

A CONTAINER-ATTACHABLE INERTIAL
SENSOR FOR
REAL-TIME HYDRATION TRACKING

By

Henry Griffith

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Electrical Engineering – Doctor of Philosophy

2019

ABSTRACT

A CONTAINER-ATTACHABLE INERTIAL SENSOR FOR REAL-TIME HYDRATION TRACKING

By

Henry Griffith

The underconsumption of fluid is associated with multiple adverse health outcomes, including reduced cognitive function, obesity, and cancer. To aid individuals in maintaining adequate hydration, numerous sensing architectures for tracking fluid intake have been proposed. Amongst the various approaches considered, container-attachable inertial sensors offer a non-wearable solution capable of estimating aggregate consumption across multiple drinking containers. The research described herein demonstrates techniques for improving the performance of these devices.

A novel sip detection algorithm designed to accommodate the variable duration and sparse occurrence of drinking events is presented at the beginning of this dissertation. The proposed technique identifies drinks using a two-stage segmentation and classification framework. Segmentation is performed using a dynamic partitioning algorithm which spots the characteristic inclination pattern of the container during drinking. Candidate drinks are then distinguished from handling activities with similar motion patterns using a support vector machine classifier. The algorithm is demonstrated to improve true positive detection rate from 75.1% to 98.8% versus a benchmark approach employing static segmentation.

Multiple strategies for improving drink volume estimation performance are demonstrated in the latter portion of this dissertation. Proposed techniques are verified through a large-scale data collection consisting of 1,908 drinks consumed by 84 individuals over 159 trials. Support vector machine regression models are shown to improve per-drink estimation accuracy versus the prior

state-of-the-art for a single inertial sensor, with mean absolute percentage error reduced by 11.1%. Aggregate consumption accuracy is also increased versus previously reported results for a container-attachable device.

An approach for computing aggregate consumption using fill level estimates is also demonstrated. Fill level estimates are shown to exhibit superior accuracy with reduced inter-subject variance versus volume models. A heuristic fusion technique for further improving these estimates is also introduced herein. Heuristic fusion is shown to reduce root mean square error versus direct estimates by over 30%. The dissertation concludes by demonstrating the ability of the sensor to operate across multiple containers.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	ix
Chapter 1 : Introduction	1
1.1 Motivation.....	1
1.2 Proposed Solution.....	2
1.3 Summary of Research Objectives.....	3
1.3.1 Summary of Sip Detection Problem	4
1.3.2 Summary of Volume Estimation Problem.....	5
1.3.3 Generalization to Additional Drinking Containers	7
Chapter 2 : Related Work	8
2.1 Introduction.....	8
2.2 Review of Hydration Tracking Sensors.....	8
2.2.1 Augmented Containers.....	8
2.2.2 Multi-Sensor Wearable Consumption Trackers.....	10
2.2.3 Single Sensor Wearable Consumption Trackers.....	14
2.2.4 Contactless and Nearable Consumption Trackers	19
2.2.5 Prior Research for Attachable IMU Sensors.....	20
Chapter 3 : A Dynamic Partitioning Algorithm for Improved Sip Detection.....	23
3.1 Introduction.....	23
3.2 Partitioning Strategies for Online Activity Classification	24
3.3 Collection Hardware	27
3.4 Signal Preprocessing.....	28
3.5 Data Collection	29
3.5.1 Overview	29
3.5.2 Training Colleciton	30
3.5.3 Temporal Resolution Testing Collection	31
3.5.4 Simulated Daily Living Test Collection	32
3.5.5 Ground-Truth Labeling.....	33
3.6 Algorithm Development	34
3.6.1 Overview.....	34
3.6.2 Dynamic Partitioning Strategy.....	35
3.6.3 Classification Algorithm.....	37
3.6.4 Performance Metrics	40
3.7 Results.....	40
3.7.1 TR Testing	40
3.7.2 DL Testing	41
3.8 Conclusions and Future Work	42

Chapter 4 : The Inclination Signature Feature Set.....	45
4.1 Introduction.....	45
4.2 Data Collection	46
4.3 Pre-processing and Drink Segmentation.....	47
4.4 Microevent Partitioning Strategy.....	48
4.5 Feature Engineering.....	53
4.6 Summary and Future Work.....	57
Chapter 5 : Drink Volume Estimation Using Regression Models.....	59
5.1 Introduction.....	59
5.2 Data Partitioning.....	59
5.3 Performance Metrics.....	60
5.4 Volume Estimation Results	60
5.5 Individual-Specific Volume Estimation Results.....	64
5.6 Discussion.....	68
5.7 Summary and Future Work.....	69
Chapter 6 : Aggregate Consumption Estimation	71
6.1 Introduction.....	71
6.2 Data Partitioning and Performance Metrics.....	72
6.3 Fill Ratio Estimation Results	72
6.4 Individual-Specific Fill Ratio Prediction Results	76
6.5 Discussion.....	79
6.6 Residual Volume Prediction Results	80
6.7 Multi-Target Estimation Frameworks	83
6.8 Summary and Future Work.....	84
Chapter 7 : Improving Aggregate Consumption Accuracy Through Heuristic Fusion	86
7.1 Introduction.....	86
7.2 Methods	87
7.2.1 Sensor-Based Fill Ratio Estimates.....	87
7.2.2 Development of Fusion Models.....	88
7.2.3 Establishment of Model Parameters	91
7.3 Results.....	94
7.4 Summary and Future Work.....	97
Chapter 8 : Verification of Inclination Estimates Using Video Motion Capture	98
8.1 Introduction.....	98
8.2 Methods	98
8.2.1 Data Collection	98
8.2.2 Video Inclination Tracking.....	101
8.2.3 Drink Event Synchronization.....	103
8.3 Results.....	103
8.4 Conclusions and Future Work	105

Chapter 9 : Feature Set Expansion Using Additional Sensor Channels	106
9.1 Introduction.....	106
9.2 Proposed Supplements to the IS Feature Set	107
9.2.1 Additions from Accelerometer Channels.....	107
9.2.2 Additions from Gyroscope Channels.....	108
9.3 Effect of Feature Set Supplementation on Performance.....	110
9.4 Effect of Inclination Estimation Technique on Performance	111
9.5 Conclusions and Future Work	111
Chapter 10 : Assessment of Sensor Performance for Alternative Drinking Containers.	113
10.1 Introduction.....	113
10.2 Methods	114
10.3 Results.....	118
10.3.1 Container-Type Classification – LOSO Training	118
10.3.2 Container-Type Classification – Subject Specific Training	120
10.3.3 Container Type Classification with Equivalent Training Samples	122
10.3.4 Fill Level Classification	122
10.4 Discussion.....	123
10.5 Summary and Future Work.....	125
Chapter 11 : Conclusions	127
11.1 Summary.....	127
11.2 Limitations	127
11.3 Summary of Key Contributions and Recommendations for Future Work	128
BIBLIOGRAPHY	132

LIST OF TABLES

Table 3-1: Daily Use Activities Considered	31
Table 3-2: Summary of Testing Collections	33
Table 3-3: Summary of DL Testing Performance	42
Table 4-1: Correlation Between Features and Volume Label	52
Table 4-2: Correlation Between Previously Reported Motion Features and Volume	53
Table 4-3: Inclination Signature (IS) Feature Set	54
Table 4-4: Correlation Between IS Feature Set and Volume/Fill Ratio Labels	55
Table 4-5: Legacy Feature Set	57
Table 4-6: Correlation Between Legacy Feature Set and Volume/Fill Ratio Labels	57
Table 5-1: Variation in Volume MOAPE for Multiple Prompt Periods.....	64
Table 6-1: Variation in MOAPE for Multiple Prompt Periods – Fill Ratio Estimation.....	75
Table 7-1: Test Set Fill Ratio RMSE.....	94
Table 7-2: Range of Fill Ratio RMSE Across Test Set	95
Table 7-3: Test Set Fill Ratio MAPE.....	96
Table 7-4: Test Set Volume MOAPE(11)	96
Table 9-1: Supplemental Features from Resultant Acceleration	108
Table 9-2: Supplemental Features from Coplanar Gyroscope Resultant.....	110
Table 9-3: Supplemental Features from Axial Gyroscope Component.....	110
Table 10-1: Container Type Classification Accuracy: LOSO Training, Half-Full Fill.....	118
Table 10-2: Container Type Classification Accuracy: LOSO Training, Full Fill	119

Table 10-3: Container Type Classification Accuracy: LOSO Training, Mixed Fill	119
Table 10-4: Confusion Matrices: LOSO Training	120
Table 10-5: Container Type Classification Accuracy: S.S. Training, Half-Full	120
Table 10-6: : Container Type Classification Accuracy: S.S. Training, Full Fill	121
Table 10-7: Container Type Classification Accuracy: S.S. Training, Mixed Fill	121
Table 10-8: Confusion Matrices: Subject-Specific Training	121
Table 10-9: Fill Level Classification Accuracy: Bottle Container, LOSO Training	123

LIST OF FIGURES

Figure 1-1: Sensor Prototype Attached to a Refillable Bottle	3
Figure 1-2: Estimated Container Inclination During Excess Discharge and Drinking.....	5
Figure 2-1: Image of an Augmented Container Using an Insertable Capacitive Sensor [23]	9
Figure 2-2: Wearable Multi-sensor Configuration for Dietary Monitoring in [24].....	11
Figure 2-3: Diagram of Feature Similarity Search Algorithm in [24]	12
Figure 2-4: Variation in Signal Morphology Depicted in [26].....	15
Figure 2-5: FluidMeter System Diagram Presented in [11].....	16
Figure 2-6: Adaptive Segmentation Scheme Employed in [27]	18
Figure 3-1: Potential Failure Modes of Static Partitioning.....	25
Figure 3-2: Inclination Estimates During Various Daily Living Activities.....	27
Figure 3-3: Bottle and Wrist Sensor Outputs for TR Trial	33
Figure 3-4: Pseudocode of TMD Partitioning Algorithm.....	36
Figure 3-5: Example DL Testing Output with Estimated Drink Intervals.....	37
Figure 3-6: Scattering of Drink and Discharge Training Instances	38
Figure 3-7: Localization Error Distributions	42
Figure 3-8: Example Error Modes – DL Experiments, SSW Algorithm.....	43
Figure 4-1: Univariate and Joint Distrubtions of Training Data.....	47
Figure 4-2: Variation in Estimated Container Inclination Over Experimental Trial	50
Figure 4-3: Variation in Coplanar Sensor Orientation During Randomly Chosen Drinks.....	51
Figure 4-4: Variation in Container Inclination During Randomly Chosen Drinks	52

Figure 5-1: Variation in Volume MAPE for Various Models Considered.....	61
Figure 5-2: Variation in Volume MAPE Across Feature Sets.....	62
Figure 5-3: Distribution of Volume MAPE for Best-Case Estimator.....	63
Figure 5-4: Distribution of Volume MOAPE(12) for the Best-Case Estimator	65
Figure 5-5: Distribution of Volume MAPE Across Trials for Subject-Specific Duration Model	66
Figure 5-6: Variation in Duration-Based Volume MAPE Across Trials.....	66
Figure 5-7: Distribution of Volume MAPE for Subject-Specific Integration Model.....	67
Figure 5-8: Variation in Integration-Based Volume MAPE Across Trials	68
Figure 5-9: Scatter Plot of Estimate Versus Ground-Truth Volumes for Best-Case Estimator ...	69
Figure 6-1: Variation in Fill Ratio MAPE for Various Models Considered.....	73
Figure 6-2: Variation in Fill Ratio MAPE Across Feature Sets	73
Figure 6-3: Distribution of Fill Ratio MAPE Across Trials for Best-Case Estimator.....	74
Figure 6-4: Distribution of Fill Ratio MOAPE(12) for the Best-Case Estimator.....	76
Figure 6-5: Distribution of Fill Ratio MAPE for a Subject-Specific LR Inclination Model.....	77
Figure 6-6: Variation in Inclination-Based FR APE.....	77
Figure 6-7: Distribution of Fill Ratio MAPE for a Subject-Specific LR Integration Model.....	78
Figure 6-8: Variation in Inclination-Based FR MAPE	79
Figure 6-9: Approximated Versus Ground-Truth Fill Ratio for Best-Case Estimator	80
Figure 6-10: Technique for Leveraging Fill Ratio for Residual Volume Estimation.....	81
Figure 6-11: Comparison of Residual and Cumulative Techniques for Aggregate Estimation ...	82
Figure 6-12: Variation in Residual Volume-Based OAPE Versus FR.....	83
Figure 6-13: Volume Estimation Accuracy Enhancement Using Fill Ratio Information	84
Figure 7-1: Variation in Test-Set RMSE for Complementary Filtering Approach	92

Figure 7-2: Variation in Training RMSE Versus Noise Multiple.....	93
Figure 7-3: Example Outputs of Prediction Techniques.....	94
Figure 7-4: Variation in RMSE Across Trials in Test Set.....	95
Figure 7-5: Variation in Volume MOAPE(11) Across Trials in Test Set	97
Figure 8-1: Sensor and Marker Configuration.....	99
Figure 8-2: Visualization of Blender Tracking Output.....	101
Figure 8-3: Video Parsing Process – Wide View	102
Figure 8-4: Video Parsing Process – Zoom View	102
Figure 8-5: Visualization of Synchronization Process.....	103
Figure 8-6: Variability in Discrepancy Metric for Varying Complimentary Filter Weights.....	104
Figure 8-7: Distribution of Discrepancy Metric for Various IMU-Based Estimations	105
Figure 9-1: Variation in Acceleration Magnitude During Drinking Events	108
Figure 9-2: Variation in Gyroscope Signals During Drinking Events.....	109
Figure 10-1: Three Container Types Considered.....	113
Figure 10-2: Inclination Signatures for the Three Container Types (Half-Full Fill Level).....	116
Figure 10-3: Partitioning the Drinking Interval Using Relative Thresholding.....	117
Figure 10-4: Variation in Container Type Classification Accuracy	122
Figure 10-5: Drink Volume Versus Maximum Inclination Angle.....	125

Chapter 1 : Introduction

1.1 Motivation

The availability of consumer-grade devices for health monitoring applications has increased substantially in recent years [1]. By enabling the prevention and early detection of disease, these products offer a promising approach for addressing escalating healthcare costs [2]. Of the many diverse monitoring applications available, those promoting adherence to positive behavioral norms, such as minimizing sedentary time and maintaining a healthy diet, have received considerable attention. This focus is merited, given the key role of lifestyle habits in determining health outcomes [3].

Amongst available dietary monitoring solutions, numerous architectures for tracking fluid intake have been proposed. These devices have tremendous potential for improving wellness, as estimates suggest that approximately 16-28% of adults are dehydrated [4]. While the health consequences of dehydration are well understood, research indicates that even slight underconsumption of water is associated with various negative health outcomes, including obesity and reduced cognitive function [5].

Maintaining appropriate hydration levels is of particular concern for the elderly population, due to the degradation of fluid regulatory mechanisms with age [6]. Elderly individuals may decrease fluid consumption due to a variety of factors, including reduced osmoreceptor sensitivity, dysphagia, cognitive impairment, as well as mobility restrictions [7]. The large-scale ramifications of elderly dehydration are considerable, especially in developed countries with aging populations [8]. For example, Medicaid expenditures in the United States associated with hospital admissions for dehydration were estimated at \$5.5 billion in 2004 [9].

To promote hydration maintenance, numerous sensing technologies have been demonstrated for tracking fluid consumption. Approaches include containers with embedded sensing functionality (often denoted as augmented or smart-containers) [10], wearable technologies [11], and video-based solutions [12]. Unfortunately, each class of sensors is characterized by some limitation which may prohibit large-scale deployment. Namely, augmented containers restrict tracking to a dedicated set of drinking vessels, which limits the feasibility of logging aggregate intake across multiple containers during daily living. Wearable technologies may not be accepted by all users due to personal preference [13]. Moreover, the at-risk elderly population may reject such devices due to various physical limitations [14]. Furthermore, video-based solutions may be viewed as excessively intrusive. A more thorough review of the various hydration tracking technologies proposed in the literature is provided in Chapter 2. As described therein, the lack of a container-agnostic, non-wearable hydration tracking sensor serves as the primary motivation for this research.

1.2 Proposed Solution

Previous work has proposed a container-attachable IMU sensor for hydration tracking [15]. This approach alleviates the restrictiveness of augmented containers by allowing for simplistic reconfiguration across multiple drinking vessels [16]. Moreover, by isolating all electronic functionality to the exterior of the container, potential exposure to water is minimized versus sensors embedded in the interior of the drinking vessel.

Similar to other wearable consumption tracking technologies employing inertial sensors, this device operates using a motion-based sensing paradigm [17]. Namely, features of the container's movement patterns are used to detect drinking events and estimate their associated volumes. Detection of drinking events for container-attached devices is simplified versus wearable sensors,

which may exhibit false alarms for arm movements exhibiting similar kinematics to drinking [18]. An image of the sensor prototype used for all experiments described in this dissertation is shown attached to a refillable bottle in Figure 1-1. Both a triaxial accelerometer and gyroscope are integrated within the sensor prototype. The broad goal of this research is to improve upon the performance previously demonstrated in [15] for this sensor architecture as specified in the forthcoming research objectives.



Figure 1-1: Sensor Prototype Attached to a Refillable Bottle

1.3 Summary of Research Objectives

Successful estimation of the fluid intake associated with a drinking event may be conceptualized as a two-stage process. Namely, the drinking event must first be segmented from the streaming sensor output, followed by the estimation of drink volume from the partitioned data. For subsequent discussion throughout this dissertation, this former problem is denoted as sip detection, and is addressed in detail within Chapter 3. The latter problem is hereby referred to as

volume estimation, and is addressed in various frameworks throughout Chapters 4 - 9. A formal discussion of these two problems, which constitute the core research objectives of this dissertation, is provided in the following subsections.

1.3.1 Summary of Sip Detection Problem

The sensor output may be represented as a sequence of tuples denoted as $\{s_1, s_2 \dots s_N\}$, where $s_k \in \mathbb{R}^6$ corresponds to the six channel output at time index k . Sip detection algorithms seek to identify pairs of indices from the above set corresponding to the initiation and termination of all drinks, hereby denoted as $(s_i, s_f)_j$, where j is an index serving to identify the drinking event. This mapping is formalized in (1.1)

$$S: \{s_1, s_2 \dots s_N\} \rightarrow (s_i, s_f)_j \quad (1.1)$$

Traditional learning-based techniques for spotting activities within streaming data employ a two-stage processing approach. Data is initially segmented into fixed duration windows. Next, a classifier is used to distinguish events of interest from other intermixed activities. This process suffers from numerous disadvantages, especially for sparsely occurring events of variable duration such as drinks. Namely, such algorithms are inherently inefficient, and are characterized by trade-offs related to accuracy and spotting precision in the selection of windowing parameters.

To address these limitations, this dissertation introduces a dynamic segmentation and classification sip detection algorithm targeted for an attachable sensor architecture. The proposed approach enhances processing efficiency, increases temporal resolution, and improves detection accuracy versus traditional fixed-duration sliding window techniques. A deterministic initial stage partitions the output into candidate drinking events based upon their distinctive motion pattern. Next, a classifier trained to discriminate between drinking events and intermixed activities

demonstrating similar kinematics is applied to the segmented output. An example of the estimated inclination of the container during drinking, along with an event exhibiting a similar inclination pattern (discharge of excess fluid), is presented in Figure 1-2.

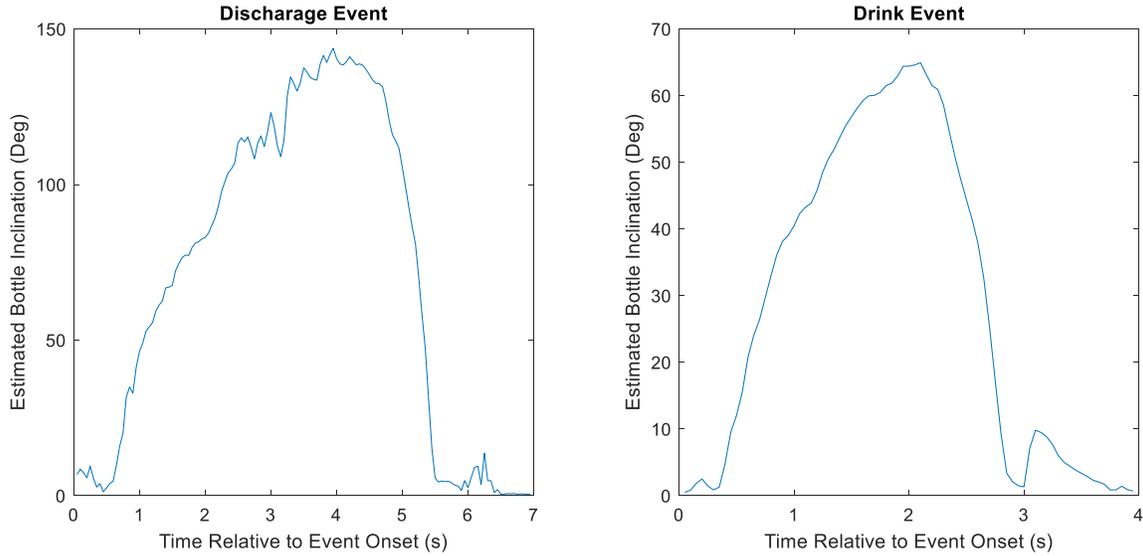


Figure 1-2: Estimated Container Inclination During Excess Discharge and Drinking

As the nature of the collection system utilized herein prohibits deployment in-the-wild as described in Chapter 2, the proposed sip detection algorithm is assessed through a series of experiments designed to test the most stringent scenarios encountered during the intended use case.

1.3.2 Summary of Volume Estimation Problem

After drinks have been segmented through application of the sip detection algorithm, a mapping between the partitioned output and estimated drink volume is developed as specified in (1.2), where \hat{V}_j corresponds to the estimated volume of the j^{th} drink.

$$\Phi^V: (s_i, s_f)_j \rightarrow \hat{V}_j \quad (1.2)$$

As reviewed thoroughly in Chapter 3, motion-based volume estimation using machine learning is a challenging problem, with prior results characterized by both limited accuracy and high inter-subject variability. The previously best-case reported mean absolute percentage error (MAPE) for drink volume estimations using a single inertial sensor was achieved by Hamatani et al. in [18]. In this research, the utilized accelerometer sensor was embedded within a commercial smartwatch. A MAPE of 58.9% was obtained for an experiment consisting of 1,069 drinks consumed by 16 individuals with ground-truth data recorded using a scale. Reported aggregate (i.e.: multiple drink) consumption estimates were slightly improved due to the cancelation of errors across adjacent drinking events. For a container-attachable inertial sensor, previously reported volume estimation results are limited to a single experiment. Namely, Dong et al. achieved an aggregate estimation error of 25% across subjects for an experiment consisting of approximately 70 drinks consumed by seven subjects [15].

As described in Chapters 4-9, numerous techniques are proposed and explored within this dissertation for improving motion-based drink volume estimation performance for the proposed sensing architecture. Drink volume is estimated in Chapter 5 using a support vector machine regression model with 33 hand-engineered features describing the estimated container inclination during drinking. Performance is improved versus the prior state-of-the art for a single inertial sensor, with MAPE reduced by 11.05% versus results from a comparable experiment presented in [11]. An alternative technique for estimating the consumption across multiple drinks based upon estimates of fill level is investigated in Chapter 6. Denoted as residual volume estimation, this process approximates aggregate consumption using fill level estimates under the assumption of known container geometry. Chapter 7 proposes a technique for further improving these consumption estimates by fusing predictions from a heuristic consumption model.

Container inclination estimates are verified using an open-source video motion capture package in Chapter 8. This chapter also introduces an approach for utilizing the gyroscope output to improve inclination estimates. Chapter 9 explores potential expansions of the proposed feature space and the resulting effects on estimation accuracy. Utilization of the alternative inclination estimates proposed in Chapter 8 are also explored within this chapter.

To support volume estimation efforts, a large-scale data collection consisting of 1,908 drinks consumed by 84 individuals over 159 trials was conducted as described in Chapter 4. All experiments were performed using a scripted protocol. Namely, participants only handled the container for purposes of drinking, resting the bottle on a stationary surface between drinking events. This protocol eliminates the complexities associated with sip detection, thereby optimizing the data format for the intended use case. Moreover, this approach allows for ground-truth data to be collected on a per-drink basis using an electronic scale, thereby eliminating reliance on commercial smart-bottle products for data labeling.

1.3.3 Generalization to Additional Drinking Containers

While all data collections supporting the aforementioned sip detection and volume estimation efforts were performed for a single container type (refillable bottle), this dissertation also provides a limited exploration of sensor performance for alternative drinking vessels. Namely, the ability of the proposed device to detect the type of container to which it is attached, along with the fill level from which a drink is consumed, is explored in Chapter 10. Dedicated experiments are conducted for both a glass and mug, in addition to the previously utilized refillable bottle.

Chapter 2 : Related Work

2.1 Introduction

This chapter reviews alternative technologies for automated fluid consumption tracking. Additional consideration is allocated for approaches employing motion-based sensing paradigms using IMU sensors. Sip detection and volume estimation results are provided where available.

2.2 Review of Hydration Tracking Sensors

Numerous hydration management technologies have been previously proposed in the literature. While complete solutions are inherently complex cyber-physical systems, which must be cognizant of individual hydration requirements, provide appropriate reminders, etc., this review focuses solely on the enabling sensing mechanisms.

2.2.1 Augmented Containers

Tracking solutions which embed sensing functionality within a dedicated drinking vessel are typically referred to as augmented or smart containers. Documentation of these technologies is largely restricted to the patent literature, thereby limiting the availability of performance data. Augmented containers for consumption tracking are currently available in the commercial market.

Augmented containers have been implemented using a variety of sensing modalities. Sensors capable of measuring the total volume of fluid contained within the vessel, such as pressure [18] and capacitive sensors [19], have been demonstrated. To form consumption estimates using this type of sensor, a reference measurement is required to assess changes in total volume. A mechanism for implementing this approach using the sensing modality considered herein is explored in Chapter 5. A capacitive fluid sensor for measuring total container volume as integrated within a current commercially available smart bottle is depicted in Figure 2-1.



Figure 2-1: Image of an Augmented Container Using an Insertable Capacitive Sensor [23]

Augmented containers estimating consumption on a per-drink basis have also been described. For example, a container with a dedicated sensor for measuring the exiting flow rate during drinking has been proposed [20]. Per-drink consumption estimation is addressed in Chapter 4 of this dissertation.

IMU sensors have been considered as an enabling sensing modality for augmented containers. Proposals for integrating IMUs within either the structure or cap of the drinking vessel have been documented in the literature. Extensions of this technology for alternative applications of benefit, such as activity tracking, have been proposed [21]. The integration of multiple sensing modalities within smart bottles, such as a touch-based sensor for heart rate monitoring, has also been suggested [22].

As the above references are largely restricted to patent literature, available performance data is limited. However, recent research has provided some independent verification of the accuracy of commercially available solutions. For example, Borofsky et al. assessed the aggregate tracking accuracy of the smart bottle previously shown in Figure 2-1. An experiment was

conducted where eight participants consumed water from the bottle over 62 twenty-four-hour intervals, with manual consumption estimates also recorded. Daily consumption estimates produced by the bottle varied from hand measurements by less than 3%. [23].

The primary disadvantage of the aforementioned technologies is the restriction of tracking functionality to a dedicated set of containers. For individuals seeking to track total daily consumption across a variety of drinking vessels, these products may be viewed as excessively restrictive. Moreover, many of the proposed approaches require the embedding of electronics within the interior of the container, thereby mandating additional design challenges to avoid water exposure. The container-attachable nature of the solution considered within this dissertation alleviates these concerns.

2.2.2 Multi-Sensor Wearable Consumption Trackers

To address the restrictiveness of augmented container tracking, various alternative sensors have been demonstrated. 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. Network sensors included IMUs placed on the upper limbs, an ear microphone, and an EMG and microphone combination configured in a throat collar [24]. This system was designed for monitoring the intake of both fluids and foods, thereby motivating the complexity of hardware employed. A schematic depicting the configuration of sensors across the body is shown in Figure 2-2.

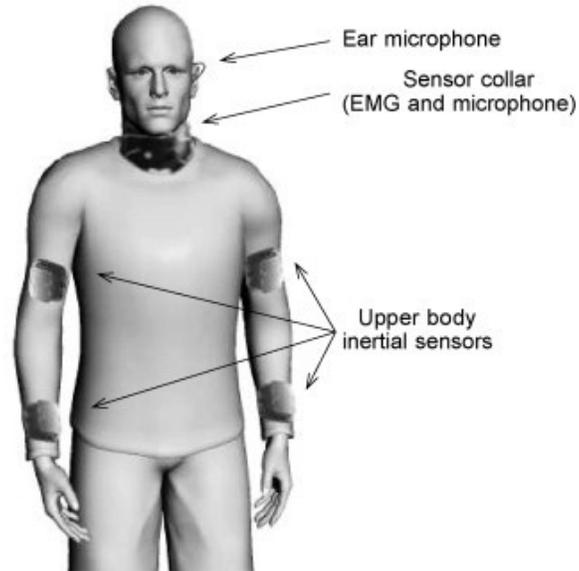


Figure 2-2: Wearable Multi-sensor Configuration for Dietary Monitoring in [24]

An experiment involving four individuals was used to assess system performance. Participants completed four consumption activities, including fetching and drinking from a glass, along with additional common hand gestures (i.e.: head scratching, using a phone, etc.). Independent detectors were used to spot the various consumption activities of interest, with individual outputs fused to improve accuracy. Detection was accomplished using a feature similarity search (FSS) algorithm on fixed duration partitions of sensor data. The FSS algorithm is summarized in Figure 2-3 and described thereafter.

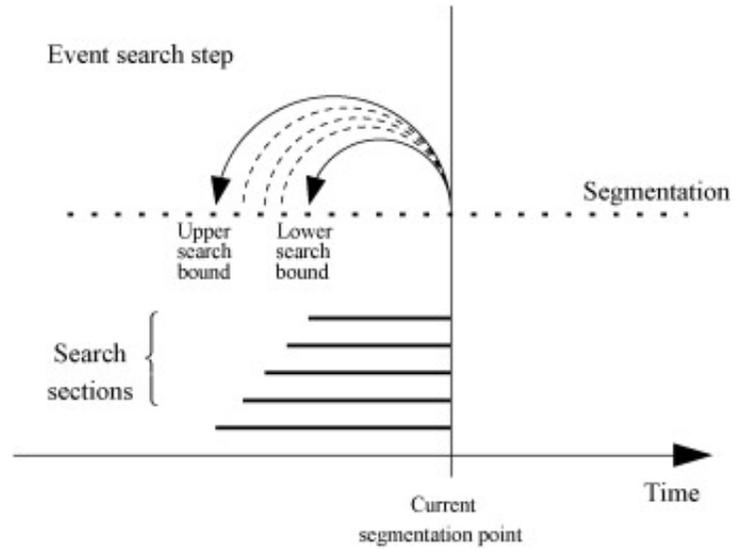


Figure 2-3: Diagram of Feature Similarity Search Algorithm in [24]

During training, the FSS algorithm determines the candidate durations for each event of interest using manually denoted video-based ground truth data. In addition, templates for each activity are formed in the utilized feature space. In inference, the FSS algorithm searches across the feasible duration for each event, computing the feature space representation of the sensor output for each search section. Similarity between the computed representation and template is measured in a Euclidean sense, with a decision formed using an event-specific threshold determined during training.

Drinking events were detected using features computed on the estimated Euler angles of the forearm. Drinks were recognized with 86% recall and 85% precision using a user-specific training procedure. Considerable (20%) confusion was demonstrated between drinking activities and those of the null class. No volume estimation was performed in this work.

While such an extensive sensor architecture may be necessary to capture the variety of dietary events considered, practical feasibility of the proposed system is limited by the number of

required sensors. The requirement of user-specific training data also limits the practical viability of this approach. Moreover, wearable sensors inherently capture signals associated with all daily living activities, resulting in a large and highly variable null class. For the attachable architecture proposed herein, the null class is restricted to only non-drinking activities for which the container is in motion with the sensor attached (i.e.: transport, handling, etc.). This reduction in problem complexity allows for the deployment of a more streamlined partitioning algorithm versus FSS as described in Chapter 3.

Mirtchouk et al. [25] performed drink volume estimation using a similar network of wearable audio and motion sensors. This effort was also focused on tracking both drinking and eating activities. An acoustic earbud, two commercial smartwatches, and a headset with embedded IMU sensors were used to collect data. While food type classification was also demonstrated, presented results below focus solely on efforts related to hydration tracking.

Six participants consumed 171 drinks of multiple types of liquids (i.e.: coconut water, coffee, etc.) over a 72-hour period in an unscripted experiment. Data was partitioned using video-based annotations on a per-intake basis. Various audio features (i.e.: energy, spectral flux, zero-crossing rate, etc.) were computed over 200 ms windows. Motion features were computed on a five second frame, and included 11 statistical features, 15 temporal shape features, and two frequency features. Random forest regression models were trained using a leave-one-drink-out approach to account for the lack of consistent consumption patterns across participants. The mass of each drink was estimated with a best-case mean absolute percentage error (MAPE) of 47.2% under the assumption of known fluid type. Similar to [24], the practical viability of this solution is limited by the number of sensors employed. Moreover, no approach for identifying drinking events from continuous sensor output was proposed within this work.

2.2.3 Single Sensor Wearable Consumption Trackers

Subsequent research has alleviated the restrictiveness of multi-sensor systems by isolating tracking functionality within a single wearable device. Amft et al. spotted drinking activities using a wrist-wearable IMU sensor containing a triaxial accelerometer and gyroscope [26]. Results were validated using 5.84 hours of data collected from six subjects during daily living. The data set included 560 drinking instances consumed from varying container types. A separate scripted experiment consisting solely of drinking events was collected for training. Sensor data was initially segmented into 2 second windows, with drinks subsequently spotted using the previously described FSS algorithm. Detection thresholds were determined on a per-subject basis during training.

Two-hundred general time-domain features were used to describe the motion pattern of the arm. The Mann-Whitney-Wilcoxon test was used to extract a subset of the 20 highest ranked features. The drinking event was partitioned into two sections, denoted as fetch (period of transport towards and away from the mouth), and sip (period of fluid intake). An image of the signal morphology during these two micro-events is depicted in Figure 2-4. The authors noted greater variability in the recorded signal for the fetch versus sip motion. A similar strategy for parsing drink events into the transport and sip phases for the sensor described herein is proposed in Chapter 4. Fetch motions were spotted with 84% precision and 90% recall, while sip motions were spotted with 84% recall and 94% precision.

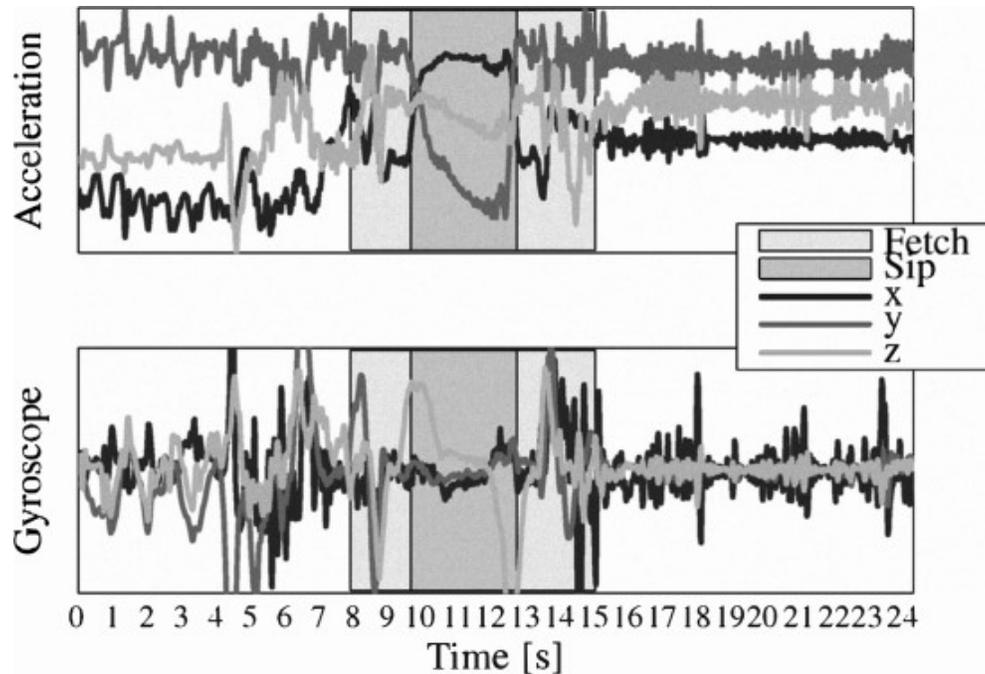


Figure 2-4: Variation in Signal Morphology Depicted in [26]

A volume estimation strategy based upon fill level detection was also introduced within this work. These experiments used a magnetic coupling sensor system attached at both the shoulder and wrist. An experiment was conducted in which three participants consumed 30 drinks from nine different container types in a scripted sequence. Drinks were consumed at three initial fill levels (full, half-full, near empty), with subjects instructed to ingest only a minimal amount during each drink to avoid overconsumption. Individual-specific classifiers achieved an average fill level classification accuracy of 72% across all subjects and container types. Classification accuracy across subjects varied considerably, ranging from 58% to 83%.

While estimation of aggregate consumption using fill level information is feasible, practical deployment requires increased resolution, along with consideration of the effect of varying drink volume on the estimation process. In addition, the requirement of individual-specific

training data limits the feasibility of employing the proposed system in practice. Both limitations are addressed in the research described within this dissertation.

Hamatani et al. [11] proposed FluidMeter, a fluid consumption tracking system utilizing the embedded IMU sensor within a commercial smartwatch. Both sip detection and volume estimation were performed as emphasized in the system design diagram shown in Figure 2-5.

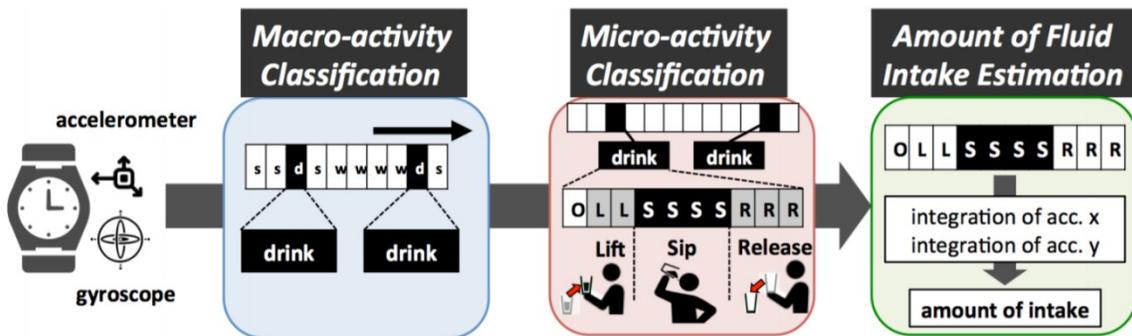


Figure 2-5: FluidMeter System Diagram Presented in [11]

Drinking events were distinguished from other arm motions using a macro-activity classification module. This module was implemented using a conditional random field (CRF) model to map features of motion to activity states. Data was partitioned using a static sliding window of 8 second duration with 0% overlap. Eight explicit activity classes were considered, including various sedentary and active states (sitting, standing still, moving, etc.), eating, and drinking. A null class was used to represent remaining motion signatures. Twenty-eight statistical features (i.e.: average, standard deviation, etc.) were used to describe the motion pattern across the six sensor channels, with backward feature selection performed to reduce dimensionality.

Once drinking events were spotted using the macro-classifier, an additional CRF classifier was used to further partition drinks into the following micro-events – 1) Lift, 2) Sip, and 3) Release. Data was segmented for the microevent classifier using a 500 ms window with 50% overlap. Sip

detection results for various collections were presented. For the Lab-macro dataset collected from 9 individuals over 1,325 minutes, drinking events were classified with 83.6% precision and 87.3% recall. The authors noted that false negatives were most commonly associated with eating due to similarity in arm movements.

An additional data collection, denoted as the Lab-micro+ dataset, was performed for assessing micro-gesture classification and volume estimation. Data was gathered from 16 individuals consuming 1,069 drinks in a laboratory setting. The ground truth weight of each drink was recorded using a digital scale, with micro-event boundaries specified by the participants using a smart-phone application. For this collection, the sip micro-gesture was classified with 90.7% precision and 96.3% recall.

While volume estimation results for multiple experiments were reported, the Lab-micro+ dataset most closely resembles the scripted experiments conducted in Chapter 4. Various linear regression models utilizing both sip duration and the integral of the accelerometer signals tangential to the wrist surface were used to estimate the mass of each drink. A best-case MAPE of 58.9% was achieved for the integration model trained using leave-one-subject-out (LOSO) validation. While variability across subjects was not reported for the Lab-micro+ dataset, models trained on this data exhibited considerable dispersion in accuracy across subjects (MAPE ranging from 57.9% to 11.0%) when applied to a dedicated in-situ collection (Wild-office dataset). MAPE for the in-situ collection using ground-truth data collected with a commercial smart bottle was 31.8%.

While FluidMeter offers an unobtrusive mechanism for consumption tracking for existing smartwatch users, some individuals may be unwilling to adopt the requisite technology to employ

this approach. Moreover, while the authors noted the influence of both fill level and drink volume on the drink motion pattern, no explicit efforts were employed to address this interdependence.

Chun et al. also detected drink episodes using a commercially available wrist-mounted inertial sensor. Drinking events were spotted with 90.3% precision and 91.0% recall for a study consisting of 561 drinks consumed by 30 participants [27]. An adaptive segmentation technique originally proposed in [28] was used within this work based upon the characteristic morphology of the accelerometer signal during drinking. Namely, data was initially partitioned into non-overlapping windows of 1 second duration. Windows were then increased bilaterally in an iterative fashion until the signal range exceeded a predefined threshold. This dynamic expansion process is summarized in Figure 2-6.

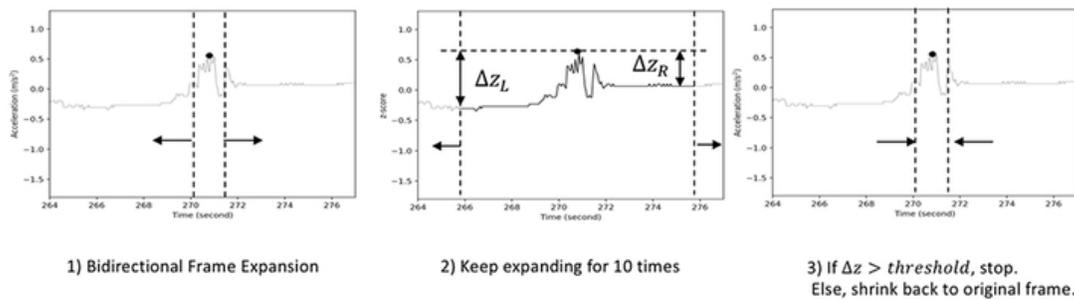


Figure 2-6: Adaptive Segmentation Scheme Employed in [27]

Once the adaptive segmentation procedure was applied, a set of 45 general features were computed on the adaptive frame duration. A random forest learning algorithm was used to classify drink events.

The adaptive segmentation proposed in Chapter 3 does not require preliminary segmentation, thereby supporting real-time implementations with minimal latency. Moreover, due to the placement of the sensor on the container, mechanistic thresholds may be established based upon container geometry (i.e.: minimum container inclination required to induce fluid flow etc.),

if available. In addition, the newly proposed algorithm utilizes additional qualifications in the dynamic segmentation process, thereby further distinguishing candidate drink events prior to classification.

Gomes and Sousa proposed a method for identifying the hand-to-mouth container movement during drinking episodes using a single IMU sensor placed on the forearm [29]. Data was partitioned into fixed duration windows of 1 second with 50% overlap. Seventeen participants performed both drinking events and other daily living activities (i.e.: walking, other hand to mouth movements, etc.), producing a dataset consisting of 1,034 drink instances versus 11,526 null class activities. A set of 10 general features were extracted using backwards feature selection for a random forest classifier. Drinking events were spotted with 85% recall and 84% precision within an experiment mimicking the daily use case of the device. While the proposed method may be useful for triggering the deployment of an additional processing stage for volume estimation, no such techniques were demonstrated within this manuscript.

Although wearable approaches are appropriate for many users, they may be excessively cumbersome for some individuals, including persons with limited dexterity and other physical limitations. This concern is alleviated for the attachable sensor placement considered herein. This advantage comes at the expense of reduced convenience, as the proposed solution must be repositioned on the container before each drinking instance.

2.2.4 Contactless and Nearable Consumption Trackers

Amongst contactless solutions, Chua et al. used a Haar-like feature set to spot drinking events by identifying the gripping posture of the hand through image processing [30]. Ienaga et al. used features related to joint position estimated using a Kinect sensor to demonstrate sip recognition for service robotic applications [31]. 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 [32]. 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. Proposed approaches include the integration of sensing functionality into coasters [33]. Alternative container-attachable sensors have also been demonstrated. Namely, an attachable passive RFID sensor for spotting drink events was proposed in [34]. Versus the IMU-based approach described herein, this technique requires additional infrastructure. Moreover, it does not support modeling the container inclination to enable mechanistic algorithms exploiting the characteristics of drink motion patterns.

2.2.5 Prior Research for Attachable IMU Sensors

The sensor architecture considered within this dissertation was originally proposed by Dong et al. in [15]. A 100 Hz accelerometer was used for data collection, which differs from the 20 Hz sampling rate employed in the current work. Both preliminary sip detection and volume estimation results were reported. Computations were performed only on the component of the accelerometer parallel to the axis of the bottle. The signal was smoothed using an 11-point moving average filter in pre-processing.

For sip detection analysis, the conditioned signal was initially partitioned using a sliding window of 30 seconds duration with 50% overlap. Within each window, an amplitude threshold of 0.2 g was applied to identify local minima values of drinking events. Local minima separated

by more than 2 seconds were subsequently extracted for classification. Classification was performed using the following four hand-engineered features – 1) signal range, 2) event duration, 3) signal mean, 4) increase-to-decrease ratio. This feature space is utilized as a benchmark in both the sip detection and volume estimation results considered herein.

Sip detection was assessed using an experiment involving seven subjects. Each subject conducted two trials for dedicated drinking collection, consuming an entire bottle in each session. In addition, two of the subjects performed a data collection solely intended to capture container motion during non-drinking events (i.e.: walking with the container in-hand, etc.). Approximately 1 hour of artifact data was collected. Data from both experiments was then parsed using the proposed segmentation algorithm. This parsing resulted in 143 drink events versus 104 non-drink events. A variety of classification models were evaluated for identifying drink events, including support vector machines, artificial neural network, and naïve Bayes classification models. All three models reported an accuracy exceeding 90%.

While valuable for initial proof-of-concept, inference regarding the generalization of these results to real-world scenarios is limited by the nature of the experiments performed. Namely, drinking events were not intermixed amongst daily activities during collection. When deployed in-the-wild, handling patterns which may result in missed drink detections for the proposed dynamic partitioning strategy may be envisioned. For example, if the container is first inclined during handling past the specified threshold, with a subsequent drink occurring less than two seconds later, the drink would be discarded due to the proposed separation criteria. The dynamic segmentation algorithm introduced in Chapter 3 addresses this concern by imposing temporal restraints on the candidate event duration, not inter-event spacing.

Moreover, the newly proposed segmentation technique is further improved through utilization of pose estimation. Namely, by estimating the container's inclination, mechanistic thresholds may be established to quantify the intensity of the drinking event (i.e.: specification in degrees, versus raw units of acceleration). The proposed algorithm is further improved through the addition of a post-thresholding merging process to support capturing of the entire drinking event

In addition, the feature space utilized in the final classification stage is modified herein to support discrimination between drinks and other motion events exhibiting similar inclination dynamics (i.e.: discharge of excess water). These improvements are assessed using a data collection specifically designed to test two challenging application scenarios – 1) closely separated drinking events, and 2) drinking events closely intermixed amongst daily living activities, including discharge events with similar motion patterns.

Volume estimation results were provided in [15] for a separate experiment where seven subjects took ten drinks from a refillable bottle with the sensor attached. Various regression models using the aforementioned four-element feature space were evaluated. A best-case average aggregate consumption estimation error of 25% across subjects was achieved using support vector machine (SVM) regression models trained in a LOSO framework. No results were provided for per-drink estimation accuracy. As noted in Chapter 4, the models considered herein are demonstrated to improve aggregate estimation accuracy relative to these results.

Chapter 3 : A Dynamic Partitioning Algorithm for Improved Sip Detection

3.1 Introduction

Traditional activity classification algorithms partition sensor output into fixed duration frames using a sliding window. While commonly employed throughout the literature, this static segmentation is characterized by notable disadvantages [35]. These concerns are especially noteworthy for sparse events of variable duration such as drinking. Under these conditions, static segmentation algorithms are inherently inefficient, suffer from trade-offs regarding accuracy and spotting precision, and may exhibit misclassification due to activity boundary effects.

The research presented within this chapter addresses these deficiencies through the development and verification of a novel two-stage sip detection algorithm. Adaptive segmentation of sensor data stream is initially performed to identify candidate drink intervals according to their unique inclination morphology. Intervals are spotted using a *Threshold-Merge-Discard (TMD)* algorithm. As this partitioning algorithm inherently discriminates against most daily use activities (i.e.: transport, maintenance, etc.), the classifier may be targeted for discriminating against actions with similar inclination kinematics to drinking (i.e.: discharging of excess water, etc.). The proposed algorithm is verified using a data set intended to thoroughly evaluate the proposed use-case of the device within the restrictions of the collection system.

The primary contribution of this chapter is the development and verification of the aforementioned two-stage temporal partitioning and classification algorithm for sip detection. The algorithm is demonstrated to improve true-positive detection rate while dramatically reducing the number of required classifier operations versus a traditional static sliding window (SSW) detection algorithm. Moreover, preliminary analysis suggests that spotting precision is also improved versus static segmentation.

A brief review of data partitioning strategies for activity recognition applications is provided at the beginning of this chapter. Next, a description of the data collection system and pre-processing workflow employed in both the current and future chapters is provided. Experimental methods and results are subsequently discussed, along with suggestions for future sip detection research.

3.2 Partitioning Strategies for Online Activity Classification

While the literature applying IMU sensors for human activity recognition (AR) is well-established [36], 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 [37]. 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 segmented into fixed length intervals (W) of pre-defined overlap (p), have been heavily explored for online AR [38-40]. 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 [41]. 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, 2) performance challenges for situations where the window encompasses signals from multiple

activities of interest (i.e.: event boundaries, cases where 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 Figure 3-1 for the estimated container inclination using the current sensor architecture.

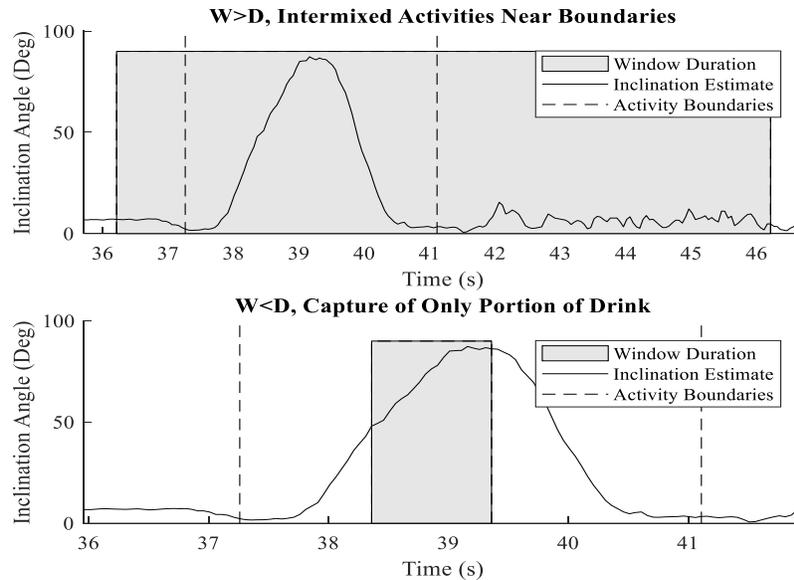


Figure 3-1: Potential Failure Modes of Static Partitioning

With respect to 2), the influence of window length on classification errors for fixed partitioning frameworks has been explored in the literature [42]. The coupling between the construction of the feature space and window parameters was investigated in [43], with adaptive selection of features and window parameters on a per-activity basis yielding optimal performance. As the current 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 segmentation proposed within this chapter.

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 [28]. As this approach requires discrete state-based sensor outputs to trigger event boundaries, it is not directly applicable for the current application.

Various other techniques which dynamically segment streaming data according to some event-specific rule have been explored. For example, Junker et al. [44] used the sliding window and bottom-up algorithm, originally proposed by Keogh et al. [45], 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 kinematics, the difference in captured signal morphology for the current events of interest renders such complexity unnecessary. Inclination estimates during various daily activities as estimated by the attachable IMU sensor are shown in Figure 3-2. As noted, the kinematics of drinking are highly distinguished from most general handling, transport, and maintenance activities.

More simplistic threshold-based partitioning approaches have been suggested for both wearable [46], and vision-based [47] 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.

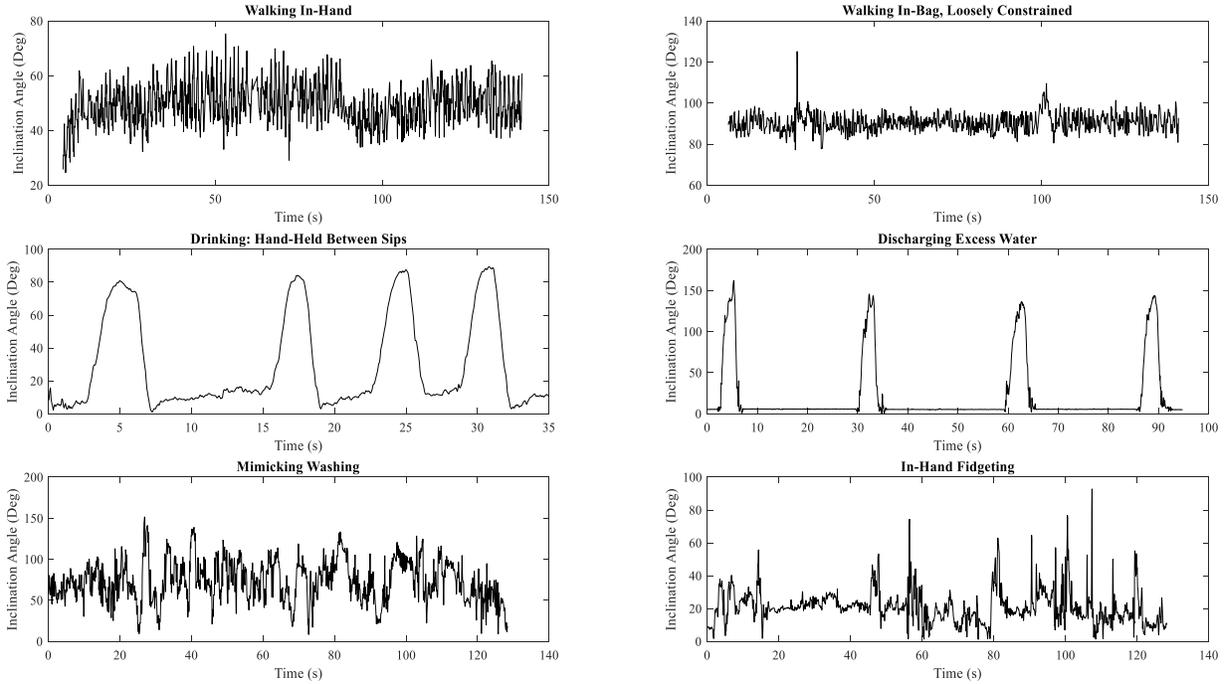


Figure 3-2: Inclination Estimates During Various Daily Living Activities

For example, Luckowicz et al. used acoustic intensities to segment accelerometer outputs for tracking assembly-related activities in a wood shop [48]. 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 [49]. As these and similar techniques require additional hardware, they are not suitable for integration within our proposed lightweight and retrofittable solution.

3.3 Collection Hardware

A three-node wireless sensor network composed of six degrees of freedom IMU sensors was used in all data collections described within this manuscript. Each IMU node contains a triaxial accelerometer (Analog Devices ADXL345), gyroscope (InvenSense IMU-3000), and 802.15.4 wireless transceiver (IRIS Mote module). Sensors were fastened in the desired configuration using

a customized elastic strap with a Velcro connector. The specific configuration of each node during the various collections performed is provided in the appropriate forthcoming sections. Only the accelerometer signal is used within the current chapter, with processing of the gyroscope output for drink spotting applications targeted for future research.

Data was transmitted from each node to a MEMSIC IRIS base-station interfaced to a PC through a USB port. This configuration demanded that the laptop be within the transmission range of the sensor during all data collection, thereby limiting in-the-wild testing. Data was polled from the sensor nodes by the base-station in a round-robin fashion at a target sampling interval of 50 ms per node. Data for each experiment was stored in a separate text file, which was controlled using a customized Python script. All files were processed offline 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, respectively, with sign convention defined according to a traditional right-handed framework. A visualization of the sensor coordinate axes was provided in Figure 1-1. It should be noted that while care was taken to maintain the stated orientation during all trials, variations may have occurred during experimentation as part of the handling process.

3.4 Signal Preprocessing

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 negligible acceleration as specified in (3.1), where a_j denotes the j^{th} component of the accelerometer output.

$$\hat{\theta} = \tan^{-1}\left(\frac{\sqrt{a_y^2 + a_z^2}}{a_x}\right) \quad (3.1)$$

Under ideal alignment, this sensor's inclination may be used to approximate that of the container. This assumption is examined in Chapter 8 using video-based positional tracking.

3.5 Data Collection

3.5.1 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.

3.5.2 Training Collection

To support the rapid acquisition of high-quality training data, individual collections were conducted for each activity described in Table 3-1. For all events other than drinking and discharging excess water, 35 minutes of data (5 minutes/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 device, hereby denoted as the bottom sensor, was placed below the hinge at the bottom of the bottle as shown in Figure 1-1. 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. 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 the acquisition of high-quality data. Isolated training collections have also been employed in related work for similar motivations (i.e.: [26]). 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 were depicted in Figure 3-2.

Table 3-1: 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-Bag, Restricted (W-IB-R)	Same as W-IB, 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

3.5.3 Temporal Resolution Testing Collection

A dedicated testing collection was conducted to assess the capacity of the algorithm to resolve closely spaced drinks. 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 contained spacings of two and 10 s, and the other two contained spacings of five and 20 s. This information is summarized in Table 3-2.

TR collections used 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 Figure 3-3.

3.5.4 Simulated Daily Living Test 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 daily living simulated (*DL*) collection is also provided in Table 3-2.

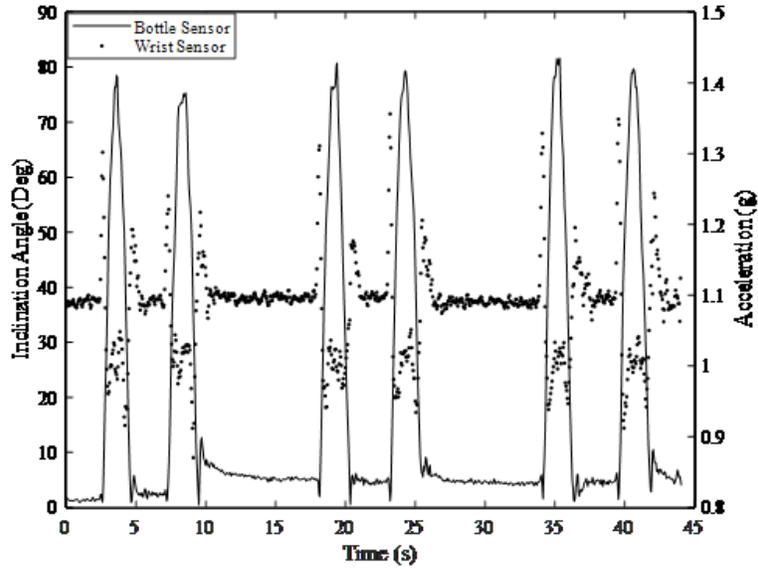


Figure 3-3: Bottle and Wrist Sensor Outputs for TR Trial

Table 3-2: Summary of Testing Collections

Collection ID	Interleaving Activities Considered	Inter-Drink Spacings Considered	Total Drinks Per Subject/Total
TR	<ul style="list-style-type: none"> • In-Hand Holding 	{2,5,10,20} s	24 / 120
DL	<ul style="list-style-type: none"> • In-Hand Holding • W-IH • W-IB • DEW • MW 	{2, 10} s	32/160

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 this chapter (Figure 3-5), after introduction of the proposed dynamic partitioning strategy.

3.5.5 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 (3.2)

$$A[n] = |a_c - 1| > \tau \quad (3.2)$$

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 (3.3) and (3.4), respectively.

$$t_s^j = \inf\{n \mid A^j[n] > \tau\} \quad (3.3)$$

$$t_e^j = \sup\{n \mid A^j[n] > \tau\} \quad (3.4)$$

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

3.6 Algorithm Development

3.6.1 Overview

Binary event detection schemes employing temporal partitioning with subsequent classification may be conceptualized as a two-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 (3.5)

$$\psi: \{x_i\} \rightarrow \{\mathbf{d}^m\} \quad (3.5)$$

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 of the partition. 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 (3.6)

$$\varphi: f(\mathbf{d}^m) \rightarrow L^m \quad (3.6)$$

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 classifiers employing hand-engineered feature spaces, f is a mapping of the raw data to the designed feature representation. The detection process may require additional post-processing, especially for schemes employing SSW segmentation with considerable overlap.

3.6.2 Dynamic Partitioning Strategy

As was exhibited in Figure 3-2, the inclination signal follows a concave morphology during drinking events. The proposed dynamic partitioning strategy seeks to identify time intervals containing candidate drink signals by exploiting this distinguished inclination signature. This process is detailed in pseudocode in Figure 3-4, with a summary description provided in the following paragraph.

To begin partitioning of the input stream, an amplitude threshold is applied to the inclination signal on a per-sample basis. This threshold is determined empirically ($\theta_{\min} = 12^\circ$) as the minimum angle required to induce fluid flow from a full bottle. Next, adjacent intervals of samples exceeding the threshold which are separated by less than a merge parameter ($s = 3$ samples) are combined. 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 .

Temporal Partitioning Pseudocode

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

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

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

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

Merge resultant thresholded subset, N^* , to form candidate output set N'_D

Initialize $N_D = N'_D = \{\}$
Set $\hat{t}_s^1 = \inf(N^*), j=1$
for $k = \hat{t}_s^1 + 1 : |N^*|$
 if $(N^*[k] - N^*[k-1]) > s$
 $\hat{t}_e^j = N[k - 1]$
 $N'_D = N'_D \cup (\hat{t}_s, \hat{t}_e)^j$
 $j = j + 1$
 $\hat{t}_s^j = N[k]$
 end if
end for

Discard events of insufficient maximum amplitude
or duration range in N'_D to form output set N_D

Set $d_{\text{allow}} = [d_{\min}, d_{\max}]$
for $j = 1 : |N'_D|$
 if $\{(\hat{t}_e^j - \hat{t}_s^j) \in d_{\text{allow}} \ \& \ \max(\hat{\theta}^j) > A_{\min}$
 $\ \& \ \text{range}(\hat{\theta}^j) > R_{\min} \}$
 $N_D = N_D \cup (\hat{t}_s, \hat{t}_e)^j$
 end if
end for

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

Figure 3-4: Pseudocode of TMD Partitioning Algorithm

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 seconds) 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 Figure 3-5.

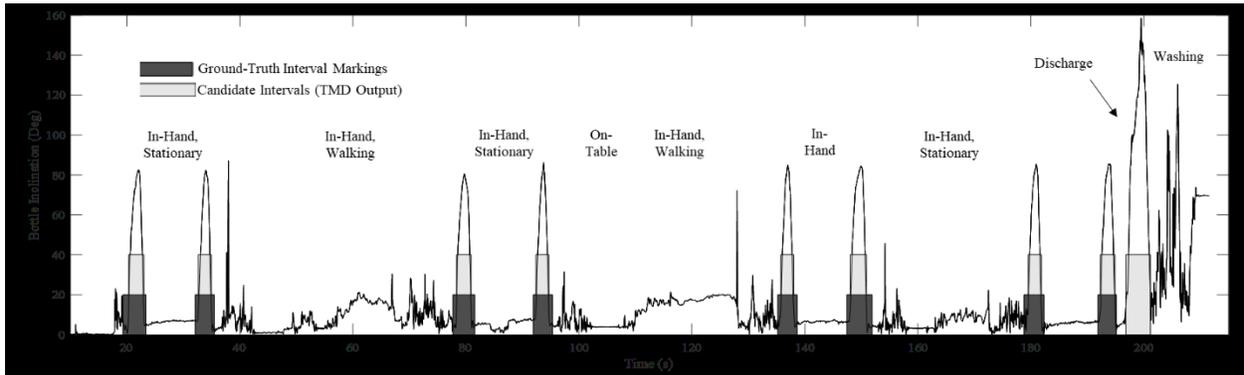


Figure 3-5: Example DL Testing Output with Estimated Drink Intervals

3.6.3 Classification Algorithm

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 concave 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 (3.7) – (3.9).

$$\hat{\theta}^j_{max} = \max(\hat{\theta}(t_s^j:t_e^j)) \quad (3.7)$$

$$\frac{d}{dt}\hat{\theta}^j_{ra} = \text{mean}(\hat{\theta}^j(t_s^j:t_{max}^j)) \quad (3.8)$$

$$\Delta e^j = \sum_{t_s^j}^{t_e^j} (\hat{\theta}[k] - s(\hat{\theta}[k]))^2 \quad (3.9)$$

where $s(\cdot)$ is a smoothing operation implemented as a third-order Savitzky-Golay filter with a nine-sample frame length (with delay compensation), 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 Figure 3-6.

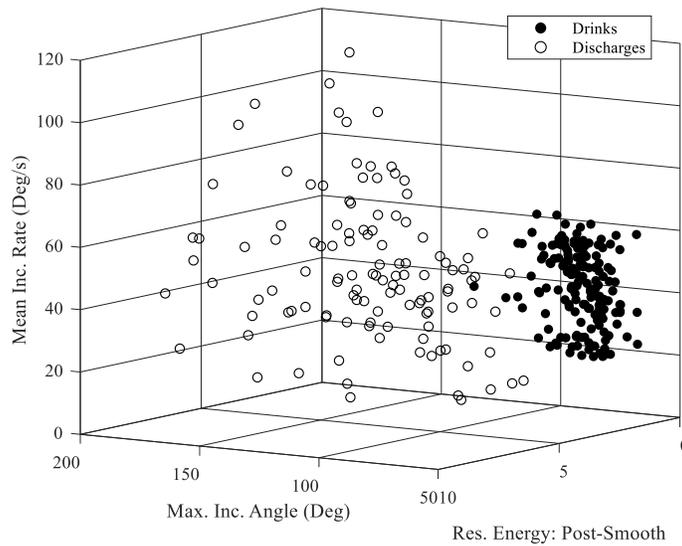


Figure 3-6: Scattering of Drink and Discharge Training Instances

Training data (D and DEW only) was partitioned using five-fold cross-validation to minimize the effect of overfitting in the model evaluation process. A variety of classifier models were then 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 the previously considered technique for a container-attachable architecture [31]. Partitioning was performed using an SSW scheme ($W = 3s, p = 75\%$). A slightly modified version of the proposed four-element feature space was employed as specified in (3.10) - (3.13).

$$\hat{\theta}_{range}^m = range(\hat{\theta}[t_i^m:t_f^m]) \quad (3.10)$$

$$N^m = nnz(\hat{\theta}[t_i^m:t_f^m] > \theta_{min}) \quad (3.11)$$

$$\hat{\theta}_{mean}^m = mean(\hat{\theta}[t_i^m:t_f^m]) \quad (3.12)$$

$$S^m = \frac{argmax(\hat{\theta}[t_i^m:t_f^m])}{N^m - argmax(\hat{\theta}[t_i^m:t_f^m])} \quad (3.13)$$

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).

Features were computed across all activity classes, excluding drink and discharge events, by sliding a window with specified SSW parameters across the training data. For pour and drink events, the window was centered at the midpoint of the GT interval label. Data was again partitioned using five-fold cross validation, with a variety of classification models evaluated. A cubic SVM classifier exhibited a maximum cross-validation accuracy of 97.5% and is used in all testing experiments. Adjacent windows identified as containing drinks were merged into a single observation interval in post-processing.

3.6.4 Performance Metrics

Performance was quantified by first mapping the midpoint of each estimated drink interval to the nearest GT interval, with each element of the GT interval considered only once. Next, error sets representing the underlap (U^j) and overlap (O^j) between the estimate and GT were defined using the non-commutative set difference operator. Localization error was measured as specified in (3.14), where $|\cdot|$ denotes the set cardinality operator.

$$E^j = \frac{(|U^j| + |O^j|)}{|\{t_s^j, t_{s+1}^j, \dots, t_e^j\}|} \quad (3.14)$$

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%. It should be noted that both the *SSW* and *TMD* algorithms were anticipated to produce some error for the 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 by the aforementioned mechanisms, potential inference regarding the estimated localization error is restricted.

3.7 Results

3.7.1 TR Testing

Both the *TMD* and *SSW* algorithms successfully 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*.

3.7.2 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 Figure 3-7. *SSW* error was again dominated by overestimation (63.5% avg.). Performance statistics for the *DL* experiments are consolidated in Table 3-3. Examples of error modes associated with *SSW* classification are depicted in Figure 3-8.

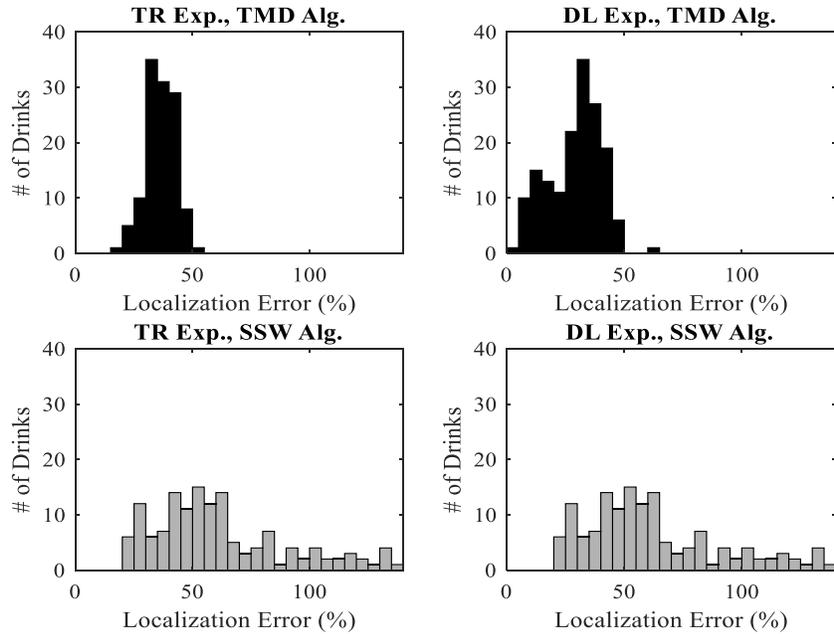


Figure 3-7: Localization Error Distributions

Table 3-3: Summary of DL Testing Performance

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

3.8 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 a previously introduced static sliding window classifier, the algorithm was demonstrated to improve sip detection performance while reducing computational overhead.

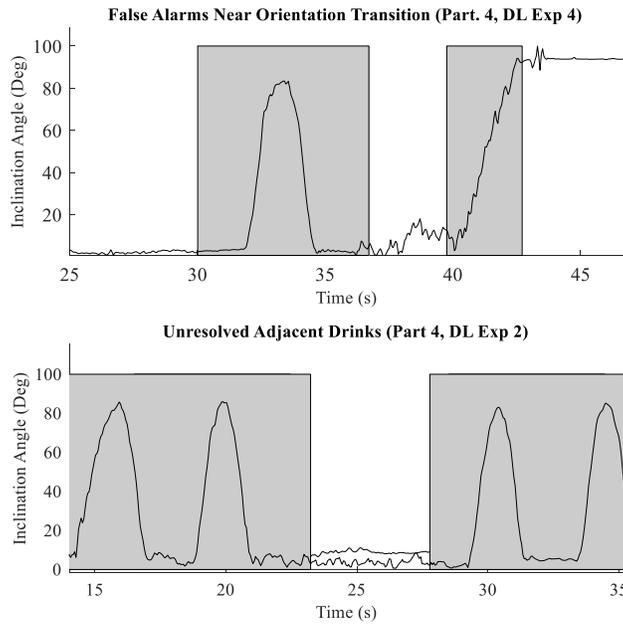


Figure 3-8: Example Error Modes – DL Experiments, SSW Algorithm

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

increases the number of required classification operations, it would likely improve generalization for larger data sets including more diverse drinks.

In addition to exploring these trade-offs, future work should also investigate the relationship between the employed drink spotting technique and the resulting volume estimations. Exploration of performance robustness with respect to sensor position, along with comparisons with wrist-worn IMU data, should be conducted. Finally, the utilization of training data obtained from daily-use scenarios should be investigated to support the deployment of models exploiting the temporal patterns of drinking events (i.e.: LSTMs, etc.).

Chapter 4 : The Inclination Signature Feature Set

4.1 Introduction

Previous motion-based approaches for estimating drink volume have achieved limited accuracy as noted in Chapter 1. Moreover, estimates have been shown to demonstrate considerable inter-subject variability. These previous models have utilized a limited description of the characteristic drinking motion pattern. For example, [11] described the drinking event using only its duration and the corresponding integral of two accelerometer channels. Research in [15] used a slightly expanded feature set for an attachable configuration. Namely, a four-element set including – 1) the duration of the drinking event, 2-3) the range and mean value of the accelerometer component parallel to the bottle’s axis, and 4) a measure of symmetry of the inclination and declination portion of the drink, was used. While both efforts qualitatively described the relationship between the reported feature space and bottle kinematics, direct estimation of the container’s inclination trajectory has not previously been explored within academic literature.

This chapter describes preliminary efforts to improve upon motion-based volume accuracy by leveraging the accelerometry-based container inclination estimation technique described in Chapter 3. In addition, a richer description of the resulting motion pattern during drinking is proposed. This representation uses both summary kinematic features, along with a low-resolution description of the variation in inclination through amplitude binning. The proposed technique is utilized throughout the remainder of this dissertation in the various estimation models explored.

This chapter begins with a review of reported volume estimation results in the literature. Approaches utilizing both volume and fill ratio estimates are presented. Next, details regarding the large-scale data collection conducted to support estimation efforts within this dissertation are

provided. A kinematically-inspired strategy for partitioning the entire captured motion sequence into transport and sip phases is also presented, followed by the proposed feature space description. Correlation with both volume and fill ratio are provided for the newly introduced feature set, along with the previously proposed four-element set in [15].

4.2 Data Collection

Eighty-four college-aged subjects (52 M, 32 F) completed 161 trials of an experiment requiring the consumption of 12 drinks from a refillable 750 mL bottle. Subjects were permitted to complete a maximum of four trials over multiple sessions. To begin the experiment, the bottle was filled to a consistent level as determined visually by the experimental proctor. To ensure that a variety of drink volumes were captured, subjects were instructed to consume either a small, medium, or large drink prior to each sip according to their personal preferences. The bottle was placed on an electronic scale following each drink, with the ground truth mass recorded manually in a spreadsheet. Variations from protocol were noted by the proctor to allow for removal in post-processing (i.e.: grasping and transporting the bottle without completing a drink, etc.). The ground-truth fill level from which each drink was consumed was estimated offline using an empirically determined mapping between changes in bottle mass and fill level reductions. Subjects consumed the entire original volume of water in seven trials, requiring refilling of the bottle during the experiment. Two trials were discarded after collection due to hardware failure, yielding a total valid data set of 159 trials (1,908 drinks).

All subject recruitment, data collection, and record storage was conducted according to protocol approved by the Institutional Research Board at Michigan State University. The univariate distribution of the initial fill ratio (fill level normalized to fillable height) and mass of each drink collected, along with their joint distributions, is depicted in Figure 4-1.

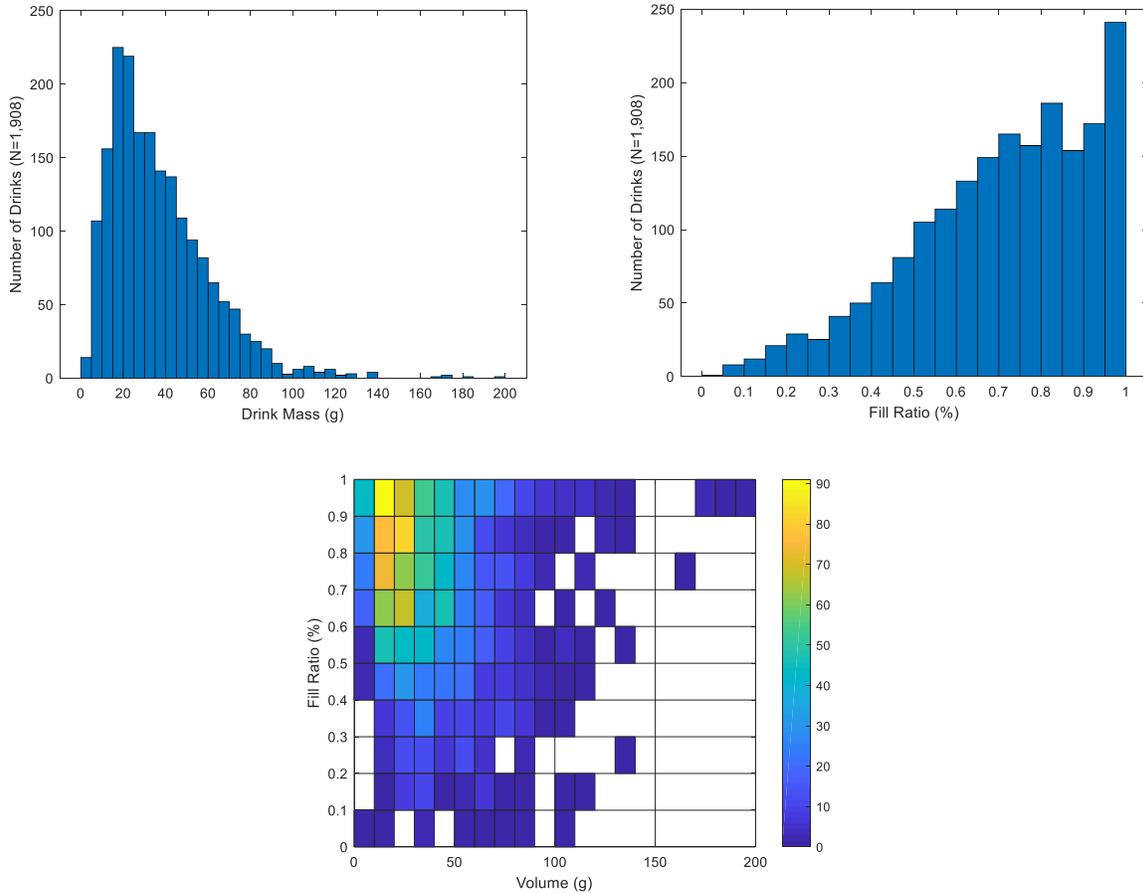


Figure 4-1: Univariate and Joint Distributions of Training Data

4.3 Pre-processing and Drink Segmentation

Data was collected using the sensor system described in Section 2.5. A sensor module was connected to the bottom of the bottle beneath the lid to avoid interference with grasping as depicted in Figure 1-1. A customized elastic strap with a Velcro connector was used to fasten the sensor to the bottle. For a subset of experiments, an additional sensor was attached midway up the bottle opposite the drinking hand. This second sensor was added to explore performance variability with respect to placement. Analysis of data from this additional sensor is reserved for future work.

To begin preprocessing, the bias of each component was estimated by averaging the initial 50 samples of each recording. During this time interval, the bottle was rested in a stationary vertical position. Portions of the signal corresponding to variations in protocol were then removed manually from the recordings using experimental annotations. Next, each file was parsed into drink events using a threshold-based algorithm exploiting the stationary placement of the bottle between drinks. This process captures the entire time interval for which the bottle was in motion (i.e.: both transport to and from the mouth, along with sipping).

After partitioning into drink events, signals were resampled to the target frequency of 20 Hz to account for variability in the base station polling interval. Smoothing was performed using a two-sample moving average filter to mimic the frequency response of the original work conducted in [15]. The sensor's inclination with respect to gravity, which is equivalent to that of the container under ideal sensor alignment, was then estimated under the assumption of negligible dynamic acceleration as specified in (3.1).

Variation in the estimated container inclination over an experimental trial is depicted in Figure 4-2. As volume is depleted from the container through sequential drinks, the maximum inclination associated with each sip increases.

4.4 Microevent Partitioning Strategy

As the parsing algorithm captures the entire motion interval of the container, further partitioning is necessary to isolate the drinking event from the transport phase. As described in similar work (i.e.: [26]), this segmentation is motivated by the substantial variation that may occur in the transport motion pattern depending upon the specific drinking scenario. For the experiments described herein, variability in handling between drinking events may be associated with the order of the drink within the trial (i.e.: more careful handling for full containers, more rapid transport as

the subject becomes familiar with protocol, etc.). In addition, differing orientation of the container upon retrieval may also introduce variability in the transport motion pattern. Due to the scripted nature of the experiments, such variability is anticipated to be negligible versus that encountered during daily living scenarios.

To isolate the drinking portion of the event, the asymmetry of the container about its axis is exploited. Namely, as the lid of the container encourages consumption from the opposite edge, we hypothesize that the transport phase will involve rotations about the axis of the bottle as necessary to achieve the desired drinking orientation. This orientation is reflected in the sensor's position within the cross-sectional plane of the bottle, and may be estimated by computing the orientation of the resultant component of the static acceleration due to gravity as specified in (4.1).

$$\alpha = \tan^{-1}\left(\frac{a_z}{a_y}\right) \quad (4.1)$$

As depicted for a random sample of drinks in Figure 4-3, α maintains a stationary value near the center of each drinking event, corresponding to the hypothesized lack of axial rotation of the container during sipping.

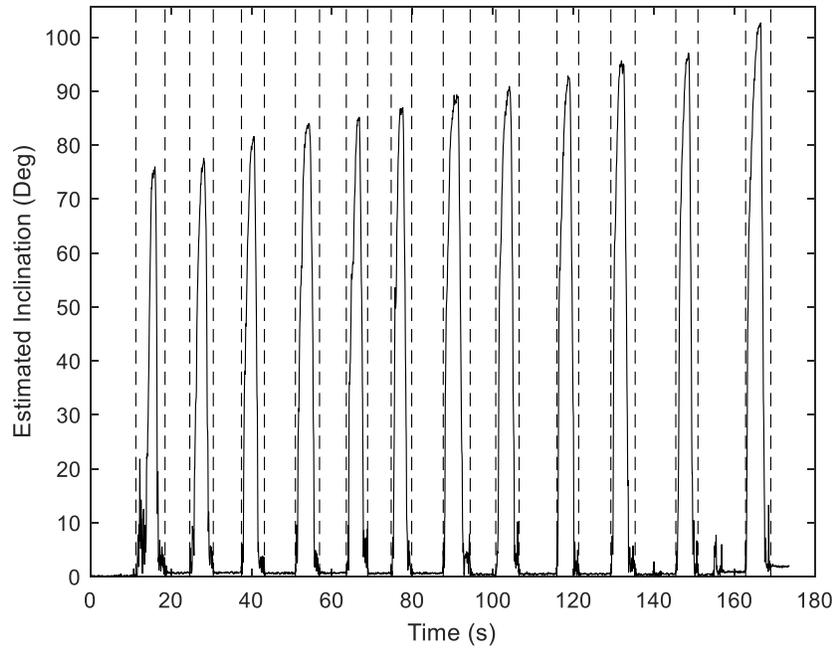
For preliminary analysis, the interval for which the sensor remains in this position is defined as the sip micro-event, yielding an aggregate micro-event partition defined as follows:

Lift: The portion of the macro-event preceding the sip micro-event

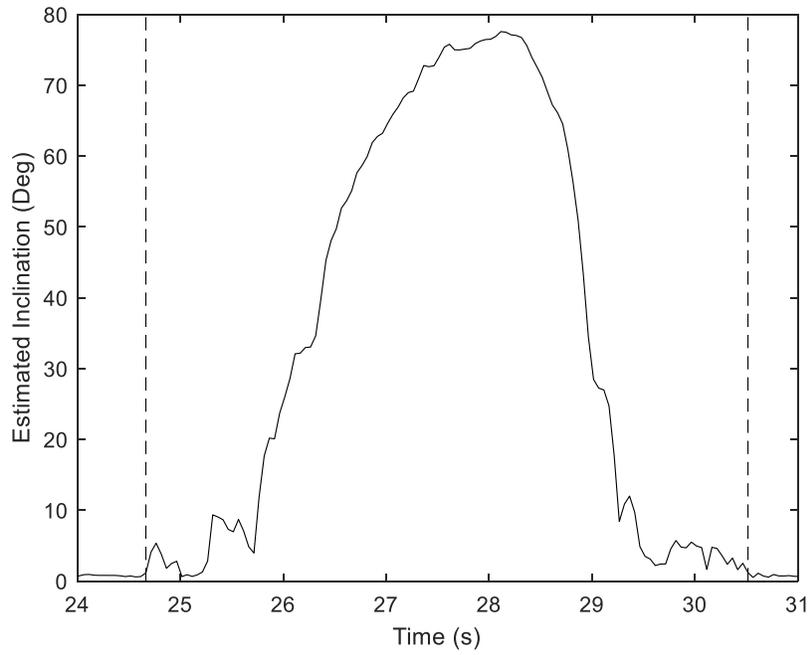
Sip: The portion of the micro-event for which the cross-sectional sensor placement is estimated as stationary

Place: The remainder of the macro-event after termination of the sip micro-event

Strategies for further isolating the time period for which fluid is entering the mouth will be explored in future work.



(a) Wide View



(b) Zoom View

Figure 4-2: Variation in Estimated Container Inclination Over Experimental Trial

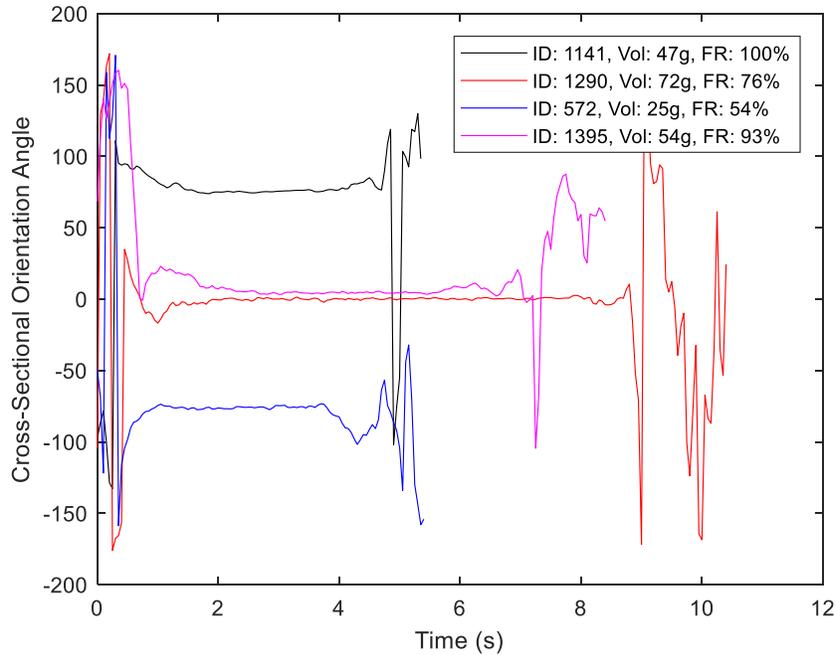


Figure 4-3: Variation in Coplanar Sensor Orientation During Randomly Chosen Drinks

To estimate the duration of the sip micro-event, a threshold-merge algorithm with empirically determined parameter values was applied on the sample-over-sample difference of α . The difference signal was initially thresholded to a maximum value of 8 degrees. All intervals meeting the threshold criteria which were separated by less than 2 samples were merged to a continuous interval, with the largest interval extracted as the sip micro-event. The resulting micro-partition for the four random drink events depicted in Figure 4-3 is shown in Figure 4-4.

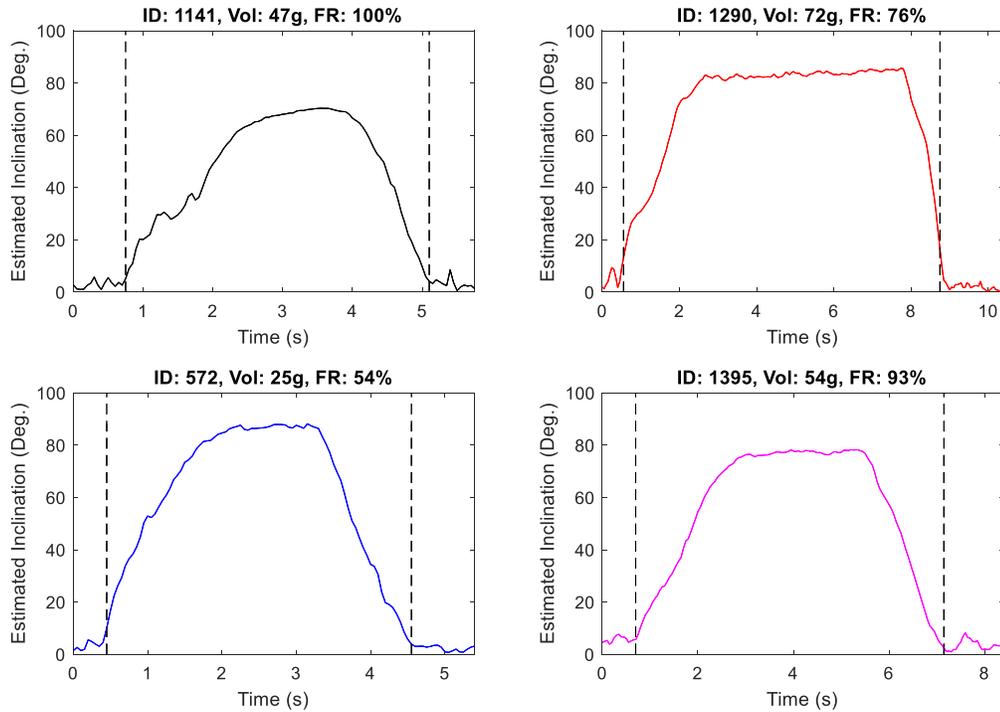


Figure 4-4: Variation in Container Inclination During Randomly Chosen Drinks

As shown in Table 4-1, sip duration is more strongly correlated with volume versus the two transport durations. The correlation between sip duration, along with the previously proposed motion feature related to the integral of the inclination [11], are shown in Table 4-2 for various ranges of controlled fill levels.

Table 4-1: Correlation Between Features and Volume Label

Micro-event Duration	Pearson Correlation Coefficient (Corr. Coeff.) (Entire Dataset)
<i>Lift Duration</i>	0.189
<i>Sip Duration</i>	0.449
<i>Place Duration</i>	0.159

Table 4-2: Correlation Between Previously Reported Motion Features and Volume

Motion Feature	Corr. Coeff. (Entire Dataset N = 1,908)	Corr. Coeff. (FR > 50% N = 1,576)	Corr. Coeff. (FR > 70% N = 1,075)	Corr. Coeff. (FR > 90% N = 413)
<i>Sip Duration</i>	0.449	0.457	0.471	0.557
<i>Integral of Inclination Over Sip Duration</i>	0.536	0.543	0.571	0.672

This two-factor description of the motion pattern captures two degrees-of-freedom which may be utilized by subjects to control the amount of fluid consumed (i.e.: drink duration and container inclination). Observations regarding the relationship between these motion factors and volume are consistent with [11], which reported a correlation coefficient with drink volume of 0.69 and -0.60/-0.55 for sip duration and the integral of accelerometer signals not parallel to the wrist. Moreover, the strength of correlation between both features and volume increases when fill level is restricted within a narrower range of values. This increasing strength of relationship supports the prior observation of the interdependence of volume and fill level on the resulting motion signature.

4.5 Feature Engineering

Based upon examination of the estimated inclination curves, along with motion observations during data collection, a set of hand-engineered features describing the drinking kinematics were hypothesized. In addition to key kinematic quantities (i.e.: maximum inclination, maximum rate of inclination, etc.) and their associated statistical moments, amplitude values of both the raw and normalized curves were binned to create a low-level time-invariant feature description of the signal. This description, hereby denoted as the inclination feature (IS) set, is summarized in Table 4-3.

Table 4-3: Inclination Signature (IS) Feature Set

Feature ID	Feature Symbol	Feature Definition	Description
1	θ^*	$\max(\theta^j)$	Maximum inclination angle during drink event
2	D	$\text{length}(\theta^j)$	Duration of drinking event
3-11	A_k	$\text{count}(\theta^j < T(k)), k = 1$ $\text{count}(T(k-1) \leq \theta^j < T(k)), k \in \{2,8\}$ $\text{count}(\theta^j \geq T(k-1)), k = 9$ $T(k) \in \{20^\circ, \dots, 90^\circ\}$	Number of samples for which inclination angle satisfies specified amplitude range criteria
12-20	AR_k	$\text{count}(\theta^j/\theta^* < P_k), k = 1$ $\text{count}(P_{k-1} \leq \theta^j/\theta^* < P_k), k \in \{2,8\}$ $\text{count}(\theta^j/\theta^* \geq P_{k-1}), k = 9$ $P[k] \in \{20\%, \dots, 90\%\}$	Number of samples for which normalized inclination angle satisfies relative amplitude criteria
21	Q_θ	$\frac{\theta^*}{\text{length}(\theta^j)}$	Ratio of maximum inclination value to duration
22	$\bar{\theta}$	$\text{mean}(\theta^j)$	Mean inclination angle
23	D_{θ_R}	$\frac{\text{argmax}(\theta^j)}{N^j - \text{argmax}(\theta^j)}$	Ratio of time for which inclination angle is increasing relative to decreasing
24-25	S_θ, S_θ^T	$\sum_{m=1}^U \theta_m^j T$	Riemann sum approximation to integral of inclination curve over entire duration ($U = N^j$) or inclination interval ($U = \text{argmax}(\theta^j)$)
26	RE_θ	$\frac{\max(\theta^j) - \theta_1}{\text{argmax}(\theta^j) - 1}$	Slope of line intersecting inclination trajectory start of trajectory time of maximum value
27	FE_θ	$\frac{\theta_{N^j} - \max(\theta^j)}{N^j - \text{argmax}(\theta^j)}$	Slope of line intersecting inclination trajectory at time of maximum value and end of trajectory
28/29	θ_T^*/θ_A^*	$\frac{\max(\theta(1:\text{argmax}(\theta^j))'^j)}{\max(\theta(\text{argmax}(\theta^j):N^j)'^j)}$	Maximum rate of inclination/declination, where θ' is a numerically estimate of the derivative of θ
30/31	$\bar{\theta}_T'/\bar{\theta}_A'$	$\frac{\text{mean}(\theta(1:\text{argmax}(\theta^j))'^j)}{\text{mean}(\theta(\text{argmax}(\theta^j):N^j)'^j)}$	Mean rate of inclination/declination
32/33	$s_{\theta_T}'/s_{\theta_A}'$	$\frac{\text{std}(\theta(1:\text{argmax}(\theta^j))'^j)}{\text{std}(\theta(\text{argmax}(\theta^j):N^j)'^j)}$	Standard deviation of inclination/declination rate

To explore the relationship of the proposed feature set with each recorded label of interest (i.e.: volume and fill ratio), the Pearson correlation coefficient between each element and the label were computed as specified in Table 4-4. As noted, the strongest correlation ($\rho = 0.544$) with volume is associated with feature 24, which corresponds to the integral of the inclination curve from the beginning of the event until the maximum value is reached. Strong correlation is also observed for feature 25 ($\rho = 0.53$), which corresponds to the integral of the inclination over the remaining portion of the event, along with feature 20 ($\rho = 0.477$), which corresponds to the number of samples for which the inclination exceeds 90% of the maximum value. The only other feature exhibiting a correlation exceeding 0.4 ($\rho = 0.460$) is feature 2, which corresponds to the entire event duration.

Table 4-4: Correlation Between IS Feature Set and Volume/Fill Ratio Labels

Feature ID	1	2	3	4	5	6	7
ρ_{Vol}	0.364	0.460	0.178	-0.057	-0.101	-0.148	-0.162
ρ_{FR}	-0.834	0.009	0.245	0.445	0.491	0.420	0.386

Feature ID	8	9	10	11	12	13	14
ρ_{Vol}	-0.062	0.126	0.336	0.376	0.215	0.061	0.006
ρ_{FR}	0.378	0.249	-0.341	-0.582	0.166	0.325	0.286

Feature ID	15	16	17	18	19	20	21
ρ_{Vol}	0.048	0.042	0.088	-0.122	0.307	0.477	-0.182
ρ_{FR}	0.108	-0.063	-0.167	-0.184	-0.155	-0.168	-0.390

Feature ID	22	23	24	25	26	27	28
ρ_{Vol}	0.384	0.316	0.544	0.530	-0.194	-0.010	0.069
ρ_{FR}	-0.659	0.085	-0.390	-0.340	-0.339	0.400	-0.421

Feature ID	29	30	31	32	33
ρ_{Vol}	0.006	-0.192	-0.013	0.028	0.086
ρ_{FR}	-0.057	-0.345	0.405	-0.388	-0.432

While the relationship between volume and both the integral of inclination and event duration have been previously noted [11], the relationship between the relative threshold feature (20) has not been reported. It is hypothesized that the strength of the relative binned value versus the absolute binned value (feature 11, $\rho = 0.376$) is associated with the previously described increase in requisite maximum inclination with declining fill ratio. As the volume remaining in the bottle decreases, the required inclination to induce fluid flow increases. Therefore, it is expected that the relationship between drink volume and inclination amplitude would be more pronounced in a relative amplitude sense.

Correlation between fill ratio and the various elements of the feature space is most demonstrated by feature 1 ($\rho = -0.834$). This observation is consistent with the qualitative observation in the prior paragraph. Namely, as the fill level of the bottle decreases upon depletion of volume, the maximum inclination associated with a drink event increases. Strong fill ratio correlation is also exhibited for feature 22 ($\rho = -0.659$), which corresponds to the mean value of inclination, along with feature 11 ($\rho = -0.582$), which corresponds to the time duration where the inclination exceeds 90 degrees. Observed correlations with the feature space are generally stronger for the fill ratio versus volume label.

For purposes of comparison, correlations with the two labels of interest are computed for the four-element feature set previously proposed for a container-attachable IMU in [15]. These features are defined in Table 4-5, with correlation values presented in Table 4-6. As noted, the observed label correlation of feature 1 in the IS set (maximum amplitude) and 1L in the legacy set (range of axial component of the accelerometer) is similar. This supports the prior observation in [15] that this quantity is related to inclination, and is verified by expressing (3.1) in terms of a decomposition involving solely this component and the resulting static acceleration due to gravity.

Moreover, this equivalent is demonstrated in comparing the observed relation for feature 23 and 4L, along with 3L and 22.

Table 4-5: Legacy Feature Set

Feature ID	Feature Symbol	Feature Definition	Description
1L	H	$range(\mathbf{a}_{ax}^j)$	Range of axial accelerometer signal during drinking
2L	D	$length(\mathbf{a}_{ax}^j)$	Duration of drinking event
3L	M	$mean(\mathbf{a}_{ax}^j)$	Mean value of axial accelerometer signal during drinking
4L	RS	$\frac{argmax(\mathbf{a}_{ax}^j)}{N^j - argmax(\mathbf{a}_{ax}^j)}$	Ratio of time for which inclination angle is increasing relative to decreasing

Table 4-6: Correlation Between Legacy Feature Set and Volume/Fill Ratio Labels

Feature ID	1L	2L	3L	4L
ρ_{Vol}	0.307	0.460	-0.416	0.316
ρ_{FR}	-0.792	0.009	0.690	0.085

4.6 Summary and Future Work

Details regarding the large-scale data collection conducted to support volume estimation efforts within the remainder of this manuscript was described herein. Moreover, a hand-engineered feature set describing the container’s inclination during drinking was introduced, with the relationship between the two labels of interest (volume and fill ratio) explored.

As quantified by the Pearson correlation coefficient, the proposed motion features generally exhibited a stronger linear relationship with fill ratio versus volume labels. Prior observations noting the relationship between both drink duration and the integral of inclination were also verified. Finally, the correlation between the labels of interest and a legacy feature set previously proposed for the attached sensor architecture was explored. Future work estimating both the

volume and fill ratio in the remainder of this dissertation compares estimation accuracy between the two feature sets.

Chapter 5 : Drink Volume Estimation Using Regression Models

5.1 Introduction

Support vector machine (SVM) models for estimating drink volume on both an individual and multi-drink basis are described within this chapter. Models utilize the hand-engineered inclination signature (IS) feature space described in the previous chapter. Results are verified using the large-scale data collection described in Chapter 4, with an analysis framework chosen to promote comparability with similar work conducted in [11]. Results are benchmarked against previously proposed linear regression (LR) motion models, along with SVMs employing the four-element benchmark feature set proposed in [15].

5.2 Data Partitioning

Leave-one-trial-out (LOTO) validation was performed for all analysis conducted as the primary method of analysis within this chapter. This approach is consistent with the target use case, where models trained on a broad pool of users would be employed on a new user absent of customization. While a LOTO approach allows for the inclusion of some subject-specific training data, the magnitude of this contribution is limited (i.e.: maximum subject-specific training data of 1.9% for scenarios where subjects completed the maximum number of trials).

A set of support vector machine (SVM) regression models with varying kernel functions were trained for both volume and fill ratio labels. Linear, medium (kernel scale = 5.7) and coarse (kernel scale = 23) Gaussian kernel functions were considered. Hyperparameters were set to the default values as established in MATLAB's Regression Learner toolbox. SVMs were chosen for initial analysis based upon their superior performance for the current sensor architecture in [15]. Alternative regressor models should be explored in future work.

For purposes of benchmarking, LR models utilizing only the previously described characteristic motion features (i.e.: sip duration and integral of inclination) are also evaluated. While motivated by the methods of [11], it should be reemphasized that direct comparison is not applicable. Namely, differences in both sensor placement (i.e.: wearable versus attachable), along with utilization of the estimated container inclination (as opposed to the raw accelerometer signals) within the integrand distinguishes the two results. In addition, SVMs using the four-element feature set proposed in [15] for an attachable architecture are also evaluated.

5.3 Performance Metrics

Multiple performance metrics are used to assess the quality of the estimation models assessed herein. Mean Absolute Percentage Error (MAPE) is employed to quantify estimation performance on a per-drink basis. MAPE was chosen over alternative measures (i.e.: root mean squared error, etc.) due to its utilization in prior work (i.e.: [11], [25]).

To assess estimation quality over a series of drinks, Mean Overall Absolute Percentage Error (MOAPE) was used. Similar to the overall error (OE) metric described in [11], MOAPE allows for cancelation of estimation errors across consecutive drinks within a single trial. However, MOAPE takes the absolute value before averaging across participants to avoid overstating performance through cancelation of errors across trials. While MAPE provides the most rigorous assessment of model performance, MOAPE is useful for exploring utility in practical scenarios where aggregate consumption is of primary concern (i.e.: estimating total daily consumption, etc.).

5.4 Volume Estimation Results

Volume MAPE is depicted in Figure 5-1 for the various models considered. Models computed on the sip interval only are labeled as Stat., with all other reported results are computed

on the entire drink duration. Consistent with wearable results in [11], LR models employing the integral of inclination outperform those using duration. The level of improvement is enhanced versus results presented in [11]. We hypothesize that this difference is associated with use of the inclination estimate of the container, as opposed to the individual accelerometer channels which are hypothesized as being related to this quantity in [11].

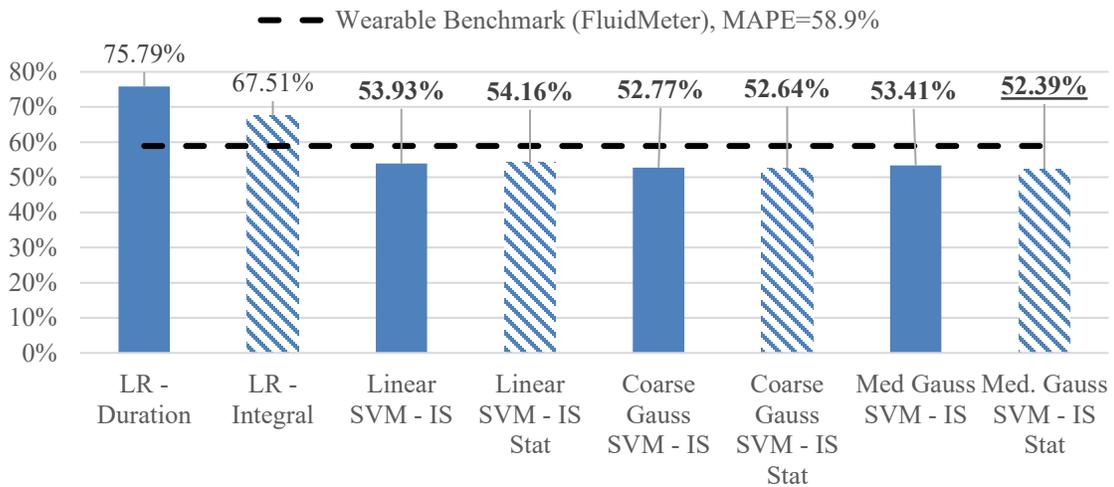


Figure 5-1: Variation in Volume MAPE for Various Models Considered

All SVM models outperformed the simplistic single factor LR motion models. Moreover, all SVM models exhibited superior performance to the previous best-case reported MAPE of 58.9% for a single wearable sensor in an experiment using scale-based ground-truth described in [11]. Only minimal differences in MAPE were observed for models utilizing the proposed sip micro-event segmentation versus those computed on the entire drinking event.

Comparison of volume MAPE for SVM models using both the IS and benchmark feature set are shown in Figure 5-2. As depicted, the expansion of the feature set improves average accuracy across kernel functions by 5.78%.

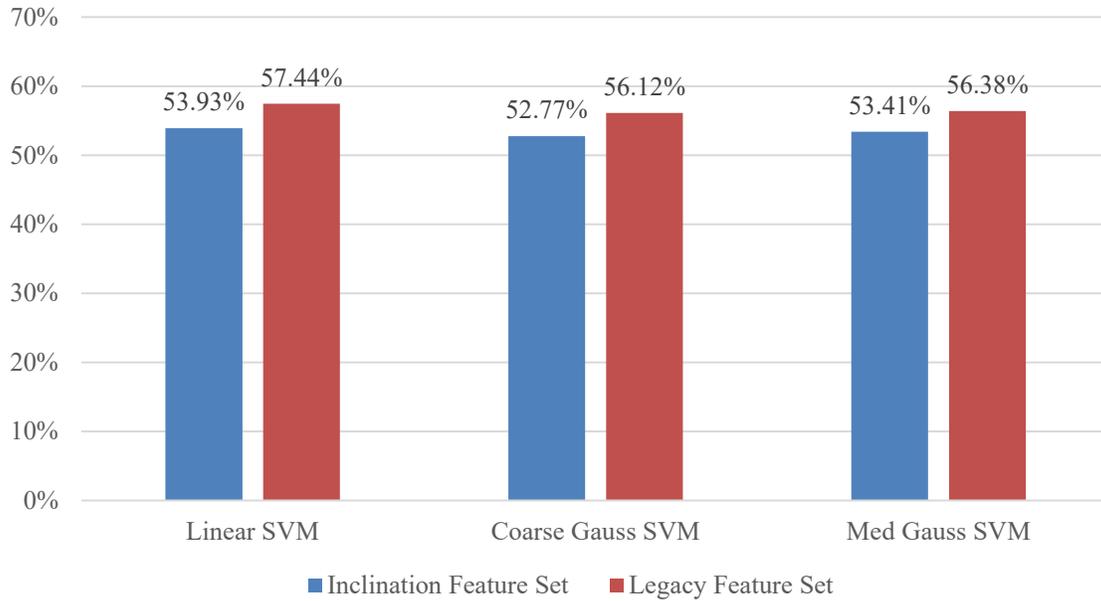


Figure 5-2: Variation in Volume MAPE Across Feature Sets

Variation in MAPE across trials is depicted in Figure 5-3 for the best-case volume estimator (medium kernel, sip micro-event partition). Consistent with prior observations [11], dispersion in the observed error metric is substantial, with a standard deviation of 28.18%.

Volume MOAPE for varying drink sequence lengths is presented in Table 5-1. Aggregate estimation accuracy generally improves with increased sequence length, with reductions more pronounced for the proposed IS-based SVM models. While not directly comparable due to the employment of the more stringent MOAPE cumulative metric herein, the best-case aggregate consumption estimation accuracy of 19.49% is improved versus the average value of 25% reported in [15] for a container-attachable IMU.

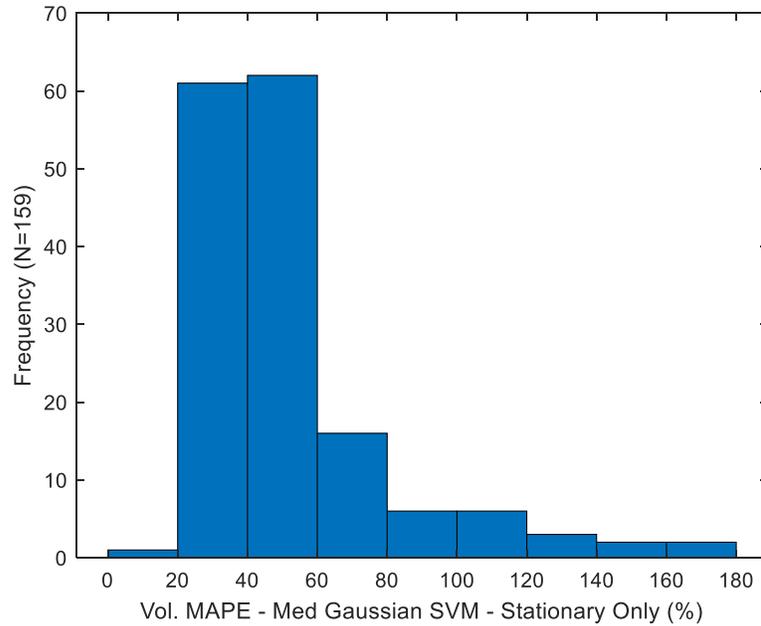


Figure 5-3: Distribution of Volume MAPE for Best-Case Estimator

Variation between SVM models employing the IS and legacy feature sets exhibited negligible difference in the MOAPE metric across the various sequence lengths considered. While the best-case MOAPE(12) value exceeds that computed for the in-the-wild data set reported in [11] (16.95%), direct comparability is limited by the inclusion of potential sip detection related errors (i.e.: both false alarms and missed drink detections) in this latter metric, along with the utilization of a commercial smart-bottle for ground-truth labeling (a description on the manufacturer’s website reports an accuracy of “a fraction of an ounce” for this device). Moreover, differences in MOAPE between the IS and legacy feature space descriptions of inclination is negligible.

Table 5-1: Variation in Volume MOAPE for Multiple Prompt Periods

Model Identifier	MOAPE(3)	MOAPE(6)	MOAPE(9)	MOAPE(12)
<i>Duration Only – LR</i>	36.74%	34.41%	33.51%	32.42%
<i>Integral Only – LR</i>	<u>28.68%</u>	27.76%	27.59%	27.79%
<i>IS – Linear SVM</i>	32.87%	26.40%	23.44%	21.46%
<i>IS Stat. – Linear SVM</i>	33.74%	26.64%	23.52%	21.56%
<i>Legacy – Linear SVM</i>	32.43%	26.95%	24.07%	22.05%
<i>IS – Coarse Gaussian SVM</i>	31.55%	25.49%	22.58%	20.75%
<i>IS Stat. – Coarse Gaussian SVM</i>	31.79%	25.39%	22.48%	20.65%
<i>Legacy – Coarse Gaussian SVM</i>	33.17%	27.46%	24.00%	21.45%
<i>IS – Medium Gaussian SVM</i>	30.52%	24.98%	<u>21.62%</u>	19.64%
<i>IS Stat. – Medium SVM</i>	30.55%	<u>24.86%</u>	21.71%	<u>19.49%</u>
<i>Legacy – Medium Gaussian SVM</i>	34.30%	27.95%	23.62%	20.92%

Variability of the best-case aggregate estimator (medium kernel, entire macro-event duration) is presented in Figure 5-4. Similar to the MAPE metric, inter-subject variability is considerable (standard deviation of 14.75%). For purposes of comparison, standard deviation across participants for the in-the-wild dataset in [11] was 14.17%.

5.5 Individual-Specific Volume Estimation Results

While not feasible for practical deployment, individual-specific models were also evaluated for purposes of comparability with [11]. These models are trained on a leave-one-drink-out (LODO) basis per trial.

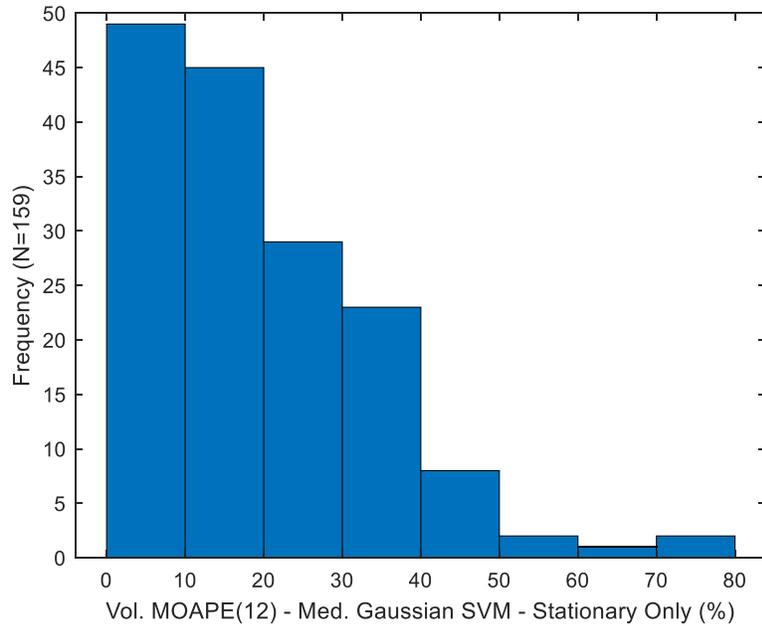


Figure 5-4: Distribution of Volume MOAPE(12) for the Best-Case Estimator

Namely, for each drink in a trial, a prediction was made using regression models trained on the additional 11 drinks. Only the two LR models were evaluated due to the aforementioned intent of this analysis.

A volume MAPE of 55.16% was observed for a subject-specific duration LR model. This corresponds to an 20.63% absolute reduction versus a duration-based LR model trained in a LOTO framework. For purposes of comparison, a duration-based volume MAPE of 64.8% was reported in [11] for a wearable IMU employing subject-specific training.

Variation in volume MAPE is depicted in Figure 5-5 for the subject-specific duration-based model. The dispersion of this metric is reduced substantially versus models trained in a LOTO framework (7.39% standard deviation for subject-specific model versus 58.98% for LOTO model). Variation in duration-based volume MAPE across trials for both training techniques considered is shown in Figure 5-6.

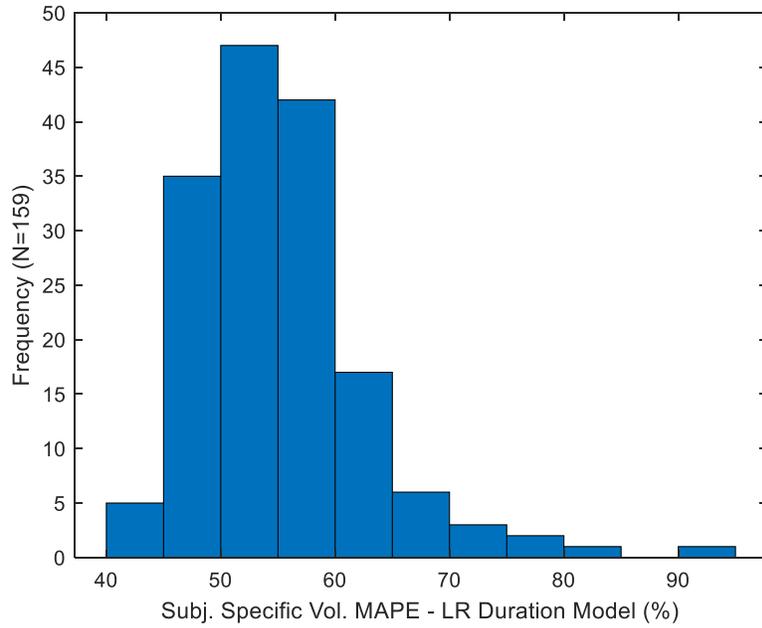


Figure 5-5: Distribution of Volume MAPE Across Trials for Subject-Specific Duration Model

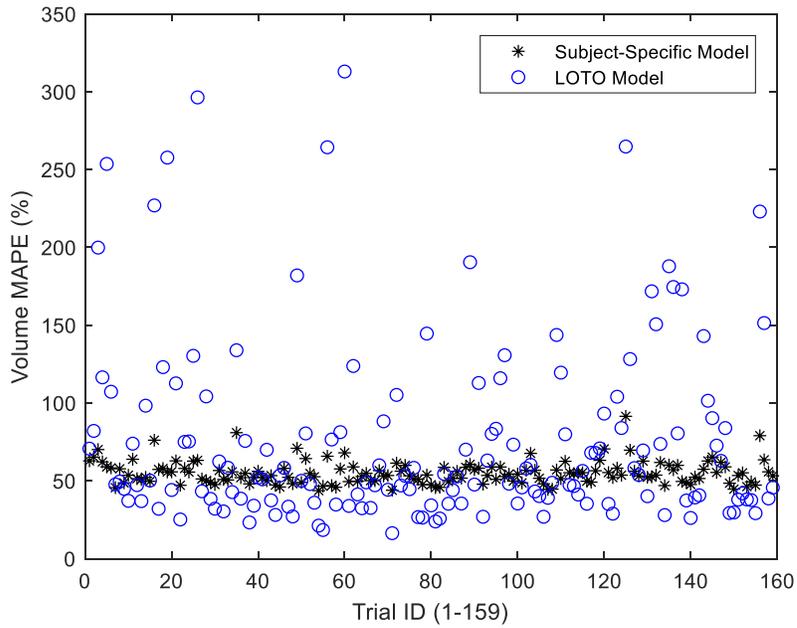


Figure 5-6: Variation in Duration-Based Volume MAPE Across Trials

A volume MAPE of 54.70% was observed for a subject-specific integration-based LR model, corresponding to a 12.81% absolute reduction versus the LOTO-trained model. For

comparison, a MAPE of 29.1% was reported for an integration-based subject-specific model for the wearable sensor in [11]. Variation in subject-specific volume MAPE is depicted in Figure 5-7. As was the case for duration-based models, the dispersion of this metric is also reduced considerably versus models trained in a LOTO framework (7.31% standard deviation for a subject-specific model versus 47.13% for a LOTO model). Variation in integration-based volume MAPE across trials for both individual-specific and LOTO models is depicted in Figure 5-8.

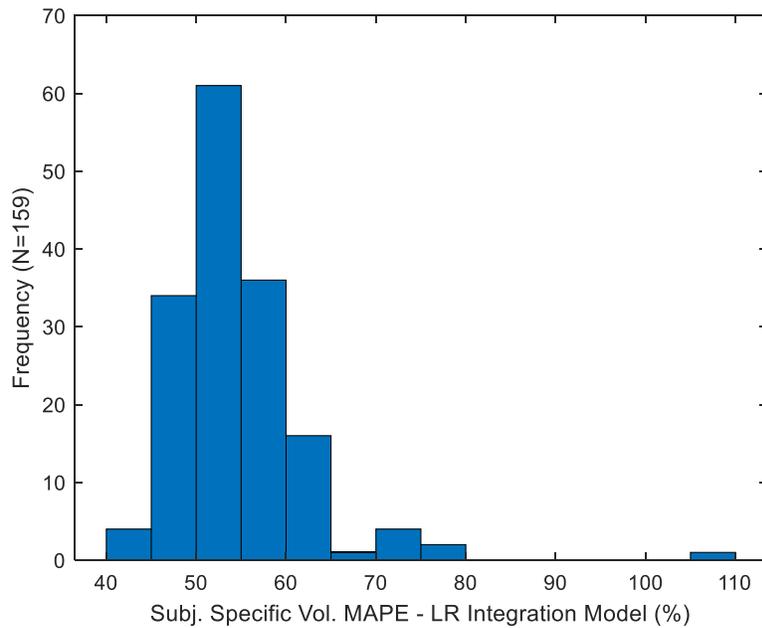


Figure 5-7: Distribution of Volume MAPE for Subject-Specific Integration Model

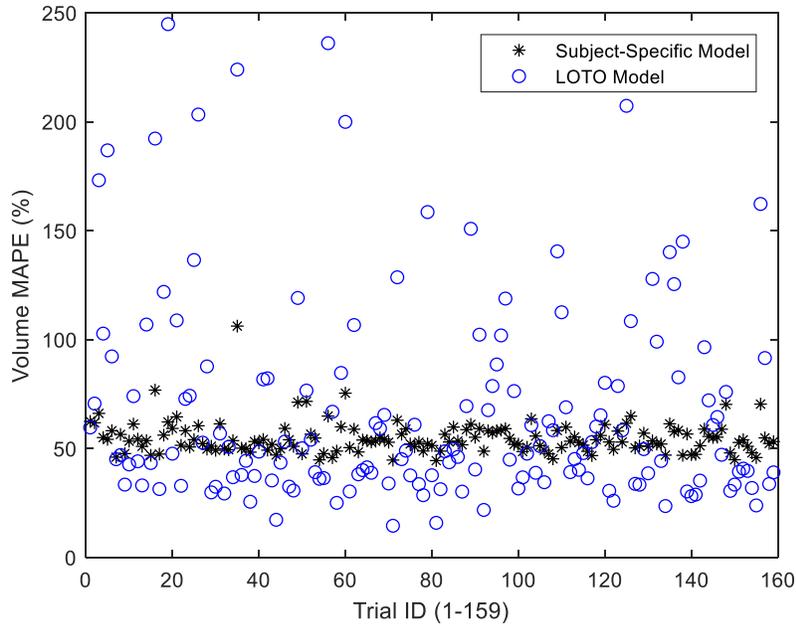


Figure 5-8: Variation in Integration-Based Volume MAPE Across Trials

5.6 Discussion

A scatter plot of the best-case (medium kernel, sip micro-event partition, LOTO training) predicted versus ground-truth volume is shown in Figure 5-9. While a general linear relationship between the estimated and ground-truth volume is observed, as quantified by a coefficient of determination of 77% for the best-fit linear mapping between the two quantities, accuracy is still limited. The relative performance improvement for subject-specific models suggests that this limited accuracy may be attributed to subject-specific factors influencing drink volume, such as the shaping of the mouth during fluid intake.

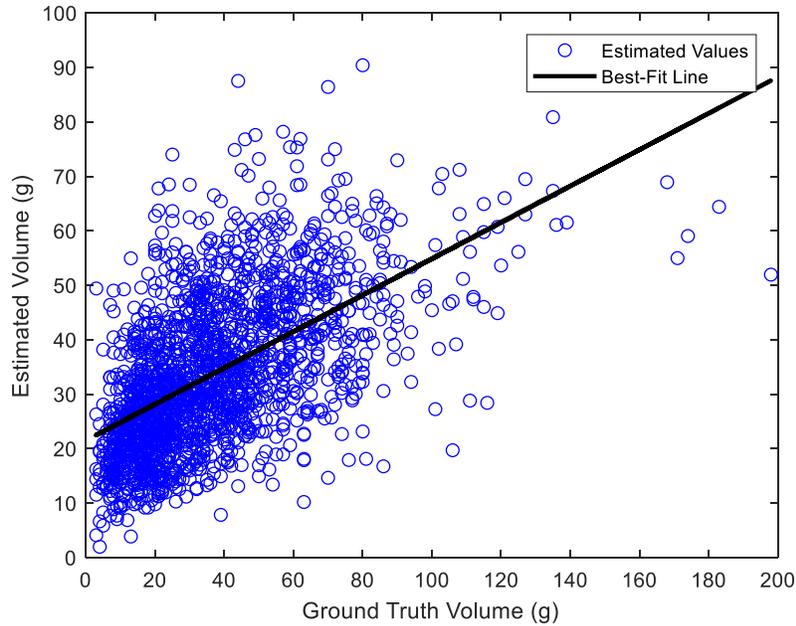


Figure 5-9: Scatter Plot of Estimate Versus Ground-Truth Volumes for Best-Case Estimator

5.7 Summary and Future Work

Support vector machine regression models for estimating drink volume were explored herein. The models utilized the hand-engineered IS feature space introduced in Chapter 3. Using a large-scale data collection consisting of 1,908 drinks consumed by 84 participants, mean absolute percentage error (MAPE) was reduced by 11.07% versus previous-state-of-the-art results for a single IMU sensor using a similar experimental set-up [11]. Moreover, measurements of aggregate consumption were reduced versus the previously reported best-case estimates for the container-attachable architecture [15]. Consistent with prior motion-based volume estimation results, accuracy was generally limited and exhibited considerable inter-subject variability. Namely, the best-case volume MAPE exhibited a standard deviation of 28.22% across trials. While subject-specific models were shown to enhance accuracy and reduce variability, the requirement of personalized training data limits the feasibility of implementing such models in practice.

Future work should focus on employing more sophisticated learning models for sip detection. While alternative models were explored as part of this research (i.e.: tree structures, Gaussian Regression Processes, end-to-end deep learning models, etc.), support vector machine approaches exhibited superior performance. This observation is consistent with the preliminary work described in [15]. The performance of more sophisticated models may improve with an expansion of training data. Specifically, collections which enhance the density in observations across fill ratio and volume may improve model generalization.

Chapter 6 : Aggregate Consumption Estimation

6.1 Introduction

As described in Chapter 1, augmented containers which estimate consumption using changes in the total amount of fluid within the container have been proposed. Technologies employing this approach are currently available in the commercial marketplace. For example, the Trago bottle cap utilizes sonar technology to estimate the fill level to a specified accuracy of fractions of an ounce [15].

The research described in this chapter explores the feasibility of this technique using learning-based fill level estimates obtained from the proposed sensor architecture. Support vector machine regressors employing the IS feature set introduced in Chapter 4 are used for fill ratio estimation. While low resolution fill level classification has been suggested in prior academic literature for more complex sensing architectures, we are unaware of the application of such techniques using a high-resolution regression framework [2]. Low-resolution fill level classification is explored in Chapter 9 for the current sensor architecture across multiple types of drinking vessels.

This chapter follows a similar structure to Chapter 4, with an initial presentation of fill ratio estimation accuracies achieved for the various models considered. The residual volume technique is then formalized, with corresponding volume estimation accuracies presented. A multi-target approach for integrating fill ratio information within the volume estimation process is then proposed. The chapter concludes with a summary and suggestions for future work.

6.2 Data Partitioning and Performance Metrics

The leave-one-trial-out (LOTO) validation approach applied in Chapter 5 for volume estimation is used in the current chapter. Limited subject-specific analysis is also provided. The various performance metrics identified in the prior chapter are also utilized.

6.3 Fill Ratio Estimation Results

Variation in fill ratio MAPE for the multiple models considered is depicted in Figure 6-1. Sip duration was replaced with maximum inclination for a single-factor LR benchmark model due to its strong correlation with fill ratio. This relationship is emphasized by the variation in this quantity over the course of an experiment as shown in Figure 4-2. Fill ratio estimation accuracy is greatly improved versus volume prediction for both the single factor regression and more complex SVM models. A comparison of accuracy for SVM models implemented using the IS and legacy feature set is shown in Figure 6-2. IS models outperform legacy models for all kernel functions considered, with an average improvement of absolute 13.0% across kernels.

Variability in MAPE for the best-case estimator (coarse kernel, entire macro-event partition) is shown in Figure 6-3. Error dispersion across trials is greatly reduced versus volume estimators. In particular, fill ratio MAPE standard deviation is 3.39%, versus 28.18% for the best-case volume MAPE.

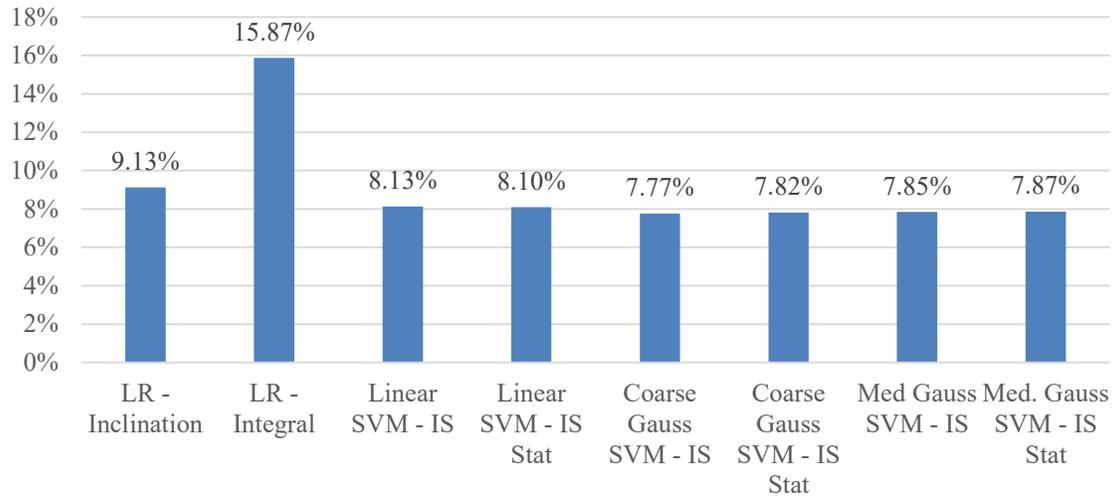


Figure 6-1: Variation in Fill Ratio MAPE for Various Models Considered

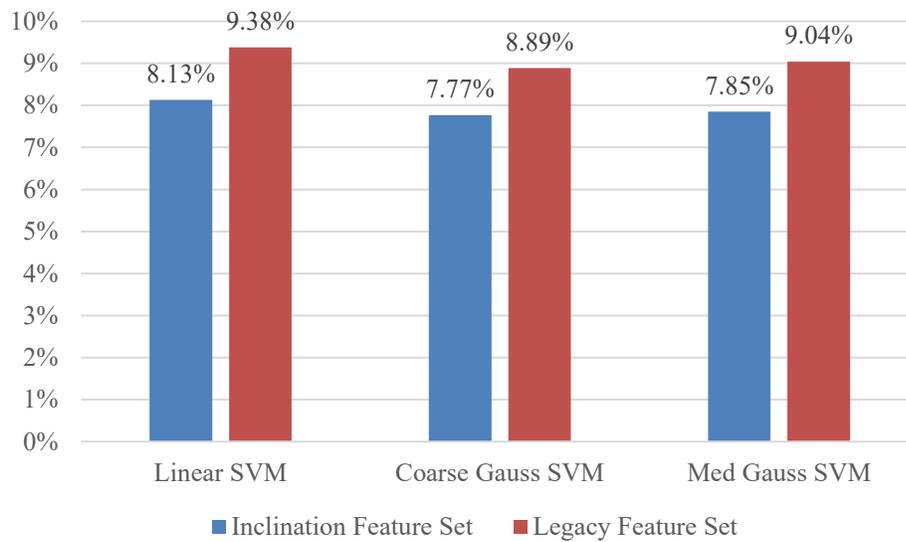


Figure 6-2: Variation in Fill Ratio MAPE Across Feature Sets

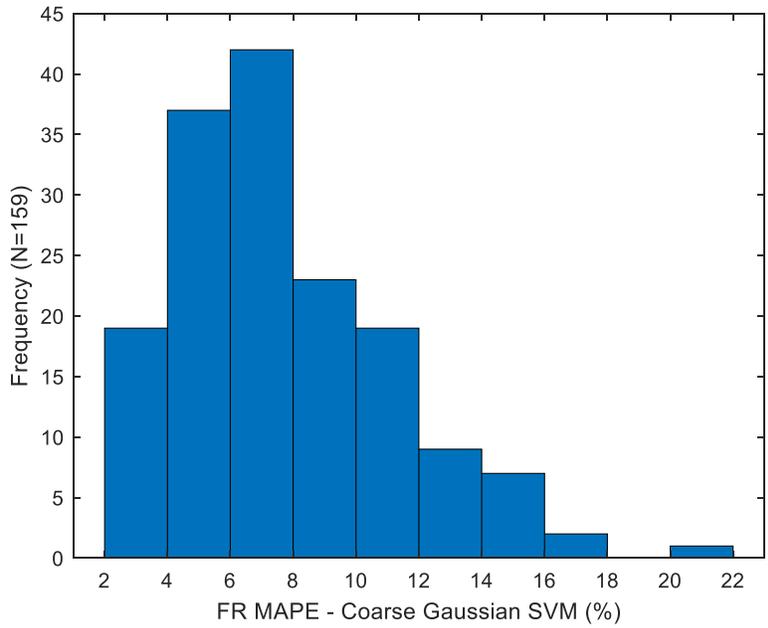


Figure 6-3: Distribution of Fill Ratio MAPE Across Trials for Best-Case Estimator

Fill ratio MOAPE is shown in Table 6-1 for varying drink sequence lengths. In contrast to volume estimation, the point nature of fill ratio estimates does not allow for sequential error cancellation across multiple drinks. Minimal accuracy is observed for the 12-drink sequence. This may be associated with the aforementioned skewing of training data towards larger fill ratios.

Variability in fill ratio MOAPE(12) estimates across trials is depicted in Figure 6-4 for the best-case estimator (coarse kernel, sip micro-event), with a standard deviation of 8.58% observed (versus 14.75% for volume MOAPE(12) estimates).

Table 6-1: Variation in MOAPE for Multiple Prompt Periods – Fill Ratio Estimation

Model Identifier	MOAPE(3)	MOAPE(6)	MOAPE(9)	MOAPE(12)
<i>Max. Inclination Only – LR</i>	10.90%	<u>7.64%</u>	8.12%	12.57%
<i>Integral Only – LR</i>	18.20%	9.14%	13.27%	22.88%
<i>IS – Linear SVM</i>	8.82%	8.18%	7.99%	9.95%
<i>IS Stat. – Linear SVM</i>	9.29%	8.30%	8.20%	10.28%
<i>IS – Coarse Gaussian SVM</i>	8.82%	8.18%	7.99%	9.95%
<i>IS Stat. – Coarse Gaussian SVM</i>	8.68%	7.99%	<u>7.96%</u>	<u>9.86%</u>
<i>IS – Medium Gaussian SVM</i>	8.24%	8.22%	8.04%	10.80%
<i>IS Stat. – Medium SVM</i>	<u>7.98%</u>	7.87%	8.08%	11.10%

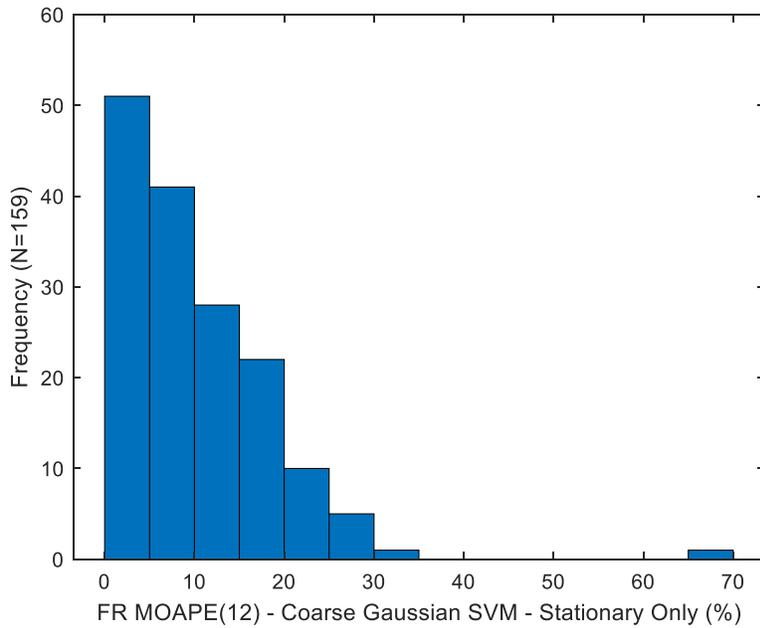


Figure 6-4: Distribution of Fill Ratio MOAPE(12) for the Best-Case Estimator

6.4 Individual-Specific Fill Ratio Prediction Results

Single factor subject-specific linear regression models were also investigated for fill ratio estimation. A FR MAPE of 8.04% was achieved for a subject-specific inclination LR model, corresponding to a 1.10% absolute reduction from models trained a LOTO framework. This reduction is minimal relative to the error reductions which were observed for volume subject-specific versus LOTO models. Variation in fill ratio MAPE is shown in Figure 6-5 for this subject-specific inclination model. A standard deviation of 4.52% is observed for this subject-specific fill ratio MAPE, versus 3.41% for LOTO models. This observation is again contrasted from the volume case, where dispersion for subject-specific models was drastically reduced. Variation in MAPE across trials for both training techniques is shown in Figure 6-6.

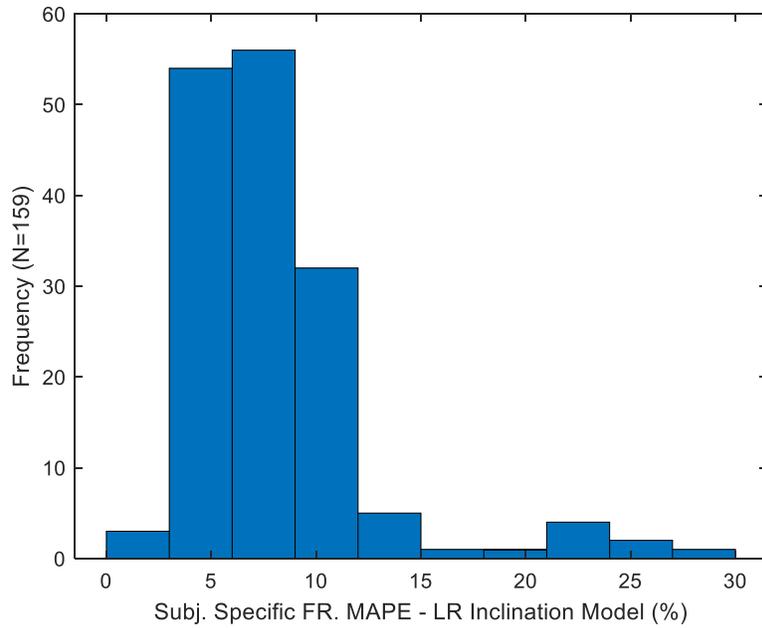


Figure 6-5: Distribution of Fill Ratio MAPE for a Subject-Specific LR Inclination Model

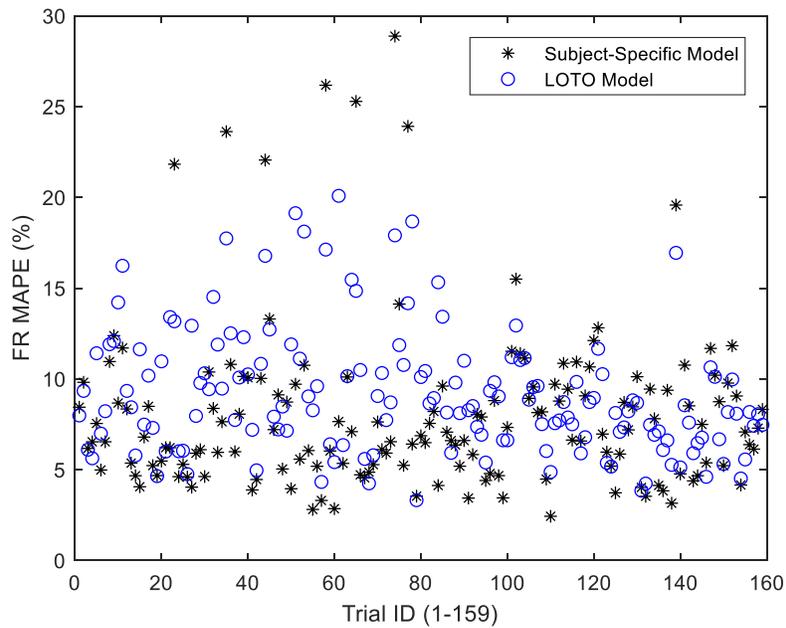


Figure 6-6: Variation in Inclination-Based FR APE

A fill ratio MAPE of 14.21% was achieved for a subject-specific integration-based linear regression model. While slightly reduced from the LOTO case (15.87%), the lack of substantial difference between the two training techniques is also distinguished from volume models. Variation in APE across trials is shown in Figure 6-7. When compared to the subject-specific APE distribution for inclination models, the presence of large additional outliers is noticeable. A comparison of MAPE achieved across trials for the two training techniques (subject-specific and LOTO) is shown in Figure 6-8. As demonstrated, models trained out-of-subject exhibit greater consistency in estimation error versus subject-specific models. We hypothesize that this improvement is associated with the substantial increase in available training data for the prior case, along with the lack of subject-specific determinants in the motion pattern for a given fill ratio.

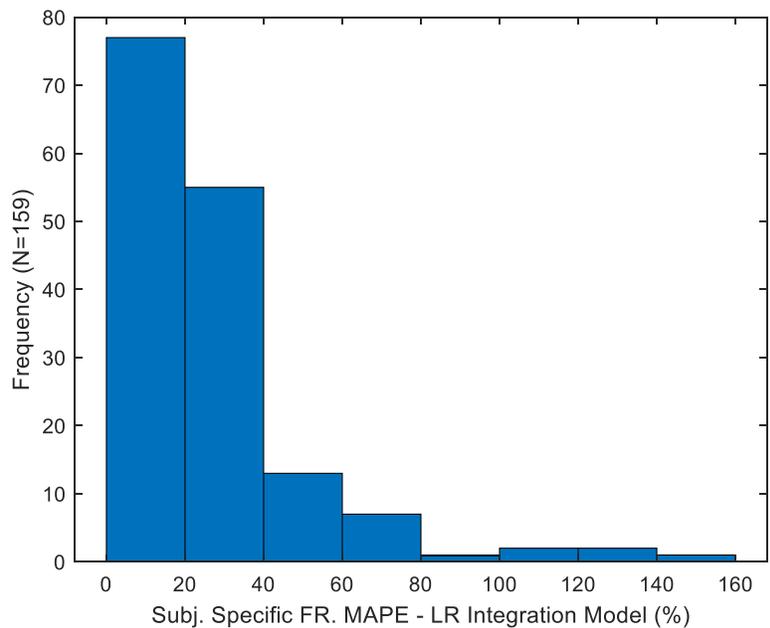


Figure 6-7: Distribution of Fill Ratio MAPE for a Subject-Specific LR Integration Model

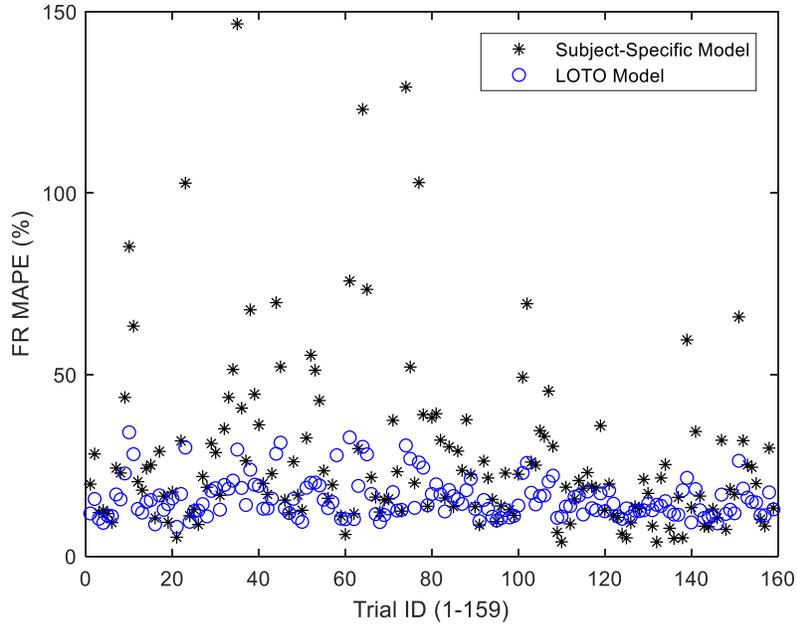


Figure 6-8: Variation in Inclination-Based FR MAPE

6.5 Discussion

A scatter plot of the best-case (coarse kernel, entire event partition, LOTO training) predicted versus ground-truth fill ratio is shown in Figure 6-9. As described in the prior sections, the accuracy of fill ratio estimates is greatly enhanced versus volume estimation. A coefficient of determination (R^2) of 77.1% is observed for the best-fit linear mapping between the best-estimate and ground-truth quantities.

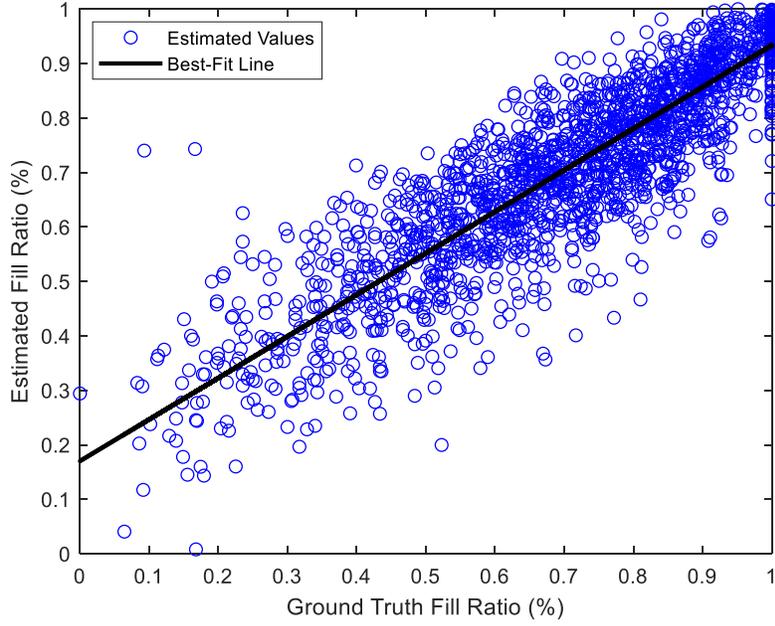


Figure 6-9: Approximated Versus Ground-Truth Fill Ratio for Best-Case Estimator

6.6 Residual Volume Prediction Results

Fill ratio estimates for pairs of drinks may be used to estimate aggregate consumption for a known container geometry as specified in (6.1), where α is a container-specific linear density parameter, $\hat{V}_{i:f}$ denotes the estimated aggregate consumption from drink i to f , and v_k and \widehat{FR}_k denote the ground truth volume and estimated fill ratio at the initiation of drink k .

$$\hat{V}_{i:f} \approx \sum_{k=i}^f v_k \approx \alpha(\widehat{FR}_i - \widehat{FR}_{f+1}) \quad (6.1)$$

This mechanism, hereby denoted as residual volume estimation, was assessed based upon the noted superior accuracy and reduced inter-subject variability of fill ratio versus volume estimators. The estimation process is depicted in Figure 6-10.

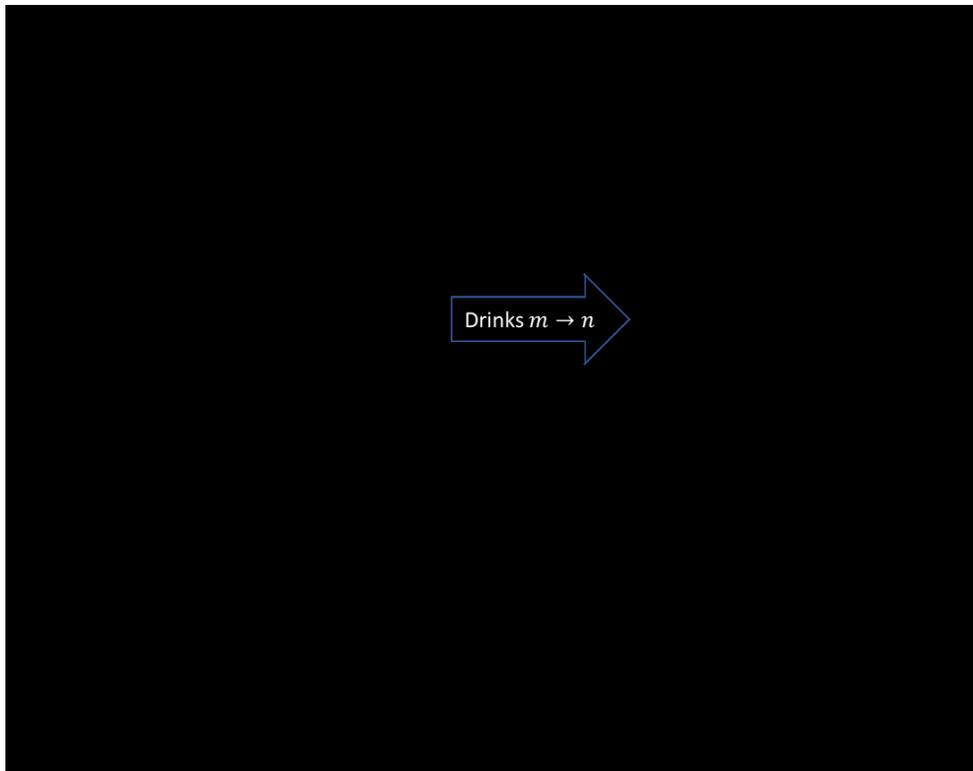


Figure 6-10: Technique for Leveraging Fill Ratio for Residual Volume Estimation

Comparison was performed using the MOAPE(11) metric. This sequence length was chosen as it represents the maximum number of drinks which can be assessed using initial fill ratio estimates for a 12-drink experimental protocol. As shown in Figure 6-11, this enhanced accuracy does not produce improved aggregate consumption estimates versus those formed through summation of drink-level volume estimates (hereby denoted as cumulative consumption estimation).

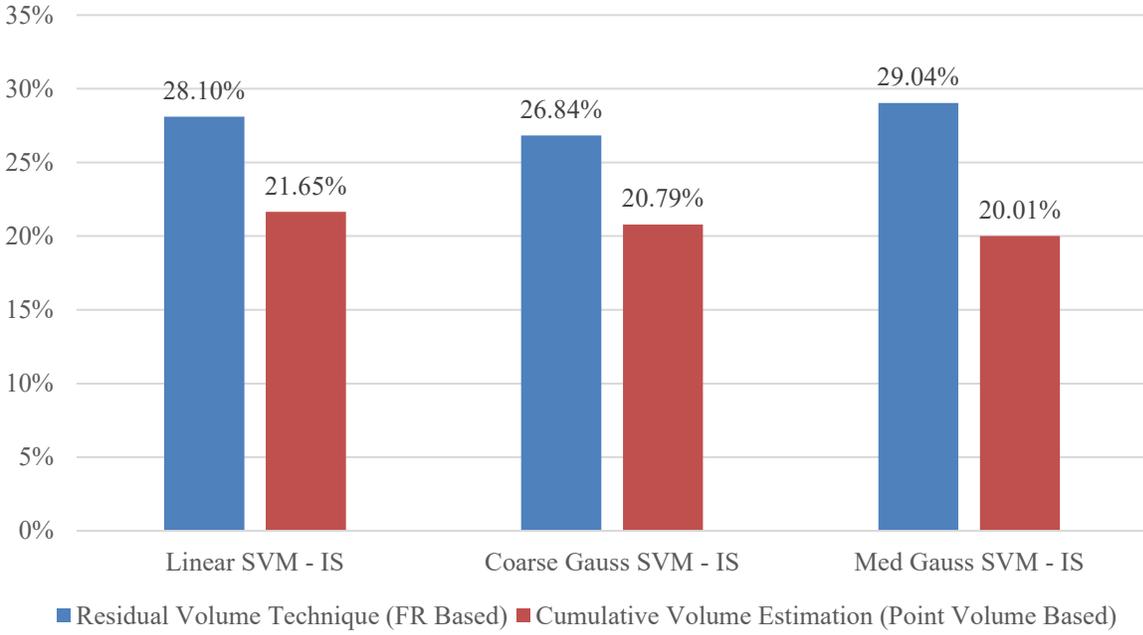


Figure 6-11: Comparison of Residual and Cumulative Techniques for Aggregate Estimation

This discrepancy may be attributed to the ability of the latter method to benefit from cancelation of sequential estimation errors within a drink sequence. Moreover, normalization effects during conversion to aggregate consumption volume (i.e.: residual volume-based OAPE) serve to distort achieved accuracy in fill ratio estimation (i.e.: fill ratio APE). This distortion is more pronounced for trials with smaller levels of aggregate consumption, as depicted in Figure 6-12 and summarized in (6.2).

$$OAPE(j) = \frac{|\widehat{FR}_{j+1} - FR_{j+1}|}{1 - FR_{j+1}} = \frac{APE_{j+1}}{1 - FR_{j+1}} \quad (6.2)$$

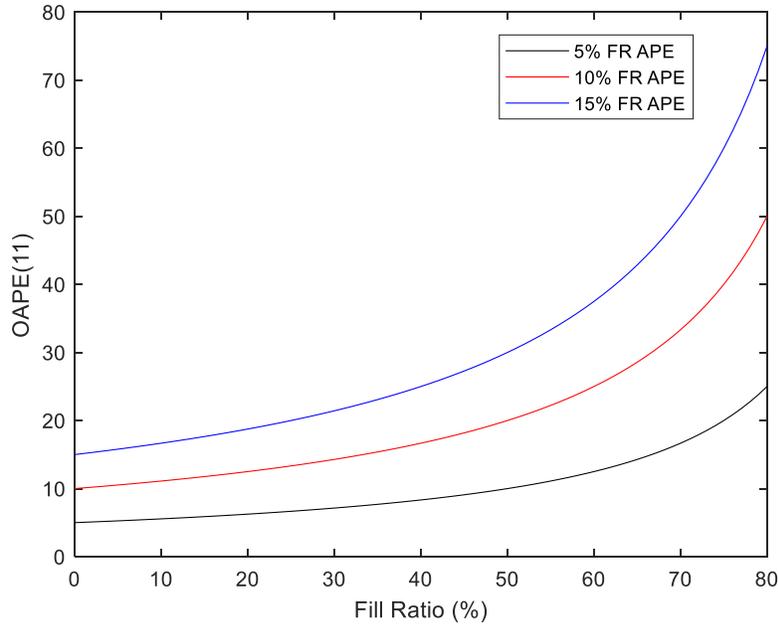


Figure 6-12: Variation in Residual Volume-Based OAPE Versus FR

6.7 Multi-Target Estimation Frameworks

As noted in Table 4-2, the relationship between sip duration and drink volume strengthens for samples with increasingly controlled fill ratio. Based upon this observation, various techniques for incorporating fill ratio information into the volume estimation process were explored. The first approach conditioned the training set using fill ratio information. Namely, training data was restricted to the 150 samples whose fill level labels were closest to the estimated fill ratio in the Euclidean sense. While the computational overhead of this approach is not feasible in practical deployment, similar techniques could be realized by instead selecting from a pretrained model library for targeted fill ratio ranges based upon estimated fill ratio.

For purposes of exploring the maximum achievable benefit using this approach, analysis was conducted using ground-truth fill ratio information in addition to estimates. Moreover, to assess the utility of explicitly mandating this form of fill ratio incorporation, a strategy of

appending the fill ratio into the feature space was also considered. Results for all four analysis combinations are presented in Figure 6-13 for the best-case macro-event volume estimator (coarse Gaussian SVM). Estimated fill ratios were obtained using the coarse Gaussian SVM regressor. As demonstrated, while ground truth fill ratio information improves estimate accuracy, no benefit is realized when noisy estimates are used. Moreover, the proposed approach of training data restriction produced only minimal error reduction versus feature space expansion. We hypothesize that this limitation is associated with the reduction in available training data using the prior method.

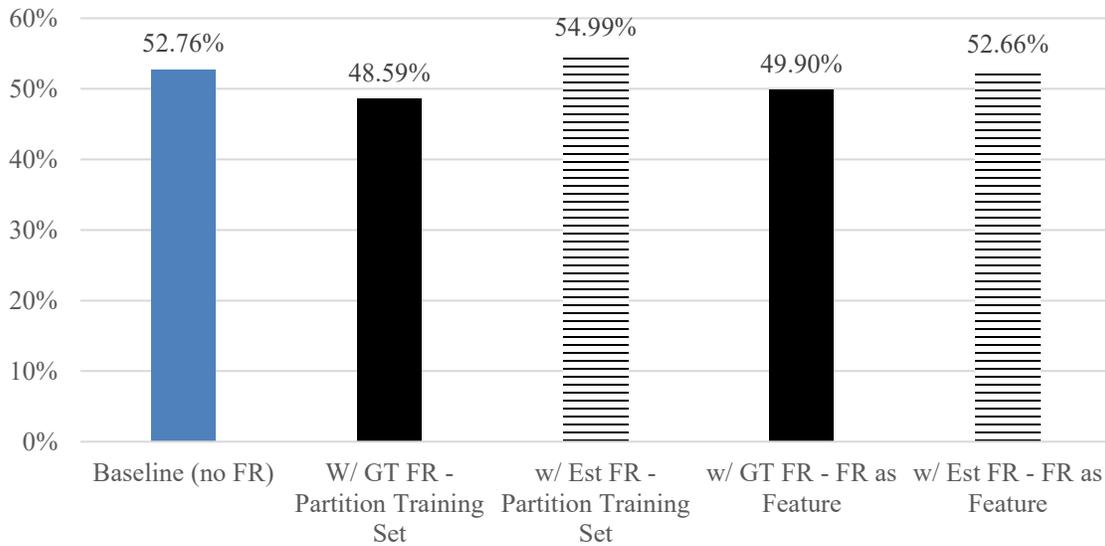


Figure 6-13: Volume Estimation Accuracy Enhancement Using Fill Ratio Information

6.8 Summary and Future Work

Support vector machine regression models for estimating fill ratio were demonstrated within this chapter. Models utilized both the newly proposed IS feature space, along with the 4-element legacy set. Estimate accuracy was improved and inter-subject variability was reduced considerably versus the volume estimators explored in Chapter 5. Models utilizing the IS feature set outperformed those employing the legacy feature set.

Contrary to the volume results presented in Chapter 5, subject-specific models did not improve fill ratio estimation accuracy. Error dispersion across trials also failed to exhibit the reduction observed in the volume case. These results indicate that fill ratio estimators exhibit less inter-subject variability, and are thus better suited for deployment without subject-specific training data.

In spite of this accuracy improvement, aggregate consumption estimates formed using computed fill ratios demonstrated reduced accuracy versus those obtained through sequential summation of volume estimates. This is attributed both to the ability of the latter approach to benefit from error cancelation across drinks, along with the described normalization effects associated with the prior approach.

In addition, a technique for utilizing fill ratio information to improve volume accuracy was presented. Namely, a strategy for conditioning the available training data distribution using fill ratio estimates was proposed. While utilization of ground-truth fill level data enhanced volume estimation accuracy, noisy estimated fill ratios did not produce an improvement.

Future work should focus on further improving the accuracy of the fill ratio estimates described herein. This approach is especially promising for the target use case, given the noted capability of these models to be trained using subject-independent data. Accuracy would also likely be improved through modification of the experimental protocol to reduce the noise of ground truth fill level labels. Namely, rather than setting the initial fill level visually, a mass reading could be used to ensure consistent initial filling across trials. It is noted that this approach would likely increase data collection time due to difficulties in filling to the requisite precision.

Chapter 7 : Improving Aggregate Consumption Accuracy Through Heuristic Fusion

7.1 Introduction

As demonstrated in Chapter 5, estimating drink volume using the characteristics of container motion is a challenging problem. This complexity is driven by the mutual influence of both fill level and volume on the resulting motion pattern. Moreover, while key kinematic parameters, such as event duration and the integral of the inclination trajectory, provide limited utility for predicting volume on an individual basis, these relationships were observed to generalize poorly across subjects.

While learned motion models offer improved fill level prediction, aggregate consumption estimates formulated using these values were less accurate than those achieved through summation of individual volume estimates as described in the prior chapter. This inferior performance may be attributed to both normalization effects, along with the ability of the latter technique to benefit from error cancelation across adjacent volume predictions. Given this observation, further improvement of fill ratio estimates is essential for achieving sufficient consumption accuracy using the residual volume approach.

One possible solution for achieving this improvement is by combining results from the learned sensor model with estimates formed using a heuristic consumption model. Under this proposed scenario, the heuristic consumption model describes the anticipated change in fill ratio over a series of drinking events. The consumption model may be designed to exploit the mandated decrement of the target variable during drinking in the absence of filling events. Moreover, knowledge of typical drink volumes may be used to reduce uncertainty in produced estimates under the assumption of known container geometry. As both the described heuristic model and

sensor estimates are characterized by some degree of uncertainty, combining both values in an attempt to improve accuracy may be viewed as a traditional sensor fusion application.

The research described within this chapter proposes a technique for implementing this proposed approach. Namely, learning-based fill ratio estimates are combined with those obtained from an empirically parameterized consumption model describing the expected drink-over-drink decrement in the target variable. Fusion is accomplished using both a complementary and Kalman filtering framework. The chapter begins with a description of the analysis methods employed. Next, the proposed fusion frameworks are introduced, followed by a discussion of the accuracy improvements achieved. Recommendations for future fusion research is provided at the conclusion of this chapter.

7.2 Methods

7.2.1 Sensor-Based Fill Ratio Estimates

The strategy for partitioning testing and training data within the current chapter is slightly modified from the previously applied LOTO approach. This modification was chosen due to the computational complexity of the brute force techniques employed for tuning filter parameters, which are described in the subsequent section. For current analysis, data was initially partitioned into an approximately 80/20% partition on a per-experiment level. The 32 testing trials were withheld for testing the proposed fusion approach, while the 127 training trials were used for both training the fill ratio regression model, along with forming parameter estimate for the proposed fusion operations. An SVM (coarse Gaussian kernel) regression model was then trained for estimating fill ratio. This model form was chosen as it exhibited the best performance amongst those for fill ratio estimation in Chapter 5. Training was performed using the default parameters as established within MATLAB's Regression Learner application.

7.2.2 Development of Fusion Models

The decrement in the bottle's fill ratio upon occurrence of a drinking event is specified as

$$FR_j = FR_{j-1} - \gamma V_{j-1} = FR_{j-1} - \Delta FR_{j-1} \quad (7.1)$$

where FR_j denotes the initial fill ratio of the j^{th} drink, FR_{j-1} and V_{j-1} denote the initial fill ratio and volume of the $(j - 1)^{th}$ drink, and γ is a geometric constant mapping volume reductions to decreases in fill ratio. For blind estimation scenarios, this quantity may be modeled stochastically as specified in (7.2)

$$FR_j = FR_{j-1} - \overline{\Delta F} + w_j \quad (7.2)$$

where $\overline{\Delta F}$ is a constant corresponding to the expected decrement in fill ratio associated with a typical drink, and w_j is a random variable reflecting variation about this assumption. For subsequent discussion, this heuristic consumption model is denoted as the *decrement model*.

As noted in the prior chapter, fill ratio may be approximated with reasonable accuracy using learned model based upon motion pattern of the sensor during drinking. This relationship may also be modeled stochastically as

$$Y_j = FR_j + v_j \quad (7.3)$$

where Y_j denotes the estimated fill ratio for the j^{th} drink using the machine learning model, and v_j is a random variable denoting the uncertainty of the estimate. For subsequent discussions, this relationship is denoted as the *measurement model*.

For the research conducted herein, w_j and v_j are modeled as independently distributed white noise Gaussian processes characterized by $w_j \sim \mathcal{N}(0, \sigma_w^2)$ and $v_j \sim \mathcal{N}(0, \sigma_v^2)$. While the physical validity of this assumption is clearly limited (i.e.: the probability of an increase in fill ratio upon occurrence of a drinking event is non-zero as specified in (6.1)), this format was chosen to

yield closed-form tractable expressions for the optimal linear estimator using a Kalman filtering framework as described below. As noted in the subsequent section, σ_w^2 was selected such that the likelihood of infeasible decrement model predictions is minimized.

Under these aforementioned assumptions, the fill ratio may be iteratively estimated by combining information from the decrement and measurement models as follows. First, the decrement model is used to obtain an a priori (i.e.: not conditioned on the current sensor prediction) fill ratio estimate. This value is determined according to the decrement model by reducing the posterior (i.e.: formulated after obtaining the sensor prediction) estimate from the prior drink by the average decrement as shown in (7.4).

$$\widehat{FR}^-_j = \widehat{FR}^-_{j-1} - \overline{\Delta F} \quad (7.4)$$

The above step is often denoted as the *prediction* stage within the sensor fusion literature. This estimate is then used to anticipate the value produced by the measurement model, which is denoted as \widehat{Y}_j . For the assumed measurement model, this is equivalent to the a priori fill ratio estimate. Upon obtaining the actual sensor estimate, the innovation is computed as specified in (7.5)

$$\Delta = Y_j - \widehat{Y}_j = Y_j - \widehat{FR}^-_j \quad (7.5)$$

This residual value may then be used to modify the a prior fill ratio estimate as shown in (7.6)

$$\widehat{FR}_j = \widehat{FR}^-_j + k\Delta \quad (7.6)$$

where k is a blending parameter specifying how new information contained within the innovation should be incorporated within the posterior estimate. Substituting (7.5) into (.6) yields the following formula expressing the posterior estimate as a convex combination of the measurement output and a priori measurement.

$$\widehat{FR}_j = kY_j + (1 - k)\widehat{FR}_j^- \quad (7.7)$$

Fusion strategies using a constant blending parameter over the entire drink sequence ($k_j = k \forall j$) are often described as complimentary filtering (CF) within the sensor fusion literature. This approach is often employed in IMU data fusion to combine information from multiple sensor modalities, and is explored for improving inclination estimates in the subsequent chapter [51].

Using a Kalman filtering framework, the optimal linear estimate of the fill ratio may be determined by minimizing the mean squared posterior error estimate through appropriate adjustment of the blending parameter k . The a priori and posterior estimate errors are defined in (7.8) and (7.9), respectively

$$e_j^- = FR_j - \widehat{FR}_j^- \quad (7.8)$$

$$e_j = FR_j - \widehat{FR}_j \quad (7.9)$$

Each of the above error quantities is a random variable characterized by a mean squared error as specified in (7.10) and (7.11), respectively

$$p_j^- = E[(e_j^-)^2] \quad (7.10)$$

$$p_j = E[e_j^2] \quad (7.11)$$

where $E[\cdot]$ denotes the expectation operator. While details of the derivation are omitted herein, the optimal value of k_j , hereby denoted as the Kalman gain, may be determined as specified in (7.12)

$$k = \frac{p_j^-}{p_j^- + \sigma_v^2} \quad (7.12)$$

The above derivations result in the following simplified recursive solution for estimating fill ratio and using the decrement and measurement model

Predict Stage

$$\widehat{FR}^-_j = \widehat{FR}^-_{j-1} - \overline{\Delta F}$$

$$p_j^- = p_{j-1} + \sigma_w^2$$

Update Stage

$$k_j = \frac{p_j^-}{p_j^- + \sigma_v^2}$$

$$\widehat{FR}_j = k_j Y_j + (1 - k_j) \widehat{FR}_j^-$$

$$p_j = p_j^- (1 - k_j)$$

Both the complementary and Kalman filtering framework were used to fuse decrement and measurement models within the subsequent analysis. Techniques for establishing model parameters are described in the following subsection.

7.2.3 Establishment of Model Parameters

For complementary filtering, the static blending parameter was set as value which minimizing root mean square (RMS) computed using the training set. This parameter was estimated using a grid search over all possible blending parameter values. Namely, k was swept through the allowable range of [0,1] at a resolution of 0.01. The RMS error of the estimate was recorded for each value of k considered. A minimum value of 8.70% was obtained for $k = 0.6869$, which is used in all subsequent analysis on the test set. Variation in test-set RMSE versus the various blending values considered is depicted in Figure 7-1.

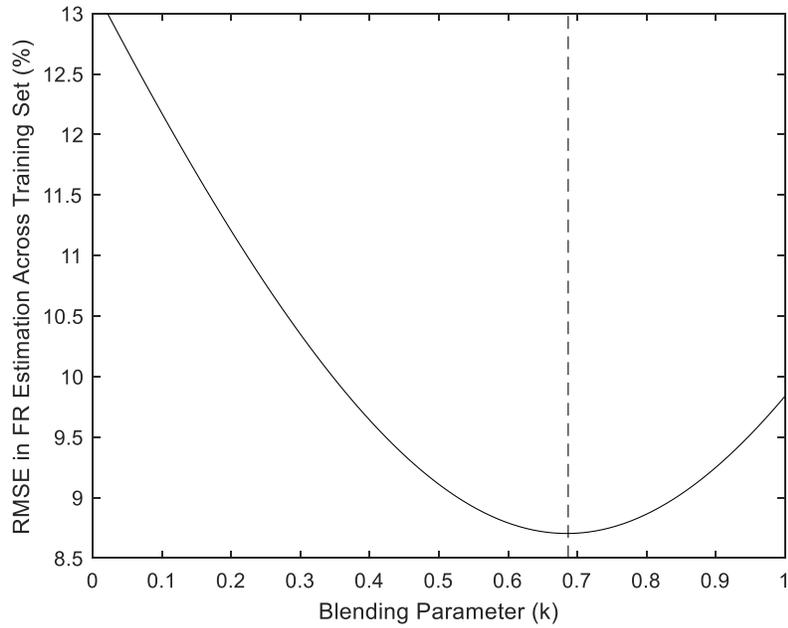


Figure 7-1: Variation in Test-Set RMSE for Complementary Filtering Approach

For the current analysis, FR_0 and p_0 were initialized to 1.0528 and 0, respectively. Model variances were obtained by setting σ_v^2 to the square of the RMS error observed during model training (0.0097), with σ_w^2 defined parametrically as $\sigma_w^2 = r\sigma_v^2$. The value of r was tuned to minimize training set RMSE using a similar brute-force approach to that used to determine the static blending parameter, yielding an optimal value of 7.67% for $\sigma_w^2 = 0.0017$. Variation in training-set RMSE versus r is depicted in Figure 7-2.

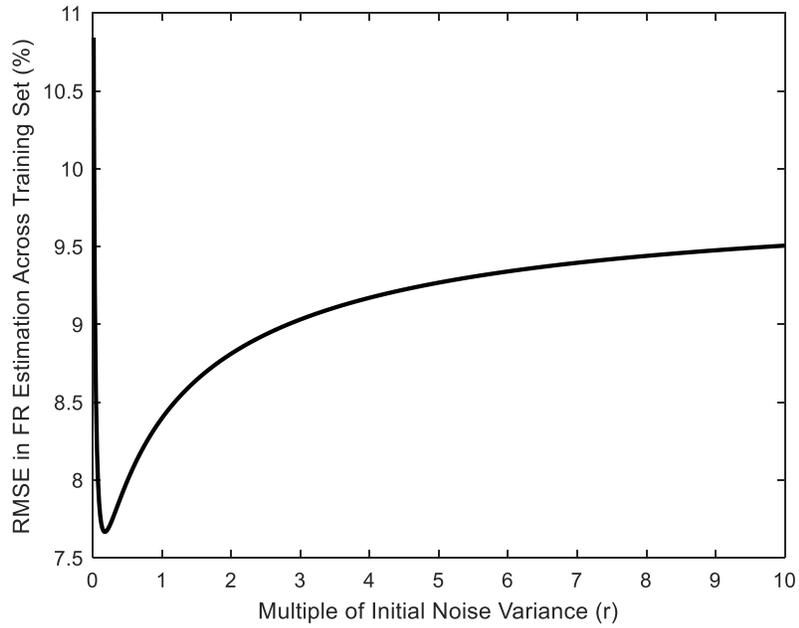


Figure 7-2: Variation in Training RMSE Versus Noise Multiple

An example of the predictions provided by each technique for a single test set experiment is presented in Figure 7-3.

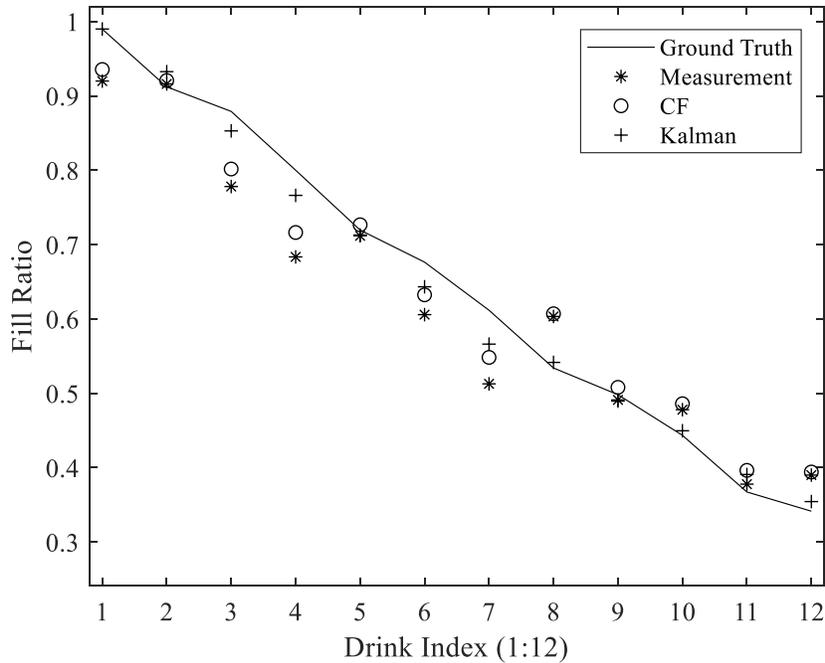


Figure 7-3: Example Outputs of Prediction Techniques

7.3 Results

Test set RMSE for each technique is reported in Table 7-1. As noted, while both heuristic fusion techniques improve estimation accuracy, the adaptive weighting (i.e.: variable across drinks) provided by the Kalman approach outperforms the static blending of the CF technique.

Table 7-1: Test Set Fill Ratio RMSE

Estimation Approach	Test Set RMSE (%)	Relative % Decrease Via Fusion
Sensor Estimate Only	9.29%	-
CF Fusion	7.33%	17.31%
Kalman Fusion	5.74%	33.67%

Variation in estimation error across experiments was considerable. The maximum and minimum RMS errors for each technique are shown in Table 7-2. Variation in RMSE for all three techniques across trials is depicted in Figure 7-4.

Table 7-2: Range of Fill Ratio RMSE Across Test Set

Estimation Approach	Min(RMSE) (%)	Max(RMSE) (%)
Sensor Estimate Only	5.43%	19.28%
CF Fusion	3.04%	15.65%
Kalman Fusion	1.32%	14.78%

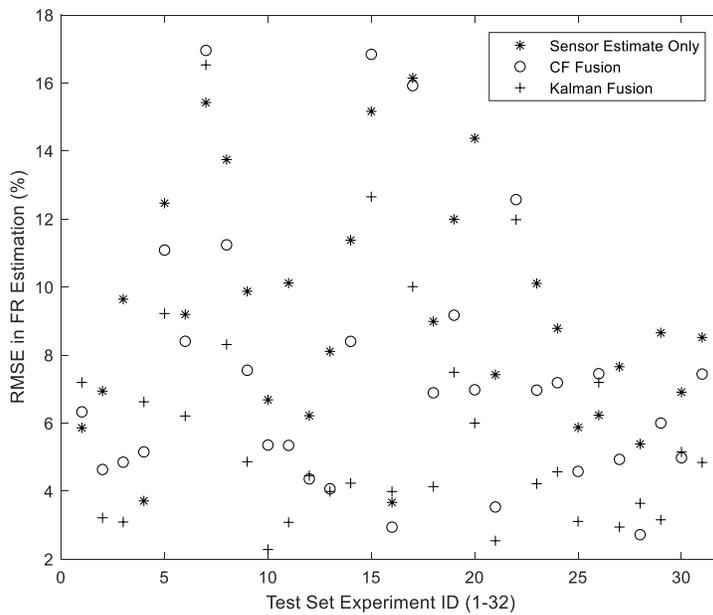


Figure 7-4: Variation in RMSE Across Trials in Test Set

To promote comparability with results from Chapter 6, error metrics were converted to those previously employed (i.e. absolute percentage error). Moreover, the residual volume estimates

corresponding to each proposed fusion technique were also computed. MAPE across trials for the three techniques considered is shown in Table 7-3. MOAPE(11) values are shown in Table 7-4.

Table 7-3: Test Set Fill Ratio MAPE

Estimation Approach	Test Set MAPE (%)	Relative % Decrease Via Fusion
Sensor Estimate Only	7.61%	-
CF Fusion	6.01%	21.02%
Kalman Fusion	4.82%	36.67%

Table 7-4: Test Set Volume MOAPE(11)

Estimation Approach	Test Set MOAPE(11) (%)	Relative % Decrease Via Fusion
Sensor Estimate Only	35.00%	-
CF Fusion	23.92%	31.67%
Kalman Fusion	15.73%	55.06%

For purposes of comparison, aggregate consumption for the test set was also computed using the cumulative technique described in the prior chapter, Namely, per-drink volume estimates were formed using the best-case SVM estimator from presented in Chapter 4, with aggregate consumption estimated through summing individual drink estimates. MOAPE(11) achieved using this approach was 20.72%. Comparative results of MOAPE(11) achieved using both the cumulative and best-case residual volume (i.e.: using Kalman filtering for fill ratio estimation) approaches are shown in Figure 7-5 across each trial in the test set.

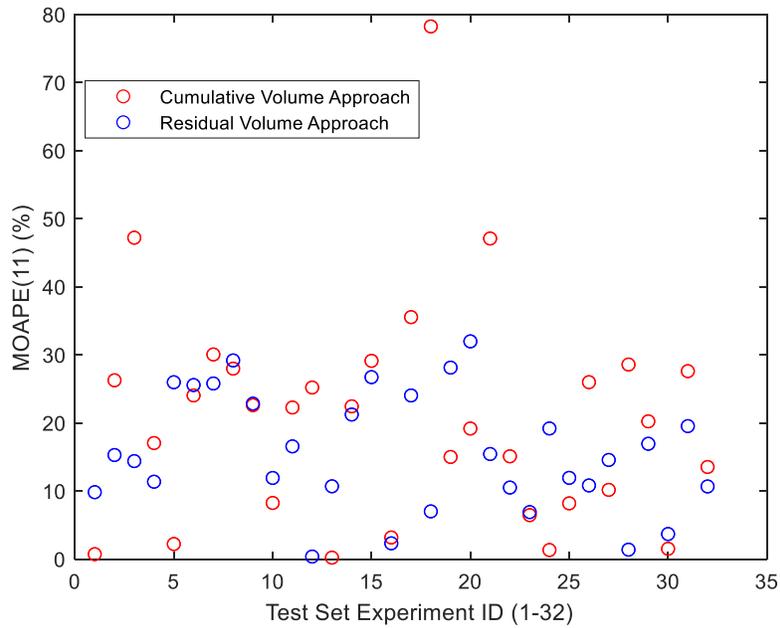


Figure 7-5: Variation in Volume MOAPE(11) Across Trials in Test Set

7.4 Summary and Future Work

The heuristic fusion approach proposed herein was demonstrated to improve the accuracy of fill ratio estimates. The dynamic blending of the Kalman framework provided superior performance compared to the static approach of the CF. Dynamic blending allows the fused model to adjust for scenarios in which the sensor-based model does not generalize well to the specific individual (i.e.: the large maximum RMS error for sensor estimates denoted in Table 7-2), as well as for cases where individual consumption varies considerably from model assumptions. Future work should assess the sensitivity of the proposed methods to variation in model parameters, explore alternative dynamic fusion strategies, and investigate utilization of volume-based predictors within the estimation framework.

Chapter 8 : Verification of Inclination Estimates Using Video Motion Capture

8.1 Introduction

This chapter describes a simplified technique for estimating the inclination trajectory of the bottle by fusing accelerometer and gyroscope data. The proposed approach isolates pertinent information in the gyroscope channels using an accelerometer-based orientation estimate previously introduced in Chapter 4. Verification of estimate quality is conducted using motion capture results obtained using Blender, an open-source computer graphics program.

The chapter begins by describing the experimental set-up and protocol utilized. The proposed techniques for estimating inclination trajectory using the IMU output are then described. Next, limited details regarding application of the motion capture software are provided. Multiple trajectory estimates developed from the IMU data are compared with video-based estimates according to a root mean squared (RMS) discrepancy metric. Conclusions and suggestions for future research are provided at the end of the chapter.

8.2 Methods

8.2.1 Data Collection

The experimental protocol consisted of consuming ten drinks of water from a refillable bottle, with activity captured by both an attachable IMU sensor and video. This identical script was completed by five participants, resulting in 50 drink events. The IMU sensor was attached to the bottle by an elastic band at a controlled position and orientation as depicted in Figure 8-1.

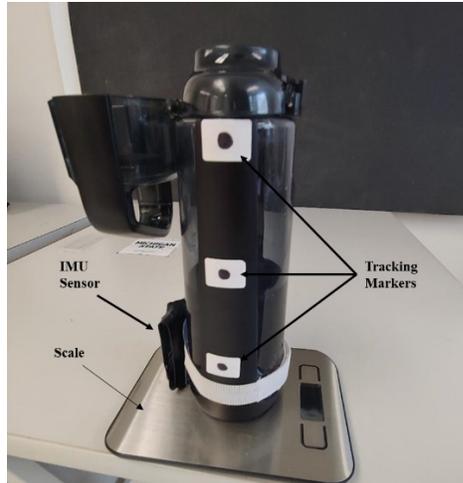


Figure 8-1: Sensor and Marker Configuration

The camera was positioned approximately 5 feet from the table, with zoom adjusted to focus the field of view on the region encompassing the expected bottle trajectory. Three markers were placed on the bottle to facilitate video tracking. Various parameters of the set-up, such as scene background and marker geometry, were determined empirically through multiple iterations before initiating the experiment. The attachable IMU sensor and supporting data collection system were described in Chapter 3.

Three estimates of the inclination trajectory were formed using the IMU outputs. Signals were preprocessed using the procedure described in Chapter 3. Drinks were parsed from the IMU data using the algorithm described in Chapter 4 for the large-scale data collection. A similar algorithm exploiting the stationary placement of the bottle between drinks was used to parse video data.

Accelerometer-based inclination estimates were computed using the technique previously introduced in Chapter 3. Under the assumption of negligible dynamic acceleration, this approach decomposes the static acceleration due to gravity in a global coordinate frame as described in (3.1).

To incorporate the gyroscope output, it is necessary to specify the axis about which subsequent rotations modify the inclination angle. This is accomplished by estimating the orientation of the resultant acceleration vector in the cross-sectional plane of the bottle as previously described in (4.1). Perturbations to the inclination occur through rotations about an axis which is perpendicular to this orientation angle α . To compute rotation about this axis using the gyroscope, sensor outputs in the local coordinate frame are projected onto the axis of rotation as described in (8.1)

$$g_{cs} = \mathbf{P}(\mathbf{g}_{yz'}) = -g_{y'} \sin \alpha + g_{z'} \cos \alpha \quad (8.1)$$

where $\mathbf{g}_{yz'}$ denotes the vector output of the gyroscope in the yz plane of the sensor, and g_{cs} denotes the projection of this output along the hypothesized axis of rotation.

The gyroscope component along the hypothesized axis of rotation is utilized to develop an estimate of the inclination trajectory through integration as specified in (8.2).

$$\widehat{\theta}_G = \widehat{\theta}_G(0) + g_r \Delta t \quad (8.2)$$

where $\widehat{\theta}_G(0)$ denotes the initial condition imposed on the inclination estimate (defined as 0 degrees at the drink parsing initiation), and Δt denotes the sampling period of 50 ms.

In addition to the individual estimates specified above, preliminary investigation was conducted exploring various fusion approaches which exploit the unique advantages of each sensing modality. Amongst simplistic fusion approaches, the complimentary filter (CF) estimates the output as a linear combination of the accelerometer and gyroscope estimates (8.3).

$$\widehat{\theta}_{CF} = h(\widehat{\theta}_G) + l(\widehat{\theta}_A) \quad (8.3)$$

where h and l are constants satisfying $h + l = 1$. To avoid errors associated with gyroscope drift during the translation portion of the drinking events, CF-based estimates only perform fusion

during the estimated stationary interval of α (i.e.: the CF output is equal to the accelerometer estimate outside of this interval) .

8.2.2 Video Inclination Tracking

The motion tracking functionality of Blender, an open-source 3-D computer graphics program, was used for estimating the inclination angle of the bottle [52]. In the processing workflow, markers are identified through selection in the graphical user-interface, and then tracked using a SIFT feature-based approach. Figure 8-2 depicts the output of the tracking process, demonstrating the estimated marker trajectory in blue.

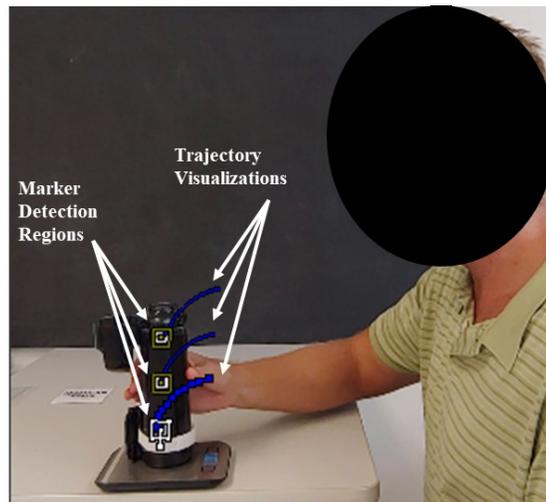


Figure 8-2: Visualization of Blender Tracking Output

The software produces estimated pixel locations, reported as ordered doubles, for each of the markers. These values were mapped to inclination estimates on a pairwise basis using trigonometry. The resulting inclination estimates were then parsed using an algorithm identical in concept to that described for the IMU data. Visualizations of the parsing process are depicted in Figures 8-3 and 8-4.

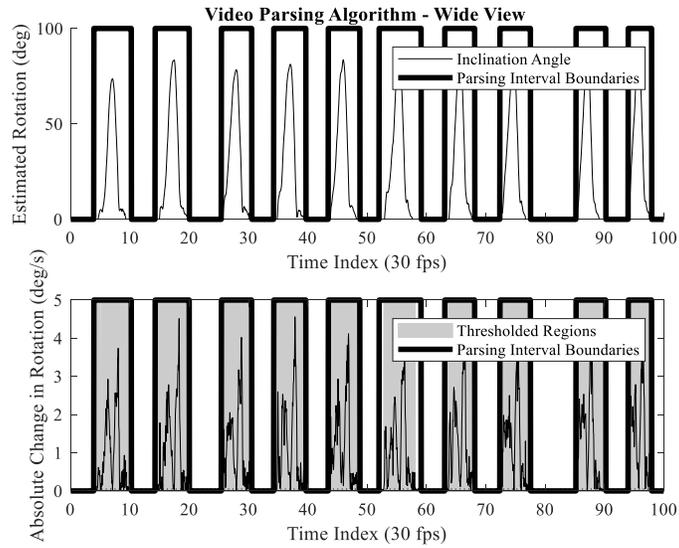


Figure 8-3: Video Parsing Process – Wide View

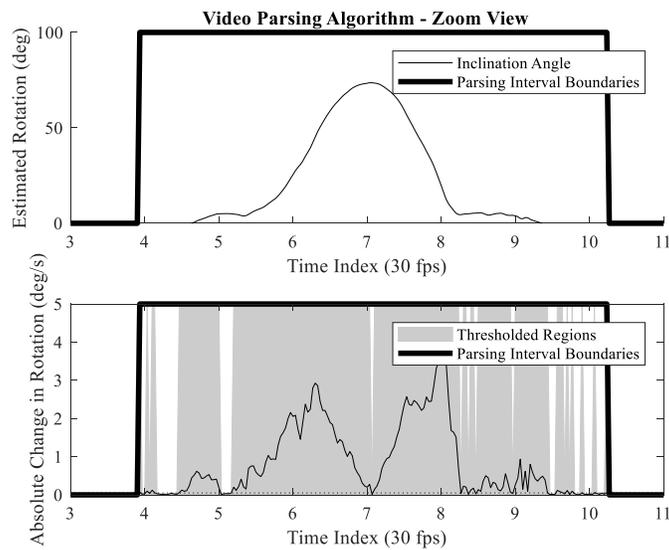


Figure 8-4: Video Parsing Process – Zoom View

As the three resulting video estimates were shown to exhibit strong correlation, all subsequent discussion references an average signal value denoted as θ_V , which was down-sampled to 20 Hz for ease of comparison with IMU data.

8.2.3 Drink Event Synchronization

To finalize the analysis framework, parsed drink events for each modality were synchronized. This was achieved by computing the cross-correlation of $\widehat{\theta}_A$ and θ_V as specified in (8.4), and subsequently shifting the signals by the maximizing lag value.

$$R[m] = \begin{cases} \sum_{n=0}^{N-m-1} \widehat{\theta}_A[n+m]\theta_V[n], & m \geq 0 \\ R[-m], & m < 0 \end{cases} \quad (8.4)$$

where N is a common duration in samples of each drink event, achieved through a-priori zero-padding as necessary. A visualization of the synchronization process is provided in Figure 8-5.

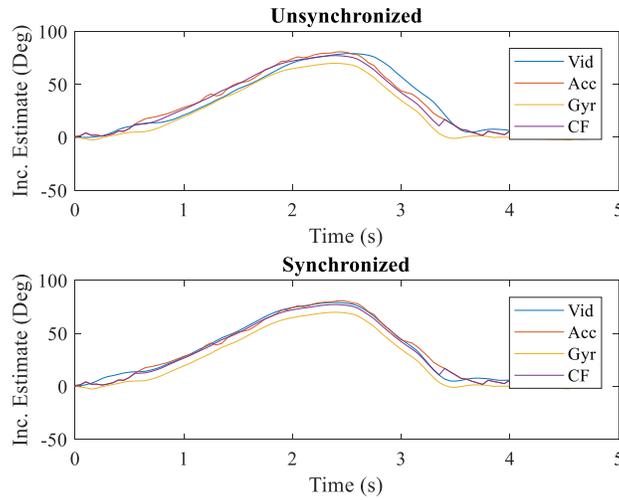


Figure 8-5: Visualization of Synchronization Process

8.3 Results

A comparison metric indicating the discrepancy between the various IMU-based trajectory estimates and the reference video estimate is defined in the RMS sense in (8.5)

$$RMSE_j = \sqrt{\frac{\sum_1^{N'} (\widehat{\theta}_j - \theta_V)^2}{N'}} \quad (8.5)$$

where N' denotes the common duration of each drink event, and $j \in \{1,2,3\}$ denotes the IMU estimation modality. Initial discrepancy estimates were computed over the entire drink event, which were synchronized according to the technique described in the previous section. A brute-force sensitivity analysis was conducted examining variability in $RMSE$ for all possible combinations of CF parameter values at a resolution of 0.001. Results are plotted versus the gyroscope weighting parameter in Figure 8-6. As noted, the error curve is convex with respect to the mixing parameter, exhibiting a minimum value of 3.85 degrees for $h = 0.425$ and $l = 0.575$.

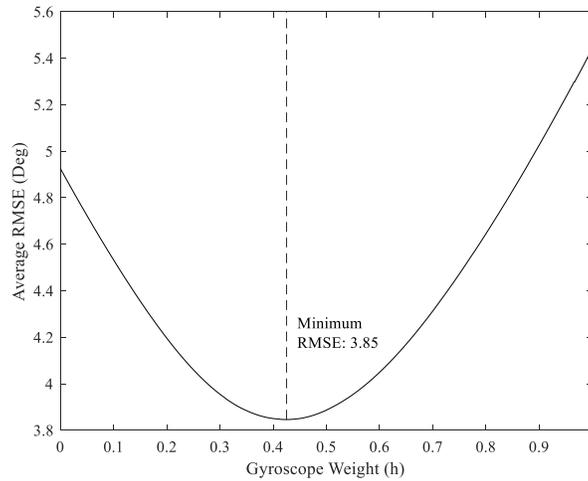


Figure 8-6: Variability in Discrepancy Metric for Varying Complimentary Filter Weights

The discrepancy metric was then computed for each of the estimation modalities, with the resulting RMS distributions depicted in Figure 8-7. The CF estimate produces the least discrepancy in the average sense, followed by the accelerometer-based estimate. The gyroscope estimate exhibits the most discrepancy, largely associated with preliminary drift error occurring during the initial lifting phase.

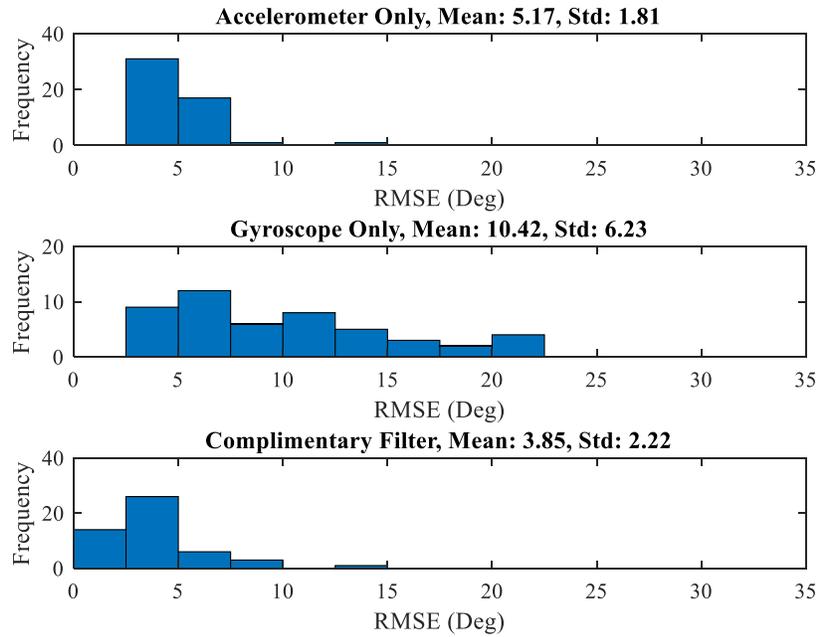


Figure 8-7: Distribution of Discrepancy Metric for Various IMU-Based Estimations

8.4 Conclusions and Future Work

The research described in this chapter proposes and verifies a simplistic approach to estimate bottle inclination by fusing accelerometer and gyroscope outputs. Results are verified through comparison with estimates produced through video-based motion capture. Employing a simplistic fusion scheme using a complimentary filter, the resulting estimation was improved by over 25% versus estimates developed solely from the accelerometer. Future research should explore alternative fusion-based approaches, along with techniques for computing drink volume using the estimated inclination trajectory.

Chapter 9 : Feature Set Expansion Using Additional Sensor Channels

9.1 Introduction

This chapter investigates various approaches for improving both volume and fill ratio estimates using information available from additional IMU channels. Namely, characteristics of the container's motion are estimated from the gyroscope sensor using the decomposition strategy introduced in Chapter 8. In addition, motion features are computed using the magnitude of the accelerometer output. This addition is intended to address limitations associated with the assumption of negligible dynamic acceleration in computing the inclination estimate. These developments produce an enriched feature set for describing the motion pattern of the container.

The performance of this feature set is assessed against all previously considered feature sets within this chapter. Comparisons are performed for both volume and fill ratio models. In addition, models utilizing the various inclination estimates introduced in the prior chapter are developed herein.

The chapter begins by formally defining each supplementary motion feature. Similar to Chapter 4, the relationship between the proposed features and labels of interest is quantified using the Pearson correlation coefficient. Next, results for volume and fill ratio estimation using the entire supplemented feature set are presented. In addition, the performance of models utilizing the fusion-based inclination estimates proposed in Chapter 8 are also presented. The chapter concludes with a summary and recommendations for future research.

9.2 Proposed Supplements to the IS Feature Set

9.2.1 Additions from Accelerometer Channels

The previously described technique for estimating container inclination in (3.1) assumes that dynamic acceleration is negligible. As the drink motion involves translation, the validity of this assumption is limited, especially for the transport portion of the motion. Based upon this observation, it was hypothesized that estimation performance may be improved by extracting information from the accelerometer channels directly.

To describe the intensity of acceleration, five features describing the morphology of the resultant accelerometer output were computed. These features are listed in Table 9-1, with correlations for the target label values also presented. Features were computed using both the entire event duration and the micro-partitioning strategy suggested in Chapter 4. For purposes of visualization, an example of variation in the accelerometer magnitude during the drinking event is depicted in Figure 9-1 for four randomly chosen drink events. Estimated inclination is also shown in this figure for purposes of comparison. As shown, while the acceleration magnitude oscillates near the assumed static value (one) during the middle of the drinking event, considerable variation is observed, especially during the transport phase.

As detailed in Table 9-1, the proposed summary features of the acceleration magnitude exhibit a stronger relationship with fill ratio versus volume labels. This is consistent with the results presented in Chapter 4, where elements of the IS feature set exhibited similar behavior.

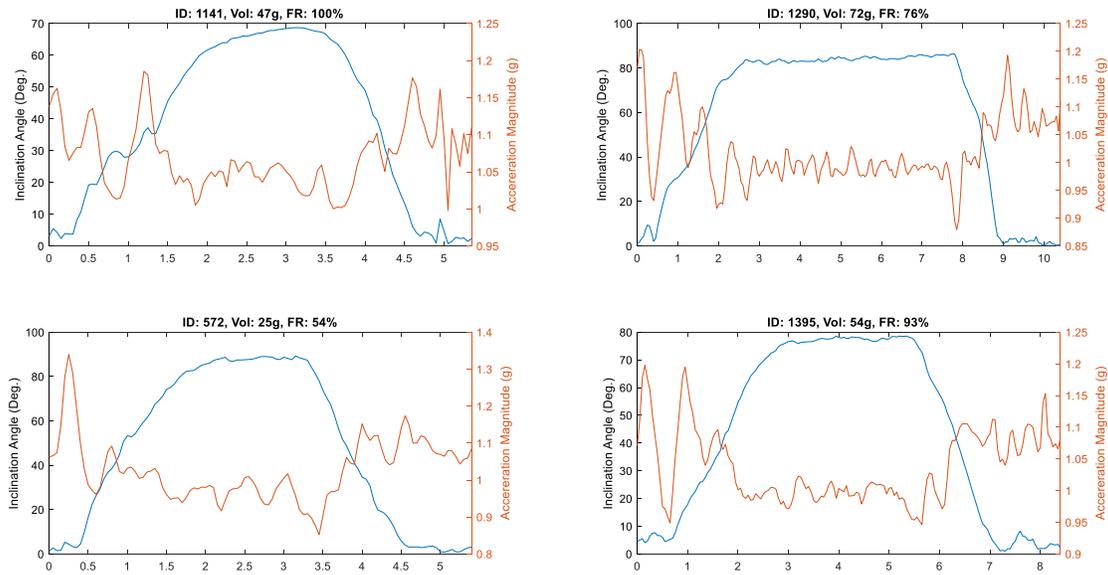


Figure 9-1: Variation in Acceleration Magnitude During Drinking Events

Table 9-1: Supplemental Features from Resultant Acceleration

Feature Definition	ρ_{vol} (Whole)	ρ_{vol} (Lift)	ρ_{vol} (Stat.)	ρ_{vol} (Place)	ρ_{FR} (Whole)	ρ_{FR} (Lift)	ρ_{FR} (Stat.)	ρ_{FR} (Place)
$max(\mathbf{a}^j)$	0.065	0.070	0.024	0.041	-0.273	-0.307	-0.114	-0.126
$min(\mathbf{a}^j)$	-0.116	-0.030	-0.129	-0.064	0.305	-0.121	0.342	0.029
$mean(\mathbf{a}^j)$	-0.239	0.027	-0.214	0.037	0.325	-0.387	0.432	-0.179
$std(\mathbf{a}^j)$	0.057	0.023	-0.030	0.038	-0.396	-0.243	-0.114	-0.127
$range(\mathbf{a}^j)$	0.098	0.067	0.087	0.062	-0.343	-0.162	-0.269	-0.118

9.2.2 Additions from Gyroscope Channels

A similar technique was used to supplement the feature set with information from the gyroscope sensor. The decomposition proposed in the previous chapter was employed, with the gyroscope output represented in terms of two components – 1) the resultant component along the axis of rotation in the bottle’s cross-sectional plane (i.e.: g_{cs}), and 2) the component parallel to the

vertical axis of the bottle (i.e.: g_x). Variation in the two quantities is depicted in Figure 9-2 for four randomly chosen drink events.

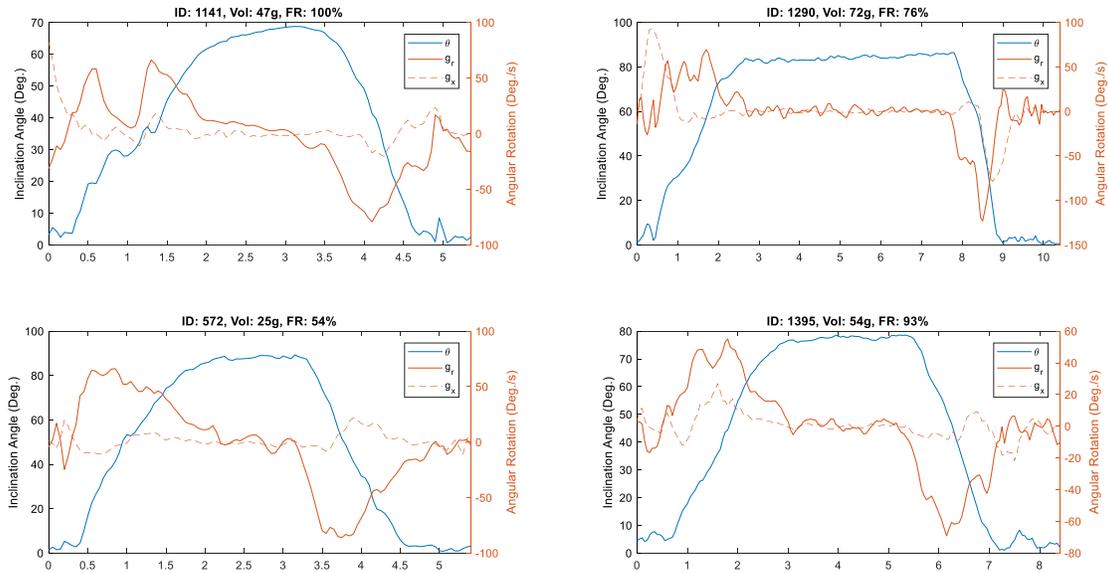


Figure 9-2: Variation in Gyroscope Signals During Drinking Events

Correlation coefficients for the newly introduced gyroscope features are summarized in Tables 9-2 and 9-3. Similar to all prior motion features evaluated, these values exhibited stronger correlation to fill ratio versus volume labels. Moreover, correlations with fill ratio were stronger for features computed using the resultant gyroscope component along the axis of rotation. These correlations were largely negative, indicating that the rate of inclination was decreased when the bottle mass was increased.

Table 9-2: Supplemental Features from Coplanar Gyroscope Resultant

Feature Definition	ρ_{vol} (Whole)	ρ_{vol} (Lift)	ρ_{vol} (Stat.)	ρ_{vol} (Place)	ρ_{FR} (Whole)	ρ_{FR} (Lift)	ρ_{FR} (Stat.)	ρ_{FR} (Place)
$max(g_{cp}^j)$	0.099	0.092	0.102	0.123	-0.430	-0.484	-0.443	-0.328
$min(g_{cp}^j)$	-0.159	-0.120	-0.214	-0.054	0.022	-0.150	0.039	0.062
$mean(g_{cp}^j)$	-0.179	0.044	-0.229	0.050	-0.396	-0.476	-0.279	-0.206
$std(g_{cp}^j)$	0.088	0.072	0.124	0.111	-0.599	-0.482	-0.541	-0.309
$range(g_{cp}^j)$	0.103	0.099	0.113	0.126	-0.430	-0.434	-0.445	-0.331

Table 9-3: Supplemental Features from Axial Gyroscope Component

Feature Definition	ρ_{vol} (Whole)	ρ_{vol} (Lift)	ρ_{vol} (Stat.)	ρ_{vol} (Place)	ρ_{FR} (Whole)	ρ_{FR} (Lift)	ρ_{FR} (Stat.)	ρ_{FR} (Place)
$max(g_x^j)$	0.101	0.051	0.047	0.091	-0.103	-0.020	-0.101	-0.097
$min(g_x^j)$	-0.179	0.044	-0.229	0.050	0.154	0.061	0.142	0.066
$mean(g_x^j)$	0.047	-0.018	0.058	0.045	0.063	0.006	0.095	-0.018
$std(g_x^j)$	0.035	0.112	-0.076	0.055	-0.104	-0.148	-0.078	-0.132
$range(g_x^j)$	0.101	0.136	0.034	0.078	-0.138	-0.075	-0.140	-0.138

In addition to the above quantities, the IS feature set was also with the various micro-event durations introduced in Chapter 4.

9.3 Effect of Feature Set Supplementation on Performance

An SVM regression model was trained to estimate volume using the LOTO technique described in Chapter 5. A medium Gaussian kernel function was employed for purposes of comparability with the previously demonstrated best-case model. A volume MAPE of 56.60% was achieved for the expanded feature set. This value is worsened from the 52.39% MAPE achieved for an identical model employed using the IS feature set.

Similarly, an SVM regression model was trained to estimate fill ratio using an identical framework to that described in Chapter 6. A coarse Gaussian kernel function was used to promote comparability with the best-case model. A fill ratio MAPE of 7.71% was achieved using the expanded feature set. This result is slightly improved versus the best-case MAPE of 7.96% achieved using the IS feature set.

9.4 Effect of Inclination Estimation Technique on Performance

The best-case SVM models described in Chapters 5 and 6 were reevaluated using the various technique for estimating the container's inclination demonstrated in Chapter 8. A volume MAPE of 57.88% was achieved using an inclination estimate formed from the gyroscope sensor only. This accuracy is decreased versus results obtained using the accelerometer-based estimate. FR MAPE using the gyroscope inclination estimate was also increased to 10.77%. A volume and fill ratio MAPE of 54.06% and 8.24% were achieved using the complementary filter-based inclination estimate. Both results are inferior versus those produced using the accelerometer-based inclination estimate.

9.5 Conclusions and Future Work

Strategies for improving volume and fill ratio estimates using supplementary motion features were explored herein. Summary features of the acceleration magnitude and various gyroscope channels were proposed. Consistent with prior features, correlation values indicated a stronger linear relationship with fill ratio versus volume labels. SVM regression models were trained using an identical approach to the previously reported best-case models presented in Chapters 5 and 6. The volume regression model using the supplemented feature set exhibited worse performance compared to the IS model. Fill ratio MAPE was slightly improved using the supplemented feature

set.

In addition, the aforementioned best-case models were reevaluated using the various inclination estimation techniques described in Chapter 8. Models utilizing the gyroscope estimate exhibited reduced estimation accuracy versus those employing accelerometer-based estimates. For the static complementary filter parameters considered, estimation accuracy was also decreased for both labels.

As only the static fusion parameters developed in the prior chapter were evaluated herein, future work should explore variation in performance for alternative fusion parameters. Moreover, performance variation for more sophisticated fusion-based inclination estimation strategies should also be explored.

Chapter 10 : Assessment of Sensor Performance for Alternative Drinking Containers

10.1 Introduction

While the reconfigurable nature of the proposed tracking solution supports deployment across multiple container types, prior research has focused solely on scenarios where the device is attached to refillable bottles. This chapter addresses this limitation by exploring placement on two additional common drinking vessels - a glass and mug. An image of all containers considered within this chapter is shown in Figure 10-1.



Figure 10-1: Three Container Types Considered

For preliminary proof-of-concept, two core sensing functions are demonstrated. Namely, the ability to classify the type of container to which the sensor is attached is shown. In practice, this functionality would support the deployment of container-specific consumption models. In addition, low-resolution fill level classification is also demonstrated. For sufficient resolution, this functionality could be used for implementing the residual volume techniques introduced in Chapter 6.

The chapter begins with a description of the experimental methods employed. Feature engineering and classifier design are then discussed, followed by the presentation of results using various training strategies and learning models. The paper concludes with a discussion of findings and recommendations for future work.

10.2 Methods

Five participants took drinks from three containers at two initial fill levels during the experiment. Subjects were instructed to consume a normal volume for each drink. The container was placed stationary on an electronic kitchen scale between drinks to simplify event parsing and ensure consistency in fill level. Drinks were consumed when the container was either completely or half-filled. For every container type and fill level combination, each participant took 5 drinks (i.e.: 30 total drinks/participant).

Data was collected using the system introduced in Chapter 3. Only data from the accelerometer is used in the current analysis, with examination of the gyroscope output reserved for future work. Results analyzed herein are obtained from a sensor placed at the bottom of each container. The sensor was oriented along the container's cross-section at a 180-degree offset from the instructed point of drinking (i.e.: on the side opposite the mouth, approximately 90 degrees offset from the grasping hand). This orientation is depicted for the refillable bottle in Figure 1-1. Data from a second sensor placed at the vertical midpoint of each container opposite the drinking hand, along with a container worn on the wrist of the participant, is not presented in the current chapter.

Data was preprocessed using the smoothing and resampling techniques detailed in Chapter 3. Sensor outputs were then transformed to a common coordinate frame (i.e.: x' component aligned with static acceleration due to gravity, y' component parallel with the surface of the table).

This was necessary to account for the slant in the glass and mug walls due to tapering of the cross-sectional area over the container height, and to adjust for any rotations of the sensor from ideal placement in the surface plane. This process was accomplished by determining offset angles during the initial portion of the recording while the containers were placed stationary on a level surface, under the assumption that only static acceleration due to gravity was present in the signal during this interval.

After transformation, each container's inclination with respect to gravity was estimated under the assumption of negligible translational forces as specified in (3.1). Drink parsing was performed using the algorithm introduced in Chapter 3. Example inclination signatures for the three containers considered are shown in Figure 10-2.

Due to motivation described in Chapter 4, further inter-event parsing was used to parse the drinking event into microevents. All resulting features were computed on the sip microevent occurring in the middle of the drinking event. Rather than employ the parsing technique described in Chapter 4, a more simplistic segmentation technique was utilized based upon the observed signal morphologies during drinking. Namely, the largest continuous interval for which the inclination exceeded 20% of the maximum value was extracted. This relative threshold was employed to reflect the variation in inclination amplitude across containers. An example of this inter-event parsing is shown in Figure 10-3.

A set of support vector machine (SVM) classifiers were trained for each application considered using the previously proposed inclination signature (IS) feature set.

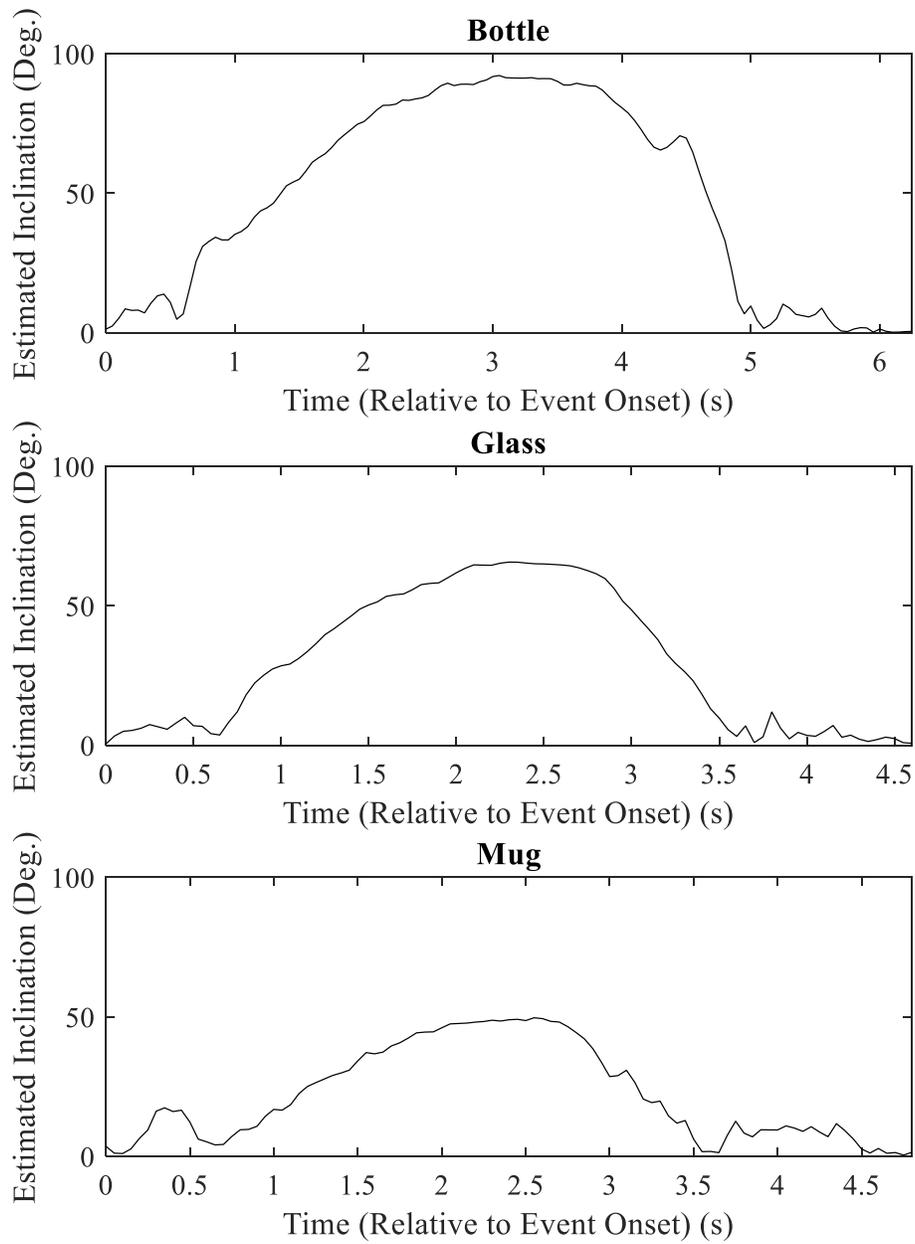


Figure 10-2: Inclination Signatures for the Three Container Types (Half-Full Fill Level)

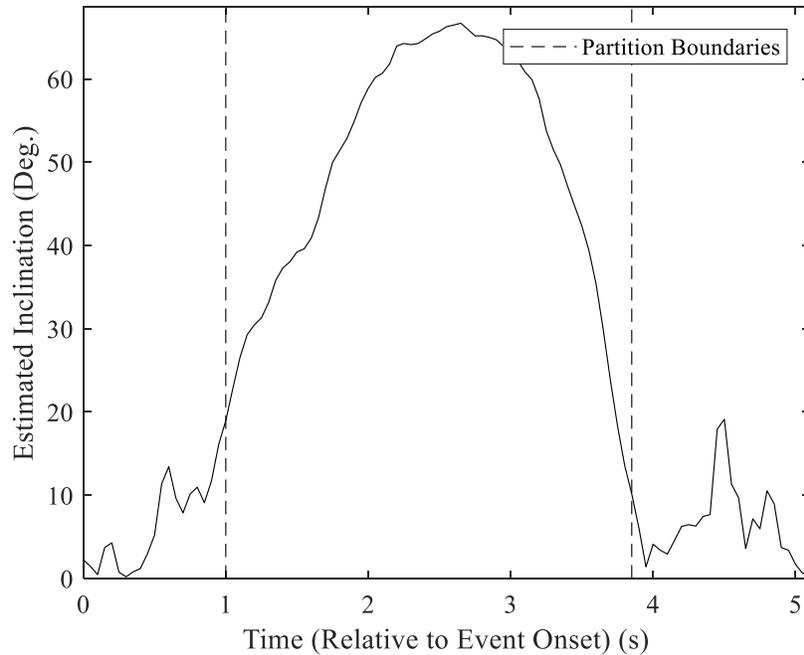


Figure 10-3: Partitioning the Drinking Interval Using Relative Thresholding

The following kernel functions were evaluated in each scenario— 1) linear, 2) cubic, 3) quadratic, and 4) Gaussian. Hyperparameters were set to the default values employed in *MATLAB*'s Classification Learner application (i.e.: Box Constraint = 1, kernel scale = 5.7 for Gaussian kernel, with values for other kernels computed automatically using a heuristic procedure implemented within the software package). Classifiers were trained for container type classification at both individual fill levels (i.e.: full and half-full), along with mixed data from both fill levels. In addition, models were trained for fill level classification for each of the three container types considered.

Two unique training scenarios were considered for each classification application. The first, hereby denoted as leave-one-subject-out (LOSO) training, trained each classifier using data exclusively from other subjects (i.e.: for testing on Subject 1, training data is gathered exclusively from Subjects 2 – 5). In the second training approach, hereby denoted as subject-specific training,

only training data from the subject under test is utilized. To maximize the use of available data for subject-specific training, a leave-one-drink-out (LODO) cross-validation strategy was employed (i.e.: for a subject specific model attempting to classify container type, models for each drink under test are trained using the 14 remaining drinks). Each SVM model was trained using the default iterative single data algorithm as implemented within the *fitsvm* function in MATLAB.

10.3 Results

10.3.1 Container-Type Classification – LOSO Training

Four SVM classifiers with varying kernels described in the prior section were used to classify container type. For LOSO training at both fill levels considered, each model was trained using the 60 drink samples gathered from other subjects, and subsequently tested on the 15 samples for the test subject. Classification accuracies for each model are presented in Tables 10-1 (half-full) and 10-2 (full fill level). For the two fill levels considered, superior classification accuracy is observed for the half-full fill level. In this scenario, differences in container geometry are more clearly reflected in the inclination signal morphology. Namely, taller containers such as the bottle require greater inclination to induce fluid flow versus shorter containers such as the mug and glass.

Table 10-1: Container Type Classification Accuracy: LOSO Training, Half-Full Fill

Subj, ID /Kernel	S1	S2	S3	S4	S5	Avg
Linear	93.3%	100%	100%	100%	93.3%	97.3%
Quadratic	73.3%	100%	93.3%	100%	93.3%	92.0%
Cubic	73.3%	100%	86.7%	100%	93.3%	90.7%
Gaussian	66.7%	100%	100%	100%	93.3%	92.0%

Table 10-2: Container Type Classification Accuracy: LOSO Training, Full Fill

Subj, ID /Kernel	S1	S2	S3	S4	S5	Avg
Linear	73.3%	80.0%	86.7%	80.0%	66.7%	77.3%
Quadratic	46.7%	73.3%	66.7%	80.0%	73.3%	68.0%
Cubic	53.3%	66.7%	66.7%	60.0%	73.3%	64.0%
Gaussian	66.7%	86.7%	73.3%	66.7%	66.7%	72.0%

Table 10-3 shows classification accuracy for models trained on a mixture of data from both fill levels (i.e.: 120 training examples/subject). While considerable variability in the inclination signal morphology versus fill level complicates this classification, best-case performance across the set of models considered is only slightly reduced from the full fill level case.

Table 10-3: Container Type Classification Accuracy: LOSO Training, Mixed Fill

Subj, ID /Kernel	S1	S2	S3	S4	S5	Avg
Linear	73.3%	76.7%	60.0%	73.3%	66.7%	70.0%
Quadratic	56.7%	73.3%	70.0%	90.0%	76.7%	73.3%
Cubic	56.7%	70.0%	76.7%	73.3%	73.3%	70.0%
Gaussian	63.3%	80.0%	66.7%	83.3%	76.7%	74.0%

Container type misclassifications are most common amongst the glass and mug samples as demonstrated in the confusion matrices presented in Table 10-4. This error type is especially prevalent for scenarios where fill level is controlled. These matrices are obtained by taking the best-case classification accuracy for each considered scenario (i.e.: linear SVM for full and half-full levels, Gaussian SVM for mixed data).

Table 10-4: Confusion Matrices: LOSO Training

Half-Full – Linear				Full – Linear			
True/ Predict	<i>Bottle</i>	<i>Glass</i>	<i>Mug</i>	True/ Predict	<i>Bottle</i>	<i>Glass</i>	<i>Mug</i>
<i>Bottle</i>	25	0	0	<i>Bottle</i>	25	0	0
<i>Glass</i>	0	24	1	<i>Glass</i>	0	14	11
<i>Mug</i>	0	1	24	<i>Mug</i>	0	6	19

Mixed – Gaussian

True/ Predict	<i>Bottle</i>	<i>Glass</i>	<i>Mug</i>
<i>Bottle</i>	40	6	4
<i>Glass</i>	2	33	15
<i>Mug</i>	1	11	38

10.3.2 Container-Type Classification – Subject Specific Training

The above process was repeated using the subject-specific training strategy described in III.C. Namely, on a per-subject basis, each drink was successively tested using a classifier trained from the remaining 14 drinks. Classification accuracies for the half-full, full, and mixed fill levels are presented in Tables 10-5, 10-6, and 10-7, respectively. As noted, although available training data is reduced from the LOSO strategy, best-case performance is improved for all three scenarios.

Table 10-5: Container Type Classification Accuracy: S.S. Training, Half-Full

Subj, ID /Kernel	S1	S2	S3	S4	S5	Avg
Linear	100%	100%	86.7%	100%	100%	97.3%
Quadratic	100%	100%	93.3%	100%	100%	98.7%
Cubic	100%	100%	93.3%	100%	100%	98.7%
Gaussian	100%	100%	100%	100%	93.3%	98.7%

Table 10-6: : Container Type Classification Accuracy: S.S. Training, Full Fill

Subj, ID /Kernel	S1	S2	S3	S4	S5	Avg
Linear	86.7%	66.7%	73.3%	80.0%	100%	81.3%
Quadratic	93.3%	80.0%	73.3%	93.3%	100%	88.0%
Cubic	86.7%	73.3%	66.7%	100%	100%	85.3%
Gaussian	80.0%	66.7%	66.7%	86.7%	93.3%	78.7%

Table 10-7: Container Type Classification Accuracy: S.S. Training, Mixed Fill

Subj, ID /Kernel	S1	S2	S3	S4	S5	Avg
Linear	70.0%	53.3%	56.7%	66.7%	60.0%	61.3%
Quadratic	93.3%	70.0%	66.7%	73.3%	70.0%	74.7%
Cubic	93.3%	83.3%	66.7%	66.7%	76.7%	77.3%
Gaussian	80.0%	70.0%	60.0%	73.3%	70.0%	70.7%

As depicted in Table 10-8, classification errors follow a similar distribution to those observed for the LOSO models.

Table 10-8: Confusion Matrices: Subject-Specific Training

Half-Full – Linear				Full – Linear			
True/ Predict	<i>Bottle</i>	<i>Glass</i>	<i>Mug</i>	True/ Predict	<i>Bottle</i>	<i>Glass</i>	<i>Mug</i>
<i>Bottle</i>	25	0	0	<i>Bottle</i>	25	0	0
<i>Glass</i>	0	25	0	<i>Glass</i>	0	18	7
<i>Mug</i>	0	1	24	<i>Mug</i>	0	2	23

Mixed – Gaussian

True/ Predict	<i>Bottle</i>	<i>Glass</i>	<i>Mug</i>
<i>Bottle</i>	41	8	1
<i>Glass</i>	5	36	9
<i>Mug</i>	0	11	39

10.3.3 Container Type Classification with Equivalent Training Samples

To facilitate fairer comparisons between the two training techniques, the LOSO approach was analyzed using only 15 randomly chosen training samples from the 60 available. This process was repeated five times for varying random seeds for the linear SVM model only. Comparative classification accuracy between the three techniques (averaged across trials for LOSO-restricted) is depicted in Figure 10-3. For an equal amount of training samples, subject-specific models outperform those trained out-of-subject in all scenarios (14.1%, 13.8%, and 11.6% for full, half, and mixed fill levels, respectively).

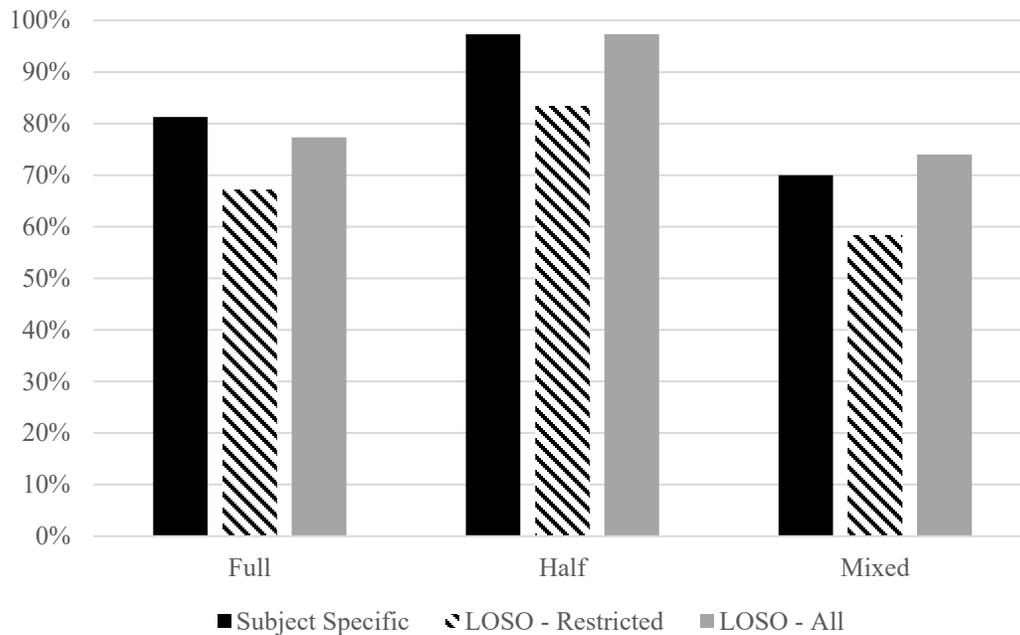


Figure 10-4: Variation in Container Type Classification Accuracy

10.3.4 Fill Level Classification

Classifiers were trained to distinguish the two initial fill levels considered for each of the three containers. Using the LOSO strategy, 100% accuracy was achieved for all subjects for both

the glass and mug for each model considered. Variability in accuracy across subjects for the various bottle models is depicted in Table 10-9. As noted, classification performance is still strong for this container type (98.0% for the best performing model), with errors isolated to only two of the five subjects.

Table 10-9: Fill Level Classification Accuracy: Bottle Container, LOSO Training

Subj, ID /Kernel	S1	S2	S3	S4	S5	Avg
Linear	100%	100%	100%	100%	90.0%	98.0%
Quadratic	100%	100%	100%	100%	90.0%	98.0%
Cubic	100%	100%	100%	90.0%	90.0%	96.0%
Gaussian	100%	100%	100%	100%	90.0%	98.0%

Subject-specific models were also trained to classify fill level. As was the case with the LOSO strategy, 100% accuracy was achieved for all subjects and models considered for both the glass and mug. While 4 of the 5 models considered for bottle data produced 100% accuracy for each subject, the Gaussian kernel model experienced some misclassifications. Namely, misses occurred for both subjects 4 and 5, yielding an average accuracy of 96.0% across subjects for this kernel.

10.4 Discussion

As shown in the confusion matrices presented in Tables 10-4 and 10-8, drinks consumed from the bottle were easily distinguished from those taken from the glass and mug at a fixed fill level. We hypothesize that this is related to the highly distinctive geometry of the prior container type versus the latter two. This belief is strengthened through examination of the improved glass-mug classification accuracy for drinks consumed from the half-full versus the full fill level. When both containers are filled, flow may be induced with only slight inclinations from both drinking vessels.

However, when each container is half-full, greater inclination is required to induce flow from the glass versus the mug due to height differences.

When initial fill levels were mixed, the accuracy of container type classification was reduced. In practical deployment, classification for a fixed fill level or limited range (i.e.: near full) is likely sufficient to provide users with the desired utility of the sensor. Namely, users could be instructed to consume an initial drink from a near-full fill level upon sensor repositioning to allow for automatic container type detection. This would support subsequent deployment of container-specific consumption estimation models, which is the primary metric of concern for the intended use case of the device.

While the fill level classification accuracies reported herein are promising, further investigation is necessary to assess the feasibility of employing this strategy for fluid consumption estimation. As the accuracy of consumption estimates using this approach is inherently limited by the resolution of fill levels which may be reliably classified, additional analysis for more closely separated fill levels is required. In addition, consideration must be given to the effect of user intent and associated drink volume within this estimation process. As was specified in Section 10.2, participants were instructed to take normal drinks in each trial. However, daily use will involve scenarios where the user intends to consume either an above or below average amount of fluid depending upon thirst. As shown in Figure 10-4, which depicts variability in drink volume versus maximum inclination angle, volume influences the amplitude of the inclination signature, further complicating classification.

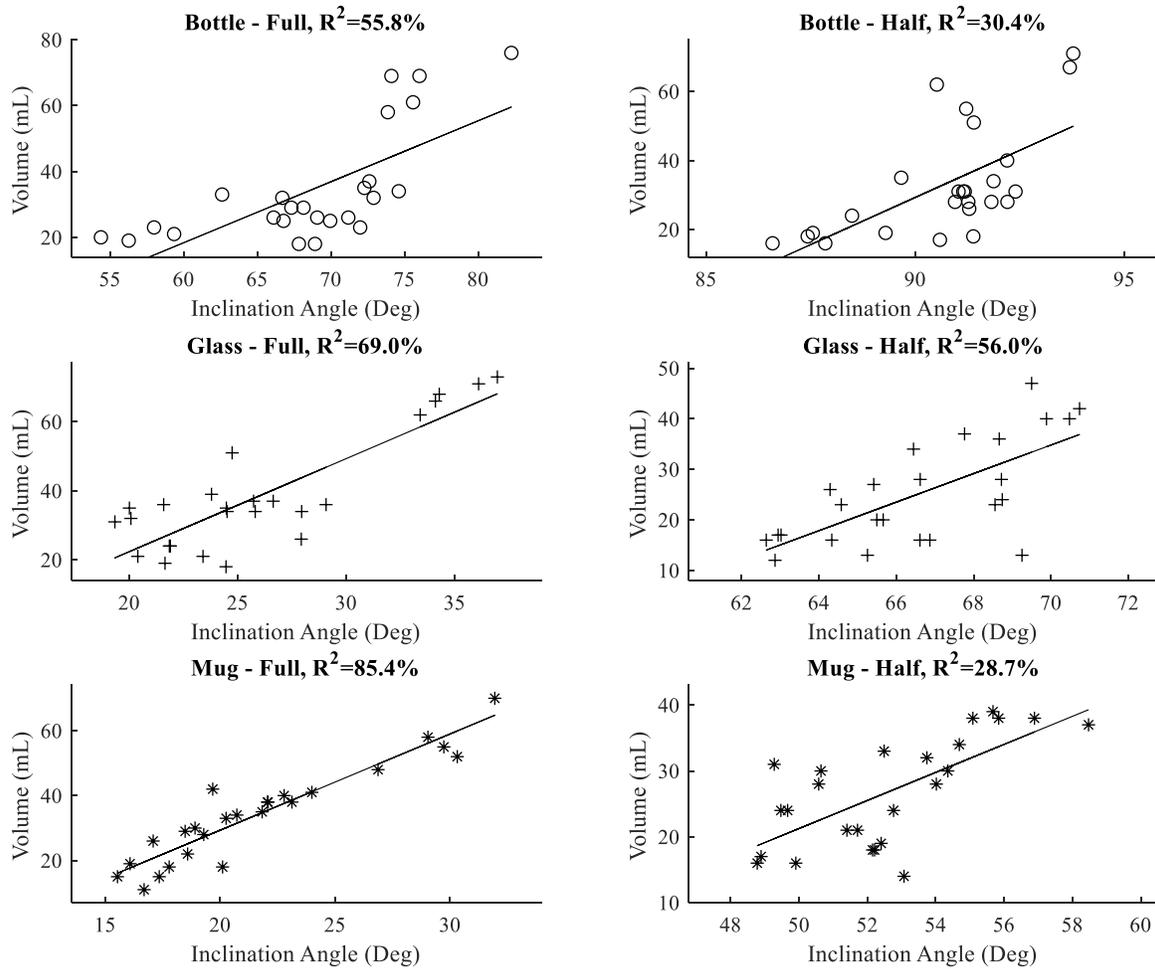


Figure 10-5: Drink Volume Versus Maximum Inclination Angle

10.5 Summary and Future Work

The ability of a bottle-attachable IMU sensor to classify both container type (bottle, glass, and mug) and fill level (full and half-full) was demonstrated herein. Classification was performed using SVMs with hand-engineered features describing each container's estimated inclination during drinking. A best-case accuracy of 98.7% was achieved for container type classification using subject-specific models at a fixed fill level. A best-case accuracy of 100% was achieved for container-specific fill level classification using subject-specific models. Variability in accuracy

versus training strategy was also explored, with subject-specific training demonstrating superior performance versus out-of-subject training for an equal amount of training data (container type classification accuracy improvement of 13.3% for full, 11.4% for half, and 10.3% for mixed fill levels using subject-specific models).

Future work should focus on analyzing the additional data collected for this experiment. Data from the second sensor attached at the vertical midline of each container will be processed to explore potential performance variability as a function of sensor placement. In addition, data from the wrist-worn sensor will be analyzed to compare achievable accuracy between the two alternative sensing strategies.

Chapter 11 : Conclusions

11.1 Summary

Various strategies for improving the performance of a container-attachable hydration tracking sensor were proposed and verified throughout this dissertation. A novel sip detection algorithm was introduced in Chapter 3. This technique was demonstrated to improve classification accuracy and enhance efficiency versus a benchmark algorithm employing static segmentation. Results were verified using a scripted experiment intended to mimic the intended daily use case of the device.

Approaches for improving drink volume estimation accuracy were explored in Chapters 4 - 9. Per-drink estimation accuracy was improved versus prior state-of-the-art results for a single inertial sensor. The accuracy of aggregate consumption estimates was also increased versus previously reported results for the sensor considered herein.

An alternative technique for estimating aggregate consumption using fill ratio estimates was proposed and explored. Fill ratio estimators were shown to exhibit improved accuracy and reduced inter-subject variability compared to volume models. A heuristic fusion approach for enhancing the accuracy of these estimates was also verified. The manuscript concluded by demonstrating the feasibility of using the sensor for multiple types of drinking vessels.

11.2 Limitations

Although the proposed attachable architecture offers notable advantages versus competitive approaches, it is characterized by some fundamental limitations. Namely, the motion-based sensing mechanism restricts use to drinking vessels in which flow is introduced through inclination (i.e.: no straw-based containers, etc.). Furthermore, the device limits ubiquity relative to wearable sensors, due to the requirement that dedicated hardware be manually repositioned on the container before each drinking episode.

Beyond these innate restrictions, generalization of the results presented herein is limited by the scripted nature of the described experiments. As noted in Chapters 3 and 4, scripted experiments were utilized due to limitations of the data collection system, along with the challenges of capturing scale-based ground truth data on a per-drink basis in an unscripted scenario. Moreover, volume prompts were used to ensure that a wide variety of drink volumes were captured to support regression model development. Further research should address these limitations by evaluating the proposed sip detection and volume estimation algorithms on data collected during free living conditions. The potential impact of evaluating the proposed techniques on such data is discussed in the following section where appropriate.

11.3 Summary of Key Contributions and Recommendations for Future Work

The key contributions of this work are summarized below. A discussion of each advancement is also provided.

- 1. Proposal and verification of a novel two-stage dynamic partitioning and classification algorithm for sip detection*

The sip detection algorithm detailed within Chapter 3 was demonstrated to improve true positive detection rate from 75.1% to 98.8% versus a benchmark algorithm employing static segmentation. This static windowing approach was chosen as a benchmark due to its prevalence throughout traditional activity detection literature. The key novelty of the proposed algorithm is the first-stage strategy for spotting drinking events using the characteristic drinking motion pattern. Versus alternative approaches relying on component-level inertial sensor outputs, this technique allows for the setting of parameters in a mechanistic sense.

The configurable nature of the sensor greatly simplifies sip detection compared to wearable architectures. Additionally, sip detection results reported in the literature for all hydration tracking technologies are generally far superior to those presented for volume estimation. Therefore, while further investigation of both the proposed (i.e.: parameter optimization, etc.) and alternative algorithms may yield slight performance improvement for the target architecture, it is recommended that future research efforts focus on enhancing consumption estimation performance as described in the following sections.

2. *Demonstration of state-of-the-art volume estimation results for a single inertial sensor on a per-drink basis*

The SVM regression model proposed in Chapter 5 was demonstrated to improve the mean absolute percentage accuracy of volume estimates by 11.1% versus state-of-the-art results for a single inertial sensor. While the proposed techniques were restricted to SVM regression models, it should be reemphasized that various other learning models (i.e.: trees, Gaussian Process Regression Models, end-to-end learning architectures, etc.) were also explored as part of this research, with the prior yielding superior performance. More sophisticated models may benefit from the enhancement of training data scale.

The comparison of subject-specific models to those trained out of subject further elucidates the complexity of the volume estimation problem. Namely, while motion characteristics (i.e.: duration, inclination kinematics, etc.) may be related to drink volume on an individual level, these relationships do not appear to generalize across a broader population based upon the experiments performed herein. One possible explanation is individual-specific shaping of the mouth during periods of fluid intake. Therefore, while improvements in volume estimation accuracy should be explored, it is recommended that

future efforts are more focused on residual volume estimation strategies using estimated fill levels.

3. *Demonstration of improved aggregate consumption results for a container-attachable inertial sensor*

Aggregate consumption estimation was improved relative to prior reported results for the attachable sensor architecture. Accuracies were comparable to those reported for other sensor modalities. Further efforts should explore potential variations in aggregate performance for consumption sequences occurring during daily use in a non-scripted environment. Aggregate consumption estimation results may differ when the variance of volumes within a sequence of drinks is reduced versus the scripted results considered within this research.

4. *Demonstration of high-resolution fill ratio estimation using drink motion patterns*

SVM regression models for estimating the initial fill level from which a drink was consumed were introduced within this dissertation. While the classification of fill level had previously been demonstrated in the literature for low resolution labels, we are unaware of any prior work using regression-based approaches for high-resolution data. With respect to volume estimators, fill level regression models were shown to exhibit considerably improved accuracy. Variability in accuracy across trials was also significantly limited relative to volume results. Subject-specific analysis suggested that the relationship between the motion pattern during drinking and the associated fill level are largely subject-independent. Given this observation, it is recommended that future research focus on residual volume techniques for estimating aggregate consumption.

While designed for the sensor architecture considered herein, this technique of training to fill ratio labels (or equivalently, aggregate container volume) could be implemented for alternative motion-based technologies. Although some limited collections were performed within this research using both a wearable and container-attachable sensor (i.e.: Chapters 3 and 10), additional large-scale data collection is recommended to fully assess the generalization of this phenomenon to alternative sensor placements.

5. *Demonstration of a heuristic fusion technique for improving fill ratio estimation performance*

A technique for fusing fill ratio estimates produced by regression models with those generated using a heuristic consumption model were demonstrated. This strategy was implemented using a Kalman filtering framework. Similar to the discussion for contribution 3, this technique should be reinvestigated for drink sequences exhibiting typical variation in volume across drinks. It is anticipated that this model will perform better for such scenarios.

6. *Demonstration of fill level and container type classification for multiple drinking vessels*

The ability of the sensor to track aggregate daily consumption across multiple types of containers is a key value proposition of the proposed device. The work presented in Chapter 10 demonstrates initial proof-of-concept of this functionality. Verification of the proposed techniques should be conducted for a large-scale data collection for all container types of interest.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] G. K. Garge, C. Balakrishna, and S. K. Datta, "Consumer Health Care: Current Trends in Consumer Health Monitoring," *IEEE Consumer Electronics Magazine*, vol. 7, no. 1, pp. 38–46, Jan. 2018.
- [2] J. Andreu-Perez, D. R. Leff, H. M. D. Ip, and G. Yang, "From Wearable Sensors to Smart Implants—Toward Pervasive and Personalized Healthcare," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 12, pp. 2750–2762, Dec. 2015.
- [3] S. A. Schroeder, "We Can Do Better — Improving the Health of the American People," *New England Journal of Medicine*, vol. 357, no. 12, pp. 1221–1228, Sep. 2007.
- [4] J. D. Stookey, "High prevalence of plasma hypertonicity among community-dwelling older adults: results from NHANES III," *Journal of the American Dietetic Association*, vol. 105, no. 8, pp. 1231–1239, 2005.
- [5] A. D. Seal, H.-G. Suh, L. T. Jansen, L. G. Summers, and S. A. Kavouras, "Chapter 11 - Hydration and Health," in *Analysis in Nutrition Research*, G. Pounis, Ed. Academic Press, 2019, pp. 299–319.
- [6] A. 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.

- [11] 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.
- [12] 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.
- [13] J. H. M. Bergmann, V. Chandaria, and A. McGregor, “Wearable and Implantable Sensors: The Patient’s Perspective,” *Sensors*, vol. 12, no. 12, pp. 16695–16709, Dec. 2012.
- [14] J. E. Lewis and M. B. Neider, “Designing wearable technology for an aging population,” *Ergonomics in Design*, vol. 25, no. 3, pp. 4–10, 2017.
- [15] B. Dong, R. Gallant, and S. Biswas, “A self-monitoring water bottle for tracking liquid intake,” in *2014 IEEE Healthcare Innovation Conference (HIC)*, 2014, pp. 311–314.
- [16] H. Griffith and S. Biswas, “Container Type and Fill Level Classification Using a Bottle-Attachable IMU Sensor,” in *Proceedings of the 23rd IEEE International Symposium on Consumer Technologies (In-Press)*.
- [17] E. Thomaz, I. Essa, and G. D. Abowd, “A Practical Approach for Recognizing Eating Moments with Wrist-mounted Inertial Sensing,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, NY, USA, 2015, pp. 1029–1040.
- [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] E. Jovanov, V. R. Nallathimmarreddygari, and J. E. Pryor, “SmartStuff: A case study of a smart water bottle,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 6307–6310.
- [23] M. S. Borofsky, C. A. Dauw, N. York, C. Terry, and J. E. Lingeman, “Accuracy of daily fluid intake measurements using a ‘smart’ water bottle,” *Urolithiasis*, vol. 46, no. 4, pp. 343–348, 2018.

- [24] 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.
- [25] 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.
- [26] 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.
- [27] K. San Chun, A. B. Sanders, R. Adaimi, N. Streeper, D. E. Conroy, and E. Thomaz, “Towards a generalizable method for detecting fluid intake with wrist-mounted sensors and adaptive segmentation,” in *IUI. International Conference on Intelligent User Interfaces*, 2019, vol. 2019, p. 80.
- [28] 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.
- [29] D. Gomes and I. Sousa, “Real-Time drink trigger detection in free-living conditions using inertial sensors,” *Sensors*, vol. 19, no. 9, p. 2145, 2019.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] G. W. Lamb, “Beverage Fill Level Detection and Indication,” US20080083475A1, 10-Apr-2008.
- [34] 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.
- [35] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, “Window Size Impact in Human Activity Recognition,” *Sensors (Basel)*, vol. 14, no. 4, pp. 6474–6499, Apr. 2014.

- [36] 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.
- [37] N. C. Krishnan and D. J. Cook, “Activity recognition on streaming sensor data,” *Pervasive and mobile computing*, vol. 10, pp. 138–154, 2014.
- [38] 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.
- [39] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in *International Conference on Pervasive Computing, 2004*, pp. 1–17.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] P. Morguet, Stochastic modeling of image sequences for the segmentation and recognition of dynamic gestures, Ph.D. Thesis, Technische Universität München, 2000.

- [48] 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.
- [49] 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.
- [50] D. M. Saltzgiver and J. T. S. III, “Liquid consumption tracker,” US20160025545A1, 28-Jan-2016.
- [51] “Using Inertial Sensors for Position and Orientation Estimation,” in *Foundations and Trends® in Signal Processing*, vol. 11, 2017, pp. 1–153.
- [52] “Introduction Blender Manual.” [Online]. Available: www.docs.blender.org/manual/en/dev/editors/movie_clip_editor/tracking/introduction.html. [Accessed: 1-Sep-2018].