart watch to spot and partition drink events into so-called microevents (lifting, drinking from, and releasing the bottle), and estimate drink volume. While wearable approaches are appropriate for many users, they may be excessively cumbersome for some individuals, including persons with limited dexterity and other physical limitations.



Fig. 1. Side and Back View of Sensor Attached to Bottle.

Amongst contactless solutions, Chua et al. used a Haar-like feature set to identify drinking events by identifying the gripping posture of the hand through image processing [26]. Ienaga et al. used features related to joint position estimated using a Kinect sensor to demonstrate sip recognition for service robotic applications [27]. Both approaches are characterized by the typical privacy concerns associated with deploying video sensors in daily living environments. Chiu et al. proposed estimating fill level using a phone camera placed adjacent to a drink container in a custom attachment, with temporal partitioning performed by fusing information from the embedded accelerometer [28]. In addition to the general privacy concerns associated with video collection, this method is also disadvantaged through its requirement of an optically transparent container, along with utilization of a custom apparatus to configure the phone in the required position.

Numerous wearable sensors have also been explored for hydration tracking. Proposals include integrating sensing functionality into areas where drinking containers are placed, such as coasters [29], [30]. Alternatively, container-attachable sensors, including the current work, have been demonstrated. While alternative attachable sensing modalities have been considered, such as RFID [31], the direct kinematic measurements afforded through IMU-based sensing allows for more accurate modeling of the governing fluid dynamics, thereby potentially aiding in sip volume estimation.

B. Temporal Partitioning of Streaming Sensor Data for Activity Recognition

While the literature applying IMU sensors for human activity recognition (AR) is well-established [32], the problem of spotting activities within streaming sensor data remains an area of active interest. This problem is distinguished from more fundamental work where classification is performed on pre-segmented data [33]. As even this subset of work is of considerable breadth, this section attempts only to provide a broad taxonomy of temporal partitioning approaches previously considered in the literature.

Static sliding window (SSW) techniques, in which streaming data is partitioned into fixed length intervals (W) of pre-defined overlap (p), have been heavily explored for online AR [34-36]. This approach offers simplicity on both a conceptual and implementation level. Algorithm parameters are typically chosen using application-specific empirical data. For example, Tapia et al. set the static window duration at half the average of the shortest event duration observed, thereby ensuring sufficient temporal spotting resolution [37]. Beyond application-specific considerations, windowing parameters should also be considered in conjunction with classifier design decisions, especially for methodologies employing hand-engineered feature spaces.

SSW temporal partitioning suffers from many disadvantages, including 1) inherent inefficiencies for scenarios requiring the spotting of sporadically occurring short-duration events, such as drinks, 2) performance challenges for situations where the window encompasses signals from multiple activities of interest, which may occur at both event boundaries, along with cases where the window duration exceeds the event duration, and 3) challenges for scenarios

where the window duration is less than the event duration. Visualizations of the segmentation cases described in 2) and 3) are shown in Fig. 2.

With respect to 2), the influence of window length on classification errors for fixed partitioning frameworks has been explored in the literature [38]. The coupling between the construction of the feature space and window parameters was investigated in [39], with adaptive selection of features and window parameters on a per-activity basis yielding optimal performance. As our work is targeted for the spotting of drinks, which may be highly sporadic and of variable duration, static windowing is disadvantaged relative to the dynamic partitioning approach proposed within this manuscript.

To address the limitations of SSW segmentation, a variety of adaptive approaches have been explored. For example, Laguna et al. identified window boundaries using sensor state changes (RFID and reed switches), thereby yielding event-specific dynamic window durations for in-home daily living activities [40]. As this approach requires discrete state-based sensor outputs to trigger event boundaries, it is not directly applicable for our application.

Various other techniques which dynamically segment streaming data according to some event-specific rule have been explored. For example, Junker et al. [41] used the sliding window and bottom-up algorithm, originally proposed by Keogh et al. [42], to partition estimates of the pitch and roll of the lower arm approximated by IMU sensors. While such complexity in partitioning may be mandated for wearable applications where multiple activities of interest exhibit similar signatures, the differentiation in morphology between the majority of our events of interest, as emphasized in Fig. 3., renders such complexity unnecessary for the current application. More simplistic threshold-based partitioning approaches have been suggested for both wearable [43], and vision-based [44] AR frameworks. Our work is distinguished from these in both sensor placement and application, along with the utilization of multiple post-thresholding qualifiers to further improve the efficiency and specificity of the partitioning process.

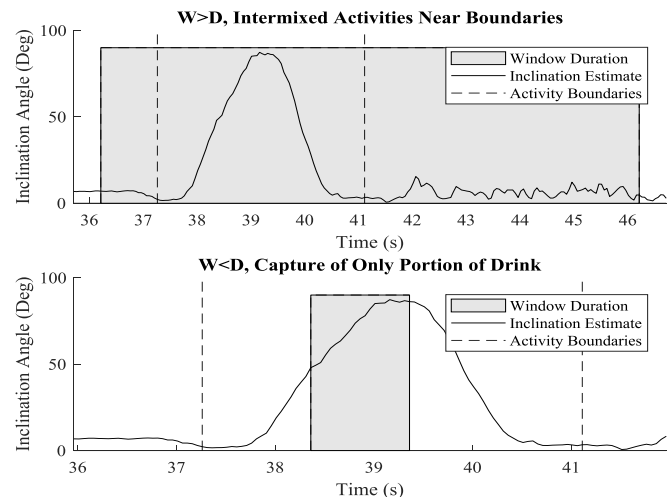


Fig. 2. Disadvantages of Static Sliding Window Architecture.

Alternative partitioning approaches have employed domain-specific sensor fusion. For example, Luckowicz et al. used acoustic intensities to segment accelerometer outputs for tracking assembly-related activities in a wood shop [45]. In relation to the current application, utilization of additional devices, such as a light sensor to indicate opening of a lid, have been proposed for providing temporal drink event markers [14]. As these and similar techniques require additional hardware, they are not suitable for integration within our proposed lightweight and retrofittable solution.

III. METHODS

A. Hardware and Pre-Processing

A wireless sensor network containing three six degree-of-freedom IMU sensors was used in all data collections. Each IMU node contains both a triaxial accelerometer (Analog Devices ADXL345), gyroscope (InvenSense IMU-3000), and IRIS Mote module. The specific configuration of each node during the various collections performed is provided in the appropriate forthcoming subsections. Only the accelerometer signal is used in the current work, with processing of the gyroscope signal targeted for future research.

Data is transmitted from each node to a MEMSIC IRIS base station through an 802.15.4 wireless link, which is interfaced to a PC through USB for subsequent data storage. Data was polled from the sensor nodes in a round-robin fashion

at a target sampling interval of 50 ms per node. All processing was performed using MATLAB. For all configurations in which a node was connected to the bottle, the relationship between the local sensor coordinate frame and bottle geometry is as follows: 1) the positive x -component of the sensor was aligned vertically along the bottle's surface, yielding a static output value corresponding to the Earth's gravitational constant when placed vertically on a surface (i.e.: $\mathbf{a} = g\hat{\mathbf{x}}$), and 2) the y and z -components were oriented parallel and normal to the bottle's surface, with sign convention defined according to a traditional right-handed framework. A visualization of the sensor coordinate axes was provided in Fig. 1. It should be noted that while care was taken to maintain the stated orientation during all trials, variations may have occurred during the experiments as part of the handling process.

Each accelerometer output was initially smoothed using a 2-sample moving average filter, and subsequently resampled using MATLAB's *resample* function to account for variability in the base station polling interval. After conditioning, the inclination angle of the bottle was estimated under the commonly employed assumption of minimal dynamic acceleration as specified in (1), where a_j denotes the j^{th} component of the accelerometer output.

$$\hat{\theta} = \frac{\sqrt{a_y^2 + a_z^2}}{a_x} \quad (1)$$

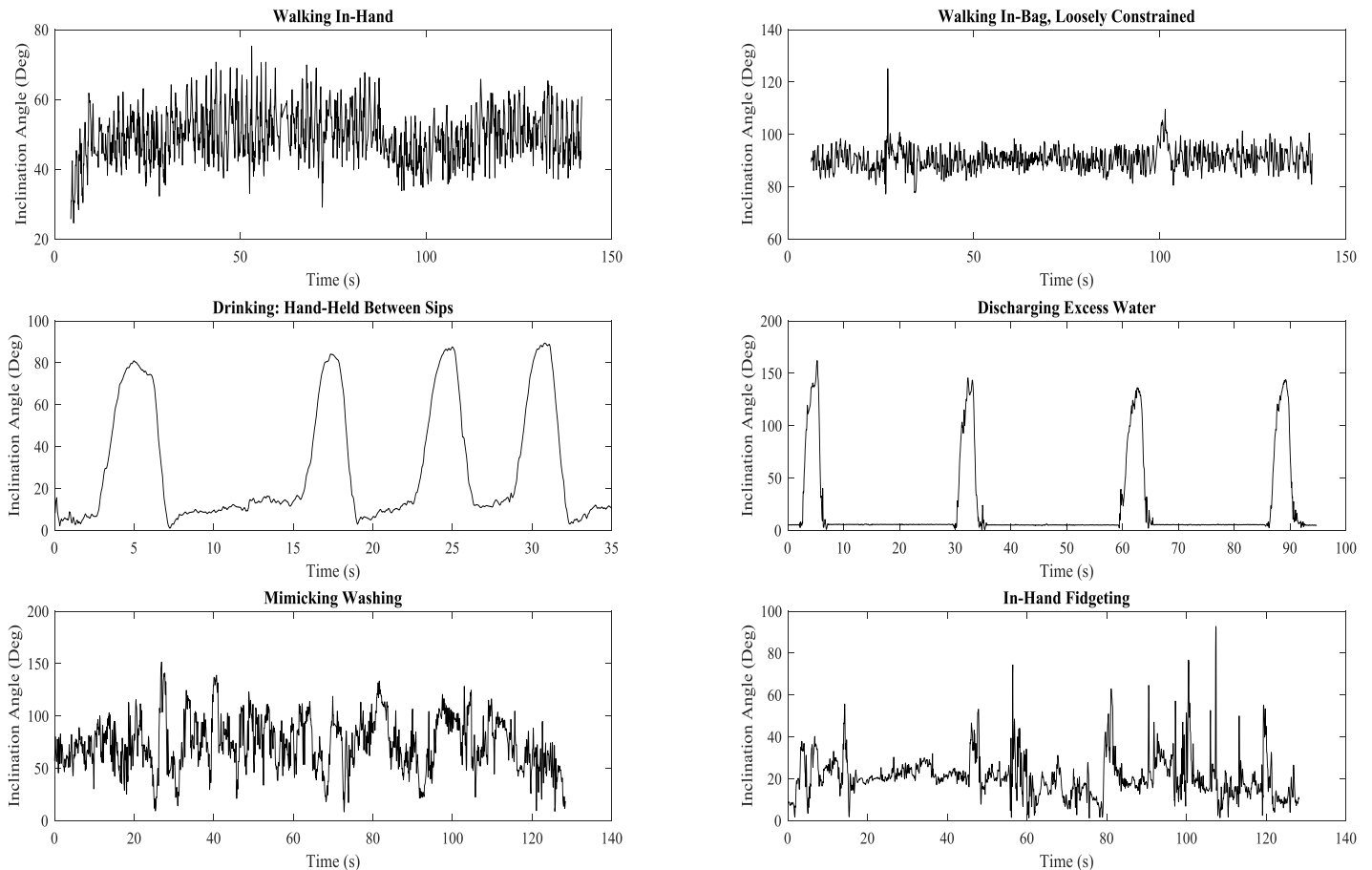


Fig. 3. Sample Realizations of Various Activities Considered.

B. Data Collectons

1) *Overview*: Experiments were designed to mimic the intended use case of the device. The following general activity classes were identified for consideration: 1) maintenance activities (i.e.: discharging excess fluid, washing, etc.), 2) transport activities (i.e.: carrying in-hand, etc.), 3) use-base handling (drinking, fidgeting, etc.), and 4) stationary placement. While the detachable nature of the sensor would ideally result in the removal of the device during maintenance activities, these were included for all current analysis.

Experiments were conducted by multiple participants to assess inter-individual variability in both handling and drinking style. Participants were directed to perform each action according to their own personal preferences. The data collection was divided into three separate sessions denoted as follows: i) Training Set (*TS*) Collection, ii) Temporal Resolution Testing Collection (*TR*), and iii) Interleaved Daily Living Testing Collection (*DL*). A brief description of each collection is provided below. The *TS* collection was completed by seven individuals, while the testing collections were completed by only five of the original seven.

2) *Training set (TS) collection*: To support rapid acquisition of high-quality training data, individual collections were conducted for each activity described in Table 1. For all events other than drinking and discharging excess water, 35 minutes of data (5 mins./participant) was collected. For drinking and discharge, 84 events (12/participant) were recorded for each activity. Two sensors were attached to the bottle during all activities in a position intended to minimize interference with handling and drinking. The first node, hereby denoted as the bottom sensor, was placed below the hinge at the bottom of the bottle as shown in Fig. 1. The second sensor was placed midway up the bottle opposite the drinking hand of each participant. The third sensor was used only for marking the initiation and termination of drink events as described in Section III.B.5. Training was performed using only bottom sensor data, with the exploration of middle sensor data reserved for future work exploring performance robustness with respect to position.

Conducting dedicated training collections where participants perform only a single activity of interest at a time offers notable advantages, including simplifying the assignment of ground-truth (GT) labels (versus data containing multiple interleaving activities). Moreover, single-activity trials simplify participant instruction, thereby ensuring data quality. Isolated training collections have also been employed in related work for similar motivations (i.e.: [25]). This strategy is not without disadvantage, as it eliminates the direct deployment of models exploiting temporal variations within the activity sequence (i.e.: HMMs, LSTMs, etc.). Sample waveforms of each activity are depicted in Fig. 3.

3) *Temporal resolution (TR) testing collection*: A dedicated testing collection was conducted to assess the capacity of the algorithm to resolve closely spaced drinks.

TABLE I. DAILY USE ACTIVITIES CONSIDERED

Activity ID	Description
Walking: Bottle In-Hand (W-IH)	Participants walked on both flat ground and stairs in a repeated loop to remain in range of base station with bottle held in hand at an unspecified orientation/grip
Walking: Bottle In-Bag (W-IB)	Participants walked in same loop at W-IH, but with bottle placed in a bag supporting vibrational, rotational, and translational degrees of freedom. Instructions for holding the bag were not specified to participants
Walking: Bottle In-Hand (W-IB-R)	Same as W-IB-L, but with additional objects placed in the bag to restrict rotational and translational degrees of freedom
Stationary Placement (S)	Bottle placed stationary in various orientations
Transport: In-Car (T-IC)	Bottle placed in various locations (floorboard, seats, etc.) in vehicle traveling in various environments (highway, city, etc.)
Fidgeting (F)	Participants held bottle in hand and were instructed to mimic activities which may occur while seated (i.e.: daydreaming, fidgeting, engaging in conversation, etc.)
Mimic Washing (MW)	Participants mimicked washing the bottle in a sink
Drinking: (D)	Participants completed 12 drinks each while standing, with the bottle retained in-hand between drinks
Discharge Excess Water (DEW)	Participants discharged excess water 12 times from various initial fill levels (full, half, and quarter filled) into a sink

Four target inter-drink spacings ($\{2, 5, 10, 20\}$ s) were considered. To avoid spilling, participants retained the bottle in-hand between drinking commands, which were provided verbally by the experimental proctor. Data was collected in a series of four trials containing six drinks each (two trials containing spacings of two and 10 s, and the other two containing 5 and 20 s spacings), corresponding to 120 total drinks across the *TR* set. This information is summarized in Table 2.

TR collections utilized a bottom sensor as previously described, a sensor placed on the wrist of the drinking hand of the participant (to be explored in future work), along with a sensor held in the hand of the proctor. Similar to the *TS* collection, this latter sensor was shaken to mark the initiation and termination of the drinking event for GT labeling. A visualization of the wrist and sensor outputs for a 2/10 s spacing trial is provided in Fig. 4.

4) *Interleaved daily living (DL) testing collection*: Further experiments were conducted to ensure algorithm viability for truncated daily living scenarios consisting of interleaved activities considered in the training collection. A series of four experiments were conducted—two employing transport in-hand, and two employing in-bag transport at two different orientations (vertical and horizontal). Each experiment contained 8 drinks with varying inter-drink separation.

Summary information for the *DL* collection is also provided in Table 2. The experiment utilized an identical hardware configuration as described for *TR* testing. A visualization of the estimated bottle inclination over the experiment is shown later in the manuscript (Fig. 6), after introduction of the proposed dynamic partitioning strategy in Section III.C.2.

5) *Ground-truth labeling*: The proctor was instructed to shake a hand-held sensor at the initiation and termination of the lifting motion for each drink. Labels were then assigned by applying an empirically determined threshold to the magnitude of the acceleration signal, a_c , with the static acceleration due to gravity removed as shown in (2).

$$A[n] = |a_c - 1| > \tau \quad (2)$$

For all samples exceeding the threshold in the local neighborhood of the j^{th} drink event (determined visually), GT values for the beginning (t_s^j) and end (t_e^j) of the drink were assigned as specified in (3) and (4), respectively.

$$t_s^j = \inf\{n \mid A^j[n] > \tau\} \quad (3)$$

$$t_e^j = \sup\{n \mid A^j[n] > \tau\} \quad (4)$$

The consistency of GT estimates across drinks is inherently limited by the subjectivity of the proctor marking. Due to this limitation, the inference which may be drawn from subsequent measurements of localization error is restricted.

C. Algorithm Development

1) *Overview*: Binary event detection schemes employing temporal partitioning with subsequent classification may be conceptualized as a three-phase processing workflow. The preliminary step involves temporal partitioning of streaming data, hereby denoted as $\{x_i\}$, where i is a time index corresponding to the sensor timestamp, by some mapping function ψ as denoted in (5).

$$\psi: \{x_i\} \rightarrow \{\mathbf{d}^m\} \quad (5)$$

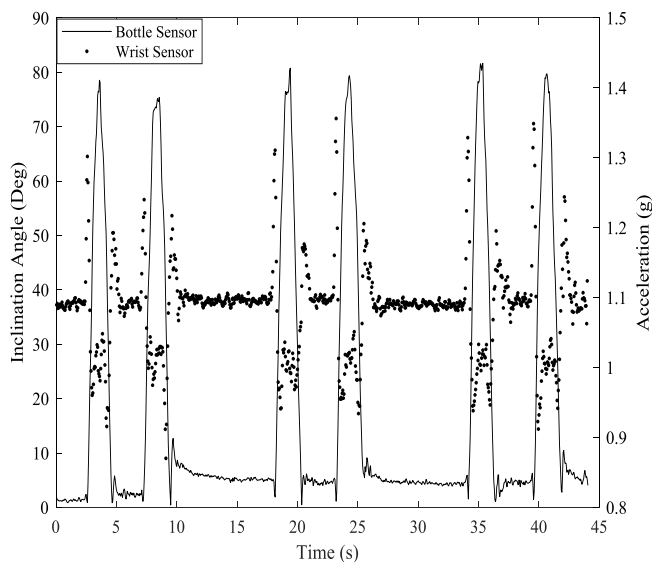


Fig. 4. Bottle and Wrist Signals for Temporal Resolution Testing Trial.

TABLE II. SUMMARY OF TESTING DATA COLLECTIONS

Collection ID	Interleaving Activities Considered	Inter-Drink Spacings Considered	Total Drinks Per Subject/Total
TR	• In-Hand Holding	{2,5,10,20} s	24 / 120
DL	• In-Hand Holding • W-IH • W-IB • DEW • MW	{2, 10} s	32/160

where $\mathbf{d}^m = \{x_1, x_2, \dots, x_n\}^m$ is the m^{th} data partition, and x_1^m and x_n^m are the starting and ending data points. For SSW approaches, ψ is a buffering process which groups input data into fixed duration intervals of specified overlap (i.e.: n^m is constant $\forall m$). For dynamic partitioning strategies, ψ exploits some characteristic of either the sensor or activity space of interest to produce variable duration partitions. Classification is performed by some learned function φ , which performs the mapping denoted in (6)

$$\varphi: f(\mathbf{d}^m) \rightarrow L^m \quad (6)$$

where $L^m \in \{0,1\}$ is a binary indicator of the presence of the event in the m^{th} partition, and f is a function computed on each data partition. For end-to-end architectures, f is the identity function (i.e.: data is fed directly into the classifier). For traditional classifiers employing hand-engineered feature spaces, f is a mapping of the raw data to the designed feature space. The detection process may require potential post-processing, especially for schemes employing SSW segmentation with considerable overlap.

2) *Proposed dynamic partitioning strategy*: As exhibited in Fig. 3, the inclination signal follows a convex morphology during drinking events. Our proposed dynamic partitioning strategy seeks to identify time intervals containing candidate drink signals by exploiting this distinguished inclination signature. This process is described in the subsequent paragraph, and presented in pseudocode in Fig. 5.

To begin partitioning of the input stream, an amplitude threshold is applied to the inclination signal on a per-sample basis ($\theta_{min} = 12^\circ$). Next, adjacent intervals of samples exceeding the threshold which are separated by less than a merge parameter ($s = 3$ samples) are combined. Merging is conducted to ensure capturing of the entire drink motion. The merging process yields candidate data partitions $\hat{\theta}^j$, with beginning and ending timestamps denoted as \hat{t}_s^j and \hat{t}_e^j . Partitions with a maximum inclination value or inclination range falling below a threshold ($A_{min} = 50^\circ$ and $R_{min} = 30^\circ$, respectively), or duration falling outside of a specified range (0.5–6 s) are discarded. This qualifying process is intended to discard events not exhibiting the desired inclination signature (i.e.: stationary placements at non-vertical orientations, etc.), which is mandated due to the collection of data even when the lid is closed. The result of applying the algorithm to a *DL* data trial is shown in Fig. 6, which shows both the data partition outputs of the *TMD* algorithm, along with the GT intervals.

Temporal Partitioning Pseudocode

Input: Accelerometer-Based Inclination Estimate,
 $\hat{\theta}[n], n \in \mathbf{N} = \{1, 2, \dots, N\}$

Output: Ordered pairs estimating the start/stop of candidate drink intervals, $\{\{\hat{t}_s^j, \hat{t}_e^j\}^j\}$

Parameters: Point Amplitude Threshold, θ_{\min} ,
Merge Parameter, s ,
Duration Criteria, d_{\min}, d_{\max} ,
Amplitude Criteria, A_{\min} ,
Range Criteria, R_{\min} ,

Threshold $\hat{\theta}[n], N^* \subset \mathbf{N} = \{n \mid \hat{\theta}[n] > \theta_{\min}\}$

Merge resultant thresholded subset, N^* , to form candidate output set \mathbf{N}'_D

Initialize $\mathbf{N}_D = \mathbf{N}'_D = \{\}$

Set $\hat{t}_s^1 = \inf(N^*), j = 1$

for $k = \hat{t}_s^1 + 1 : |N^*|$

 if $(N^*[k] - N^*[k-1]) > s$

$\hat{t}_e^j = N^*[k-1]$

$\mathbf{N}'_D = \mathbf{N}'_D \cup (\hat{t}_s^j, \hat{t}_e^j)^j$

$j = j + 1$

$\hat{t}_s^j = N^*[k]$

 end if

end for

Discard events of insufficient maximum amplitude or duration range in \mathbf{N}'_D to form output set \mathbf{N}_D

Set $d_{\text{allow}} = [d_{\min}, d_{\max}]$

for $j = 1 : |\mathbf{N}'_D|$

 if $\{(\hat{t}_e^j - \hat{t}_s^j) \in d_{\text{allow}} \ \& \ \max(\hat{\theta}^j) > A_{\min} \ \& \ \text{range}(\hat{\theta}^j) > R_{\min}\}$

$\mathbf{N}_D = \mathbf{N}_D \cup (\hat{t}_s^j, \hat{t}_e^j)^j$

 end if

end for

Return candidate drinking events, $\mathbf{N}_D = \{(\hat{t}_s^j, \hat{t}_e^j)^j\}$

Fig. 5. Threshold-Merge-Discard (TMD) Dynamic Partitioning Pseudocode.

3) *Classification architecture:* As the TMD algorithm was designed to discard most confounding daily living activities, the subsequent classification process was targeted to differentiate solely between drinks and other events exhibiting a convex inclination (i.e. excess discharges, etc.). Data visualization and domain knowledge were used to develop a candidate feature set suitable for distinguishing these events under normal operation (i.e. users not attempting to spoof the device). As drinking is subject to somatosensory feedback and involves careful handling to avoid spills, it was hypothesized that the motion should be more controlled versus discharge and other pouring events away from the mouth. To reflect this hypothesis, features describing the maximum inclination angle, mean inclination rate through the maximum angle, and residual energy after smoothing were used as defined in (7)–(9)

$$\hat{\theta}^j_{\max} = \max(\hat{\theta}(t_s^j : t_e^j)) \quad (7)$$

$$\frac{d}{dt} \hat{\theta}^j_{ra} = \text{mean}(\hat{\theta}(t_s^j : t_{\max}^j)) \quad (8)$$

$$\Delta e^j = \sum_{t_s^j}^{t_e^j} (\hat{\theta}[k] - s(\hat{\theta}[k]))^2 \quad (9)$$

where $s(\cdot)$ is a smoothing operation implemented as a third-order Savitzky-Golay filter with a nine-sample frame length, and t_{\max}^j is the time index of the maximum inclination angle. A scatter plot showing the clustering of drink and discharge training instances in this feature space is depicted in Fig. 7.

Training data (D and DEW only) was partitioned using five-fold cross-validation to avoid overfitting. A variety of classifier models were evaluated using MATLAB's *Classification Learner* Application. Cross-validation accuracy exhibited minimal variation across the various models considered (K-NNs: 98.2% for fine clustering, SVMs: 98.2% for various kernels (linear, quadratic, etc.), etc.). A linear SVM was used for all subsequent analysis.

The proposed algorithm was benchmarked against a slight variation of our previously considered technique [17]. Partitioning was performed using an SSW scheme ($W = 3s, p = 75\%$). A slightly modified version of the four-element feature space used in [17] was employed as specified in (10)–(13).

$$\hat{\theta}^m_{\text{range}} = \text{range}(\hat{\theta}[t_i^m : t_f^m]) \quad (10)$$

$$N^m = \text{nnz}(\hat{\theta}[t_i^m : t_f^m] > \theta_{\min}) \quad (11)$$

$$\hat{\theta}^m_{\text{mean}} = \text{mean}(\hat{\theta}[t_i^m : t_f^m]) \quad (12)$$

$$S^m = \frac{\text{argmax}(\hat{\theta}[t_i^m : t_f^m])}{N^m - \text{argmax}(\hat{\theta}[t_i^m : t_f^m])} \quad (13)$$

where nnz is a function counting the number of non-zero samples satisfying the threshold criteria, and t_i^m and t_f^m are the initial and final timestamps in the m^{th} window. Slight modifications of the feature space were necessary to reflect utilization of the inclination estimate in the current work (versus the axial component of acceleration in the prior). Moreover, window duration was reduced (to the mean duration of training instances) and percent overlap was increased to improve the temporal resolving capacity of the algorithm versus previously considered settings.

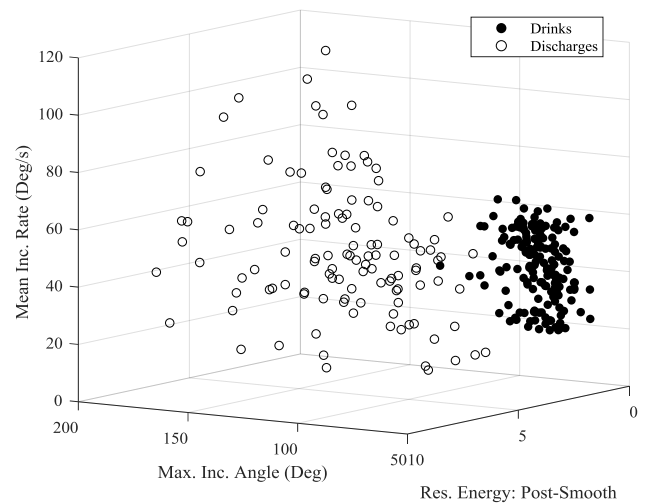


Fig. 6. Scattering of Drink and Discharge Training Instances.

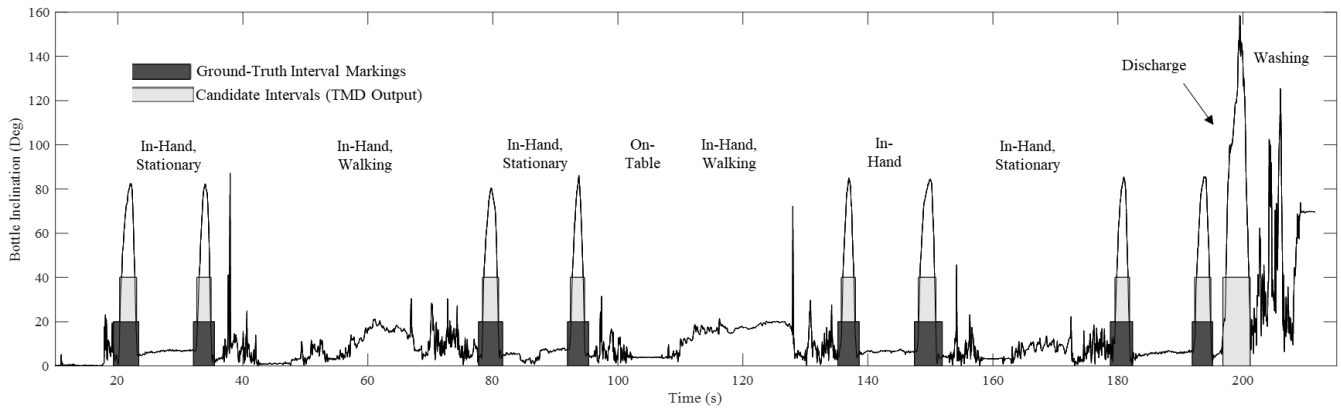


Fig. 7. Example *DL* Output with *TMD* and GT Drink Interval Labels.

Features were computed across all activity classes excluding drink and discharge events by sliding a window using the specified *SSW* parameters across the training data. For pour and drink events, the window was centered at the midpoint of the GT interval label. A cubic SVM classifier (chosen to maximize cross-validation accuracy) was trained using five-fold cross validation, yielding an average accuracy of 97.5%. Adjacent windows classified as containing drinks were merged into a single observation interval in post-processing.

4) *Analysis metrics*: Performance was quantified by first mapping the midpoint of each estimated drink interval to the nearest GT interval. Sets representing the underlap (U^j) and overlap (O^j) between the estimates were defined using the non-commutative set difference operator. Localization error was then computed as specified in (16), where $|\cdot|$ denotes the set cardinality operator.

$$E^j = \frac{(|U^j| + |O^j|)}{|[t_s^j, t_{s+1}^j, \dots, t_e^j]|} \quad (14)$$

To account for the expected variability in GT marking, successful detection was declared when the normalized intersection between the estimate and GT interval exceeded 50% of a single drink. It should be noted that both the *SSW* and *TMD* algorithms were anticipated to produce some error for the ideal GT marking protocol used herein. For the prior, the post-classification merging of adjacent windows is expected to produce overestimations. In contrast, thresholding to the minimum inclination angle in *TMD* does not necessarily allow for capturing of transport to and from the mouth, thereby resulting in potential underestimations. As consistency in GT estimates is limited, potential inference regarding the estimated localization error is restricted.

IV. RESULTS

1) *TR testing*: Both the *TMD* and *SSW* algorithms detected each of the 120 drinks in the *TR* experiments. Total localization error for *TMD* was $36.3 \pm 6.2\%$ (mean \pm standard deviation), versus $59.6 \pm 23.7\%$ for *SSW*. Error sources were consistent with those hypothesized based upon the mechanism of each algorithm as described in the prior section (average overlap of *SSW*: 58.9%, average underlap of

TMD: 36.3%). The total number of classifications performed for *TMD* processing was 120, versus 1,749 for *SSW*.

2) *DL testing*: The *TMD* algorithm detected 162 drinks through 172 classification operations across the *DL* experiments. Of these detections, 160 corresponded to true positives, with two false positives produced (True-Positive Rate (TPR): 98.8%). Total observed localization error was $31.4 \pm 23.1\%$. Consistent with *TR* experiments, localization errors largely resulted from underestimates of the GT interval (29.2% average).

In contrast, the *SSW* algorithm detected 197 drinks through 4,310 classification operations. Of these, 148 were true positives, 43 were false positives, and six contained unresolved adjacent drinks (i.e.: two drinks in one interval), corresponding to a TPR of 75.1%. Total observed localization error was $65.3 \pm 34.0\%$, with distributions for both testing trials shown in Fig. 8.

SSW error was again dominated by overestimation (63.5% avg.). Performance statistics for the *DL* experiments are consolidated in Table 3. Examples of error modes associated with *SSW* classification are depicted in Fig. 9.

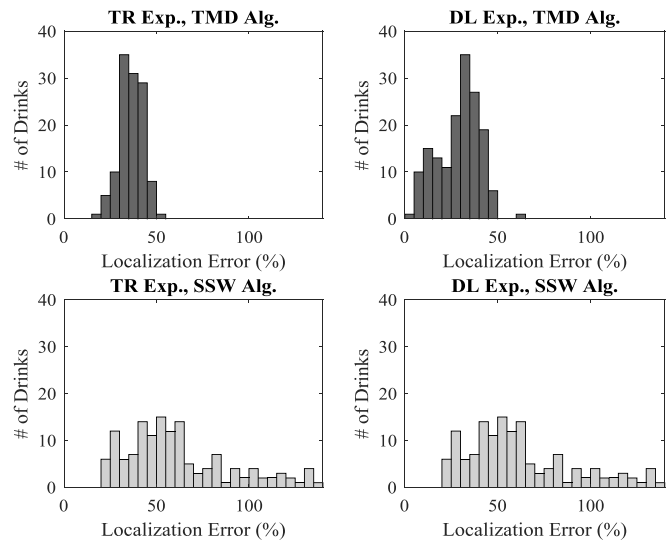


Fig. 8. Localization Error Distributions.

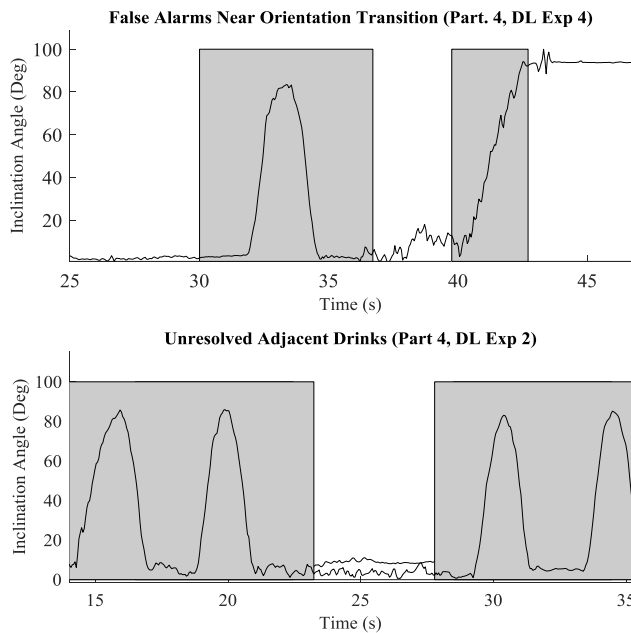


Fig. 9. Example Error Modes, DL Experiments, SSW Algorithm.

TABLE III. SUMMARY OF DL TESTING PERFORMANCE

Algorithm ID	True Positive Detection Rate	Mean Localization Error	Total # of Classifications
TMD	98.8%	31.4%	172
SSW	75.1%	65.3%	4,310

V. CONCLUSIONS AND FUTURE WORK

A novel dynamic temporal partitioning and classification algorithm for drink spotting was proposed herein. This approach is designed for implementation on streaming accelerometer data generated from a bottle-attachable IMU sensor. Benchmarked against a slightly modified version of our previously introduced static sliding window classifier, the algorithm was demonstrated to improve sip detection performance while reducing computational overhead. Namely, for a series of simulated daily living activities containing 160 intermixed drinks, true-positive detection rate was improved from 72.9% to 98.8%, while the total number of required classification operations was decreased from 4,310 to 172. Preliminary analysis also suggests improved spotting precision, although inference is limited by the subjectivity of the employed GT labeling process.

Further investigation should be conducted to assess potential trade-offs between the design of the individual stages of the proposed algorithm. Namely, the current implementation imposes several qualifying criteria on the inclination signal in the discard stage of partitioning. These could be relaxed in alternative implementations, with discrimination against the target activities for which the criteria were implemented instead performed through classification. While this approach would increase computational overhead, it would likely

improve generalization for larger data sets including more diverse drinks.

In addition to exploring these trade-offs, future work will investigate the relationship between the employed drink spotting technique and the resulting volume estimations. Moreover, exploration of performance robustness with respect to sensor position, along with comparisons with wrist-worn IMU data will be conducted using the data gathered within these experiments. Finally, the utilization of training data obtained from daily-use scenarios will be investigated to support the deployment of models exploiting temporal dependencies within the event sequence.

REFERENCES

- [1] M. El-Sharkawy, O. Sahota, R. J. Maughan, and D. N. Lobo, "The pathophysiology of fluid and electrolyte balance in the older adult surgical patient," *Clin Nutr*, vol. 33, no. 1, pp. 6–13, Feb. 2014.
- [2] M. N. Begum and C. S. Johnson, "A review of the literature on dehydration in the institutionalized elderly," *e-SPEN, the European e-Journal of Clinical Nutrition and Metabolism*, vol. 5, no. 1, pp. e47–e53, 2010.
- [3] H. J. Miller, "Dehydration in the older adult," *Journal of gerontological nursing*, vol. 41, no. 9, pp. 8–13, 2015.
- [4] F. Gankam Kengne, C. Andres, L. Sattar, C. Melot, and G. Decaux, "Mild hyponatremia and risk of fracture in the ambulatory elderly," *QJM*, vol. 101, no. 7, pp. 583–588, Jul. 2008.
- [5] D. R. Thomas, S. H. Tariq, S. Makhdomm, R. Haddad, and A. Moinuddin, "Physician Misdiagnosis of Dehydration in Older Adults," *Journal of the American Medical Directors Association*, vol. 4, no. 5, pp. 251–254, Sep. 2003.
- [6] M. Frangeskou, B. Lopez-Valcarcel, and L. Serra-Majem, "Dehydration in the elderly: A review focused on economic burden," *The journal of nutrition, health & aging*, vol. 19, no. 6, pp. 619–627, 2015.
- [7] H. Xiao, J. Barber, and E. S. Campbell, "Economic burden of dehydration among hospitalized elderly patients," *Am J Health Syst Pharm*, vol. 61, no. 23, pp. 2534–2540, Dec. 2004.
- [8] A. M. El-Sharkawy et al., "Hydration and outcome in older patients admitted to hospital (The HOOP prospective cohort study)," *Age and ageing*, vol. 44, no. 6, pp. 943–947, 2015.
- [9] A. M. Namasivayam-MacDonald et al., "Inadequate fluid intake in long term care residents: prevalence and determinants," *Geriatric Nursing*, vol. 39, no. 3, pp. 330–335, May 2018.
- [10] D. Bunn, F. Jimoh, S. H. Wilsher, and L. Hooper, "Increasing fluid intake and reducing dehydration risk in older people living in long-term care: a systematic review," *Journal of the American Medical Directors Association*, vol. 16, no. 2, pp. 101–113, 2015.
- [11] E. Messaris et al., "Dehydration is the most common indication for readmission after diverting ileostomy creation," *Diseases of the Colon & Rectum*, vol. 55, no. 2, pp. 175–180, 2012.
- [12] M. A. Khan, F. S. Hossain, Z. Dashti, and N. Muthukumar, "Causes and predictors of early re-admission after surgery for a fracture of the hip," *The Journal of bone and joint surgery. British volume*, vol. 94, no. 5, pp. 690–697, 2012.
- [13] J. F. Kreutzer, M. Pfitzer, and L. T. D'Angelo, "Accuracy of caring personnel in estimating water intake based on missing liquid in drinking vessels," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 4682–4685.
- [14] G. Zhang, R. Xu, Y. Jiang, and C.-C. Chiang, "Smart cup, drinking amount detecting method for smart cup and system therefor," *US20180045547A1*, 15-Feb-2018.
- [15] T. Hamatani, M. Elhamshary, A. Uchiyama, and T. Higashino, "FluidMeter: Gauging the Human Daily Fluid Intake Using Smartwatches," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, p. 113:1–113:25, Sep. 2018.

- [16] J.-L. Chua, Y. C. Chang, M. H. Jaward, J. Parkkinen, and K.-S. Wong, "Vision-based hand grasping posture recognition in drinking activity," in *Intelligent Signal Processing and Communication Systems (ISPACS), 2014 International Symposium on*, 2014, pp. 185–190.
- [17] B. Dong, R. Gallant, and S. Biswas, "A self-monitoring water bottle for tracking liquid intake," in *Healthcare Innovation Conference (HIC), 2014 IEEE*, 2014, pp. 311–314.
- [18] M. Marjanovic and I. Marjanovic, "Hydration Monitor," US20140311239A1, 23-Oct-2014.
- [19] A. Hambrock et al., "Wireless drink container for monitoring hydration," CA2979209A1, 15-Sep-2016.
- [20] R. Gellibolian and J. Stump, "Water bottle with flow meter," WO2013181455A1, 05-Dec-2013.
- [21] G. Sweeney, C. McCluskey, and J. W. Pfeiffer, "Activity and volume sensing beverage container cap system," US9320375B2, 26-Apr-2016.
- [22] O. Amft and G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artificial intelligence in medicine*, vol. 42, no. 2, pp. 121–136, 2008.
- [23] C. Merck, C. Maher, M. Mirtchouk, M. Zheng, Y. Huang, and S. Kleinberg, "Multimodality sensing for eating recognition," in *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2016, pp. 130–137.
- [24] M. Mirtchouk, C. Merck, and S. Kleinberg, "Automated estimation of food type and amount consumed from body-worn audio and motion sensors," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 451–462.
- [25] O. Amft, D. Bannach, G. Pirkl, M. Kreil, and P. Lukowicz, "Towards wearable sensing-based assessment of fluid intake," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on*, 2010, pp. 298–303.
- [26] J.-L. Chua, Y. C. Chang, M. H. Jaward, J. Parkkinen, and K.-S. Wong, "Vision-based hand grasping posture recognition in drinking activity," in *Intelligent Signal Processing and Communication Systems (ISPACS), 2014 International Symposium on*, 2014, pp. 185–190.
- [27] N. Ienaga, Y. Ozasa, and H. Saito, "Eating and Drinking Recognition via Integrated Information of Head Directions and Joint Positions in a Group," in *ICPRAM*, 2017, pp. 527–533.
- [28] M.-C. Chiu et al., "Playful bottle: a mobile social persuasion system to motivate healthy water intake," in *Proceedings of the 11th international conference on Ubiquitous computing*, 2009, pp. 185–194.
- [29] G. W. Lamb, "Beverage Fill Level Detection and Indication," US20080083475A1, 10-Apr-2008.
- [30] A. B. Chan and R. Scaer, "Hydration Tracking Coaster with BLE Android App," 2018.
- [31] A. Jayatilaka and D. C. Ranasinghe, "Towards unobtrusive real-time fluid intake monitoring using passive UHF RFID," in *RFID (RFID), 2016 IEEE International Conference on*, 2016, pp. 1–4.
- [32] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.
- [33] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive and mobile computing*, vol. 10, pp. 138–154, 2014.
- [34] M. Stikic, T. Huynh, K. Van Laerhoven, and B. Schiele, "ADL recognition based on the combination of RFID and accelerometer sensing," in *Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on*, 2008, pp. 258–263.
- [35] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *International Conference on Pervasive Computing*, 2004, pp. 1–17.
- [36] N. C. Krishnan and S. Panchanathan, "Analysis of low resolution accelerometer data for continuous human activity recognition," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 3337–3340.
- [37] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *International conference on pervasive computing*, 2004, pp. 158–175.
- [38] T. Gu, Z. Wu, X. Tao, H. K. Pung, and J. Lu, "epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition," in *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, 2009, pp. 1–9.
- [39] T. Huynh and B. Schiele, "Analyzing Features for Activity Recognition," in *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-aware Services: Usages and Technologies*, New York, NY, USA, 2005, pp. 159–163.
- [40] J. O. Laguna, A. G. Olaya, and D. Borrajo, "A dynamic sliding window approach for activity recognition," in *International Conference on User Modeling, Adaptation, and Personalization*, 2011, pp. 219–230.
- [41] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recognition*, vol. 41, no. 6, pp. 2010–2024, 2008.
- [42] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, 2001, pp. 289–296.
- [43] C. Lee, X. Yangsheng, Online, interactive learning of gestures for human/robot interfaces, in: N. Caplan, C.G. Lee (Eds.), *ICRA 1996: Proceedings of the IEEE International Conference on Robotics and Automation*, of IEEE Robotics and Automation Society, vol. 4, IEEE Press, New York, 1996, pp. 2982–2987.
- [44] P. Morguet, Stochastic modeling of image sequences for the segmentation and recognition of dynamic gestures, Ph.D. Thesis, Technische Universität München, 2000.
- [45] P. Lukowicz et al., "Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers," in *Pervasive Computing*, 2004, pp. 18–32.