Eating and Drinking Gesture Spotting and Recognition Using a Novel Adaptive Segmentation Technique and a Gesture Discrepancy Measure

Dario Ortega Anderez[1], Ahmad Lotfi*, Amir Pourabdollah

School of Science and Technology, Nottingham Trent University,
Clifton Lane, Nottingham, NG11 8NS,
United Kingdom
Email: {ahmad.lotfi; dario.ortegaanderez2013; amir.pourabdollah}@ntu.ac.uk

---

Abstract

Despite the increasing developments on human activity recognition using wearable technology, there are still many open challenges in spotting and recognising sporadic gestures. As opposed to activities, which exhibit continuous behaviour, the difficulty of spotting gestures lies in their rather sparse nature. This paper proposes a novel solution to spot and recognise a set of similar eating and drinking gestures from continuous inertial data streams. First, potential segments containing an eating or a drinking gesture are found using a Crossings-based Adaptive Segmentation Technique (CAST). Second, further to the long-established range of features employed in previous human activities recognition research work, a gesture discrepancy measure is proposed to improve the classification performance of the system. At the final step, a range of state-of-the-art classification models is employed for evaluation. Various conclusions can be drawn from the results obtained. First, given the 100% recall achieved at the segmentation step, the CAST can be considered a reliable segmentation technique for spotting drinking and eating gestures which may be employed in future gesture spotting work. Second, the addition of gesture discrepancy as a feature descriptor consistently improves the classification performance of the system. Third, the reliability of the food and drink intake monitoring approach proposed in this work finds support on the out-performance of previous similar work.

*Corresponding author: Email: *ahmad.lotfi@ntu.ac.uk*
[1]PhD research student who has conducted the research as part of his thesis.

Keywords: Gesture recognition, Gesture spotting, Wearable Sensors, Adaptive signal segmentation

1. Introduction

Current wearable and portable technologies such as smart phones, smart watches or fitness trackers incorporate a great array of sensors, allowing for human behaviour analysis in different applications. Examples include fitness (Wundersitz et al., 2015), rehabilitation (Billiet et al., 2016), security (Mahler et al., 2017) or health care (Chernbumroong et al., 2013). In line with the latter application, the efforts of this paper are given to gain insights into the fluid and food intake, which as suggested in Zhang et al. (2009), can be crucial for many applications related to measuring wellness and/or support for independent living.

Dietary behaviour plays an important role in our day to day lives and health. While obesity is a major risk factor for heart diseases, stroke, high blood pressure or diabetes (Wellman and Friedberg, 2002), malnutrition is considered as a confounding factor for developing chronic diseases (Amft et al., 2007). Dietary behaviour is normally tracked in the form of self-assessment questionnaires. However, two major drawbacks are found in the use of conventional dietary tracking approaches. First, the data entry process may result cumbersome, since questionnaires have to typically be filled manually by the subjects. Second, numerous studies indicate self-reported estimates of daily activities are subjective and variable (Smith et al., 2005; C. rush et al., 2008). With reference to dietary behaviour, people tend to under-report their food consumption (Schoeller et al., 2013).

Additionally, maintaining an adequate hydration level is an important aspect in dietary management (Sawka et al., 2005). Particularly, fluid intake is a severe issue in elderly care, where diminished thirst perception is frequently related to reduced cognitive capabilities, leading concurrently to difficulty at remembering to drink enough (Kenney and Chiu, 2001). Approximately 17 million people suffer a stroke yearly (Mackay, 2004), with 77% of them enduring an upper extremity disability or a function loss of the limb upper motor (Lawrence et al., 2001). Such function loss may lead stroke patients to difficulty at performing basic actions like eating or drinking, therefore limiting their own independence (Chen et al., 2017).

Increasing developments have been achieved in Human Activity Recognition (HAR) with the use of inertial sensors, however, efforts are primarily given to the recognition of quasi-periodic activities such us climbing stairs, walking or running (Mannini and Sabatini, 2010; Liu et al., 2016). As opposed to activities, which exhibit continuous behaviour in time, the difficulty of spotting gestures lies in their rather sparse nature. Further, spotting naturally learned gestures such as grasping a fork, has been shown to be harder than detecting gestures which have been purposely trained within a constrained environment, e.g. human-machine interaction gestures (Junker et al., 2008).

To clarify the distinct terminology used in this paper, it should be noted that the conducted research related to gesture recognition undertakes the classification of already labelled signal segments or windows while gesture spotting attempts the identification of potential segments containing one of the gestures within the sought gesture set. Although some previous studies have only undertaken gesture recognition, a complete gesture tracking system should include both gesture spotting and recognition since continuous data streams not only include gestures within the gesture target set but also a 'Null' class composed of other gestures/activities as well.

The issues mentioned above alongside the various open challenges in spotting and recognising naturally learned gestures motivate the search for solutions towards the development of an automatic non-invasive fluid and food intake monitoring system. In line with this, this paper proposes a novel and comprehensive approach to spot and recognise a set of four different eating and drinking gestures using a single wrist-worn inertial unit. Based on the analysis of previous work and the results achieved in this work, the following contributions are made:

1. Evaluate and validate the segmentation technique proposed in our previous work (Ortega-Anderez et al., 2018a) on a larger data set which includes additional intra-person and inter-person variability, as well as a more extensive 'Null' class (activities without the sough gesture set). An outstanding 100% recall is achieved at the segmentation stage, supporting the reliability of this segmentation technique.

2. Propose the addition of a Soft Dynamic Time Warping (Soft-DTW) gesture discrepancy to activity/gesture recognition systems. To the best of our knowledge, previous published papers on HAR have not considered the use of gesture discrepancy. Given the intra-person and inter-person variability as well as the duration intra-variability of the studied gestures, we believe the addition of gesture discrepancy to long-established HAR feature vectors can increase the classification performance of current systems. The results achieved in this work go in accordance with the above intuition.

3. Present a reliable fluid and food intake tracking solution which finds support on the out-performance of previous similar work.

Recent statistics show eating difficulties are a prevalent issue among the elderly population (Westergren et al., 2002). Furthermore, in many cases, individuals require some form of eating assistance (Lohrmann et al., 2003). The lack or diminution of performance of fluid or food intake by a subject, can potentially indicate the need for peripheral support or the inability for independent living. The above contributions not only imply a great step forward towards the development of an intelligent system for the identification of eating difficulties or eating neglect from subjects, but a valuable input in the form of an adaptable and flexible novel segmentation technique (CAST) and the introduction of a feature descriptor based on gesture discrepancy for their employment in future work on intelligent systems for activity and gesture recognition as well.

The rest of the paper is organised as follows: Section 2 reviews previous work on segmentation of time series as well as on gesture spotting and recognition. Section 3 presents the method proposed for the development of a fluid and food intake tracking system. Section 4 presents the results achieved and compares them to those of previous similar published works. Section 5 reports the conclusions drawn from the achieved results.

2. Previous Work

This section provides a review of the published works on time series segmentation as well as on gesture spotting and recognition. A discussion on the findings that motivates the proposed research work is given at the end. For reading convenience, this section has been divided according to the aforementioned topics.

2.1. Time Series Segmentation

Despite the increasing achievements in HAR using wearable sensors, efforts are principally given to the recognition of quasi-periodic activities such as walking, stairs climbing or running (Mannini and Sabatini, 2010; Liu et al., 2016). Given the continuous nature of the studied activities, an artificial segmentation technique, whereby the collected time streams are divided into consecutive (often overlapping) time windows or fundamental motion segments of equal length, is normally applied (Kwon et al., 2014; Wen and Wang, 2017; Ronao and Cho, 2016; Chernbumroong et al., 2013; Ortega-Anderez et al., 2018b). Typically, the window length is either decided based on previous HAR work or calculated as a hyper-parameter of the classification problem. This, as suggested in Anderez et al. (2018), indicates the window length is dependent on the activity set is studied. That is, windows must be sufficiently long to capture fundamental characteristics of the signal but sufficiently short to avoid capturing signal from multiple activities.

Given the sparse distribution in time of sporadic actions or gestures, adaptive segmentation techniques have been shown to offer better performance (Noor et al., 2017). Within adaptive segmentation techniques, Piecewise Linear Representations (PLRs) are well-known techniques (Keogh et al., 2004; Lovrić et al., 2014). In PLRs, segments of time series are approximated to a line either by the application of linear regression or interpolation, until a customised threshold error is exceeded. A posteriori, a Feature Similarity Search (FSS) is normally used to narrow down the number of segments (Junker et al., 2008). In point of fact, the work in (Junker et al., 2008), employed a PLR, namely the Sliding Window and Bottom-up (SWAB), to spot a set of fluid and food intake gestures.

Besides PLRs, various customised segmentation approaches have been proposed for spotting sporadic gestures or actions from continuous inertial data streams. Noor et al. (2017) used an extendable Gaussian Probability Function-based window. Parate et al. (2014) employed a segmentation approach based on a re-adjustable resting position and a distance peak detector from the most current resting position. Dong et al. (2014) utilised a wrist motion energy threshold-based approach. Xu et al. (2012) used sign changes on the accelerometer signal to divide it into different segments.

2.2. Gesture Spotting and Recognition

Numerous solutions for spotting and recognising gestures have been proposed in recent years. Chen et al. (2017) studied the recognition of drinking gestures using a single wrist-worn inertial sensor. A recall of 91.3% was achieved using an SVM classifier on a feature vector calculated over windows of 0.25 seconds.

Xu et al. (2012) proposed a solution to recognise a set of seven basic hand gestures for human-machine interaction purposes using bi-axial data from a tri-axial accelerometer. A set of ten features was used to determine the gesture termination points. Once segments were found, three different models were proposed for the recognition of the gestures. Among the three models, the best results were achieved by a template matching model (95.6% classification accuracy). Similar work by Tai et al. (2018) employed an LSTM network to recognize a set of six different hand gestures using tri-axial accelerometer and tri-axial gyroscope data from five users. The proposed LSTM-based approach achieved a classification accuracy of 95.85%.

Schiboni and Amft (2018) developed a Gaussian Mixture Hidden Markov Models (GMM-HMMs) network for spotting drinking gestures. The experimental data were collected from 7 users following their usual daily activities while wearing a single wrist-worn inertial sensor which included a tri-axial accelerometer, a tri-axial gyroscope and a tri-axial magnetometer. An average precision of 75.2% and recall of 76.1% were achieved.

An adaptive segmentation technique to spot a set of four transitional activities (sit-to-stand, stand-to-sit, sit-to-lie and lie-to-sit) was developed by Noor et al. (2017) using data from a waist-worn accelerometer. First, a set of thirteen features was used on windows of fixed length to determine whether the different windows contained a transitional, a dynamic or a static activity. Windows classified as a transitional activity were extended until a decrease in likelihood for a particular transitional activity, given by the Gaussian probability density function, was identified. The results demonstrated an improvement in classification recall from 89.9% using an artificial segmentation approach to 93.0% with the adaptive segmentation technique.

A drinking spotting solution based on a Feature Similarity Search (FSS) was proposed by Amft et al. (2010). Data was collected from six users wearing a single wrist-worn inertial unit containing a tri-axial accelerometer, a tri-axial gyroscope and a tri-axial compass while performing a set of various free-living scenarios. A classification recall of 84.0% was achieved.

A solution for spotting and recognising smoking gestures using data from a wrist-worn quaternion was proposed by Parate et al. (2014). First, gestures were detected using a rest position tracking algorithm alongside a peak detector used to detect peaks on the distance between the most recent rest position and the current position. Further to the spotting stage, a feature vector from the extracted segments was calculated and used to train a Conditional Random Field (CRF) classifier. A precision of 91.0% and a recall of 81.0% were achieved by the proposed system.

Junker et al. (2008) proposed a solution for spotting and recognising a set of four dietary gestures (cutlery, drink, spoon and hand-held) using 5 inertial sensors (two on each arm and one on the trunk). To do so, a two-stage spotting approach was first developed by the combination of a sliding-window and bottom-up (SWAB) and a FSS. Once potential segments were identified by the two-stage gesture spotting technique, a Hidden Markov Model (HMM) was used to classify the gestures, achieving a precision of 73.0% and a recall of 79.0%.

Dong et al. (2014) presented a two-stage approach for spotting periods of eating using data from a single wrist-worn inertial sensor. First, a custom-peak algorithm based on wrist motion energy was used as a mean of segmentation. The intuition behind this approach is that periods of eating are preceded and followed by periods of higher wrist kinetic energy. Once the potential periods of eating were identified, a range of four features was extracted across those periods to train a naive Bayes classifier, by which a classification recall of 81.0% was achieved.

2.3. Justification and Motivation

Various limitations are found in the reviewed articles. First, some studies rely on extremely constrained environments. For example, in the published paper related to drinking spotting by Chen et al. (2017), chairs were height-adjusted to individuals. In addition, individuals were told how to perform the drinking actions and the data set only included drinking gestures. The work by Tseng et al. (2018) on recognising door opening gestures makes no mention of a 'Null' class. The 'Null' class in a gesture recognition problem is the class composed by gestures outside the studied gesture set. This fact implies the data set was built only with door opening gestures. In research conducted by Xu et al. (2012) on the recognition of a set of seven hand gestures, participants were told to hold the accelerometer horizontally during the experiments. To our view, gesture spotting and recognition should be undertaken in realistic scenarios where participants perform the studied actions freely. In addition, the resultant data sets should include a reasonable 'Null' class with a range of additional gestures outside the sought gesture set.

Second, the classification performance of gesture spotting and recognition systems under unconstrained environments still lies far away from that in HAR systems. The main reason is that given the sparsity of gestures and the resultant difficulty at developing accurate adaptive segmentation techniques, a great number of true positives are missing at the segmentation (spotting) step. For example, the work presented by Junker et al. (2008) resulted in a recall of 80% at the segmentation stage. The results in (Amft et al., 2010) indicate an 84% recall at spotting drinking gestures.

Besides, various fluid and food intake tracking solutions proposed have been found to require the use of several sensor units (Junker et al., 2008; Ortega-Anderez et al., 2018a). This could make such solutions be excessively intrusive for a daily use. Overall, the drawbacks above suggest there are still many open challenges in gesture spotting and recognition. The mitigation of the above drawbacks has motivated the development of the fluid and food intake tracking system presented in this work.

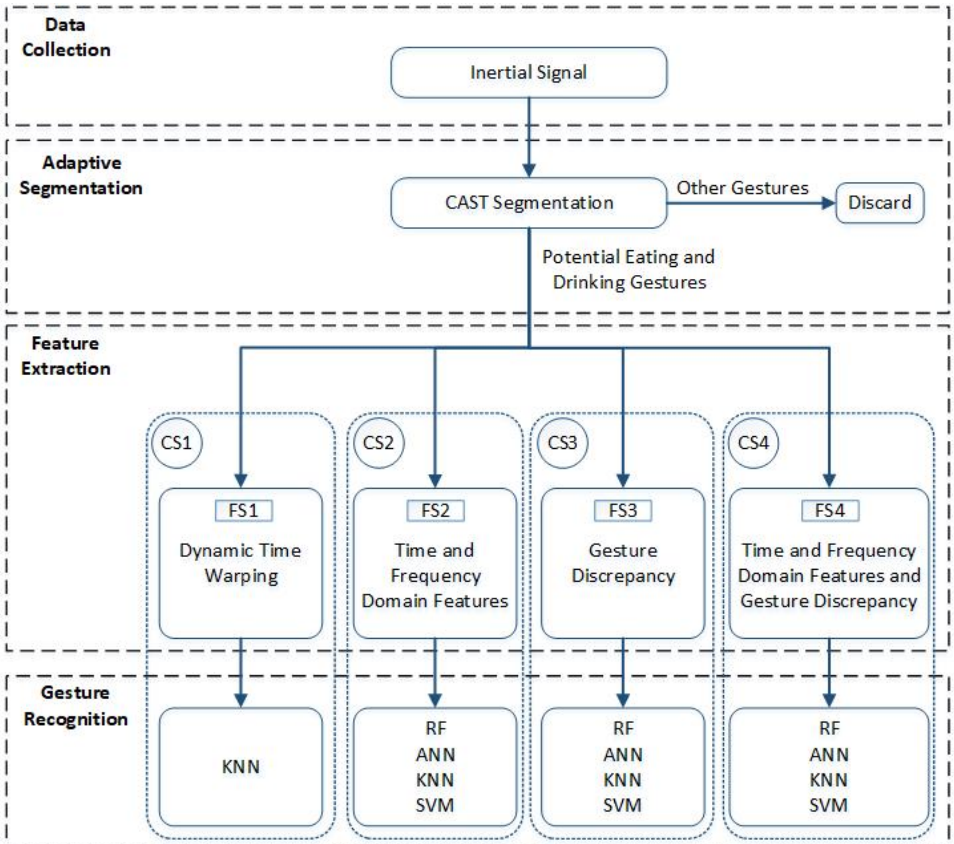

Figure 1: Schematic diagram of the proposed methodology to spot and recognise eating and drinking gestures.

3. Methodology

This section presents the steps undertaken to develop the fluid and food intake system. The different stages of the proposed system are illustrated in Figure 1. First, potential segments containing an eating or a drinking gesture are identified using a Crossings-based Adaptive Segmentation Technique (CAST). A posteriori, four different Computational Solutions (CS) are proposed as follows:

CS1:- Dynamic Time Warping (DTW) Distance + K-Nearest Neighbours (KNN)

CS2:- Feature set + range of state-of-the-art classification models

CS3:- Gesture discrepancy + range of state-of-the-art classification models

CS4:- Feature set+ gesture discrepancy + range of state-of-the-art classification models

The above computational solutions are used to methodically justify the addition of a gesture discrepancy measure to long-established features used in previous HAR work. In CS1, the use of Dynamic Time Warping is evaluated. Given the challenging gesture set proposed, modest results are expected from CS1, however, this serves as a basis to justify a further exploration of DTW as feature descriptor as well as to validate the CAST on the identification of eating and drinking gestures. CS2 explores the use of long-established features employed in previous HAR applications for the recognition of eating and drinking gestures. CS3 introduces the use of gesture discrepancy as feature descriptor. Ultimately, CS4 evaluates the combination of the long-established range of features with the gesture discrepancy measure proposed. The achievement of an improvement on the classification performance of CS4 as compared to previous computational solutions will justify the addition of the gesture discrepancy measure in future activity and gesture recognition work. The performance of the proposed computational solutions was studied across three different gesture sets as follows:

2-Class: Null, Drinking or Eating

3-Class: Null, Drinking, Eating

5-Class: Null, Drinking, Spoon, Fork, Hand

where 'Null' refers to any gesture within the 'Null' class. That is, any gesture which is not an eating or a drinking gesture.

3.1. Experimental Procedure

The fluid and food intake system was evaluated on a total of 0.93 hours of data, which included a total of 226 relevant eating and drinking gestures. Considering an approximated duration of 2 seconds per eating or drinking gesture, the data set was composed of 0.125 hours of relevant gestures and a total of 0.805 hours of 'Null' class. Six participants were asked to wear a wrist-mounted Meta Motion R inertial unit (Mbientlab, 2018) on their dominant hand. The inertial unit was programmed to provide tri-axial accelerometer and tri-axial gyroscope data at 25 Hz.

The scenario proposed was designed to provide data from the participants having a full meal, while facing the challenges one would expect to encounter in real life. First, before starting the meal, participants were asked to act freely within the house for an unlimited time. This included activities like walking, walking upstairs, hands washing or chatting to other participants. This ensured the data contained a reasonable 'Null' class. The resultant temporal ratio Null/Relevant was 6.44. Second, the utensils differed for some participants, e.g. some participants used a glass to drink water while other participants used a mug. Third, a great variety of dishes was provided to ensure various utensils were utilised. Concretely, participants were provided with crisps, soup, chicken breast and cake. Fourth, participants were not given any instructions as to how or when to eat or drink the different dishes. In addition, a left-handed participant took part in the experiment, introducing more variability into the data set.

3.2. Signal Processing

In order to minimise the computational cost of the system, a limited initial pre-processing was carried out on the raw inertial signals. The directions of the y-axis accelerometer and the z-axis gyroscope for the left-handed participant were shifted 180° , given the opposite orientation of these two signals when the sensor unit is worn on the left hand.

3.3. Signal Segmentation and Gesture Spotting

Spotting sporadic gestures requires the implementation of an adaptive segmentation technique, whereby the extracted segments are determined by changes in the signals themselves. Three main constraints are identified on the segmentation of eating and drinking gestures. First, an eating or a drinking gesture can exhibit different length in time. This implies the segmentation has to adapt to such variability to extract the fundamental characteristics of each gesture. Second, segments need to be adjusted as new incoming data is received. Third, the impact of the well known long-term drift of gyroscopes must be mitigated to avoid inaccurate measurements.

The concept of a segmentation technique to spot eating and drinking gestures while overcoming the above constraints was presented in our earlier work (Anderez et al., 2018). Given its good performance in a narrow data set, the segmentation technique is further validated in this work using a greater number of participants as well as an extensive 'Null' class.

The crossings of two moving averages are used to determine the potential segments containing an eating or a drinking gesture. Given its functionality,

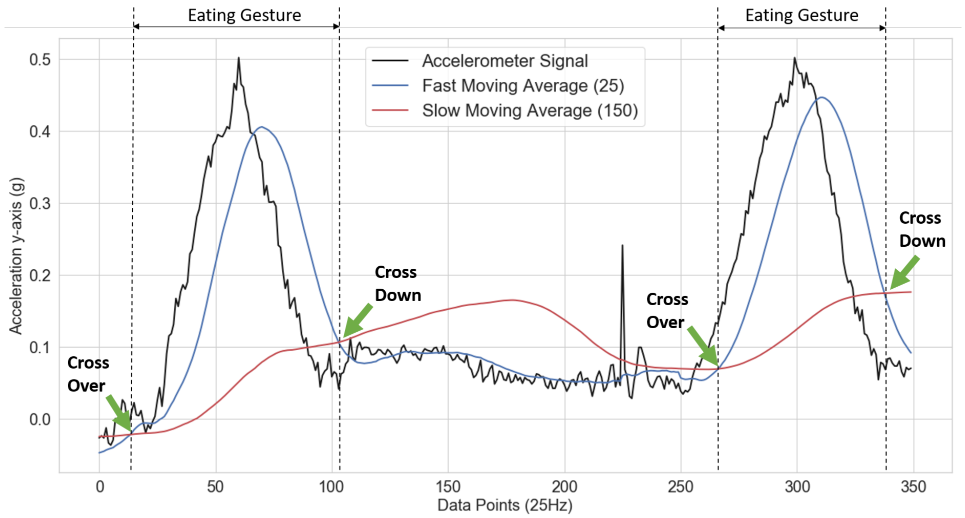

Figure 2: Crossings-based adaptive segmentation technique applied to a sample signal with two consecutive eating gestures.

the technique is referred to as Crossings-based Adaptive Segmentation Technique (CAST). The intuition behind CAST is the sequence of hand motions involved in an eating or a drinking gesture. First, the corresponding tool (e.g. a glass) is taken to the mouth. This is followed by a movement of the hand back to the rest position. Such a sequence of motions leads to a rapid increase on the fast moving average when food or a drink are taken to the mouth, crossing over the slow moving average. A hand movement to the rest position will follow, producing a rapid decrease on the fast moving average and the consequent cross down of the slow moving average. This is illustrated in Figure 2, where the segmentation of two consecutive eating gestures using the CAST is shown.

The CAST can be explained as follows. Consider a signal $y[t]$. The moving average $\bar{y}[t]$ of $y[t]$ is defined as:

$$\bar{y}[t] = \frac{1}{n}\sum_{i=0}^{n-1} y[t-i] \tag{1}$$

where n is the number of data points over which the moving average is calculated. Two moving averages $\bar{y}_1[t]$ and $\bar{y}_2[t]$ are calculated over the intervals T_1 and T_2 respectively, such that $T_2 > T_1$. If $y[t]$ increases, the CAST moving average $\bar{y}_1[t]$ will react faster to that increase on $y[t]$. The same will happen when a decrease is applied on $y[t]$.

The optimal values for T_1 and T_2 as well as the accelerometer axis (y-axis) over which the moving averages were calculated, were experimentally determined in our previous work (Anderez et al., 2018). Given that more computational intensive tools are to be applied after the segmentation step, T_1 and T_2 were calculated so as to optimise the classification recall. The resultant values for T_1 and T_2 on the y-axis accelerometer signal are $n = 25$ and $n = 150$ respectively. Considering the sampling frequency of 25 Hz, $\bar{y}_1[t]$ and $\bar{y}_2[t]$ are the moving averages of the acceleration on the y-axis over 1 second and 6 seconds respectively.

Overall, the CAST overcomes the challenges exposed at the beginning of this section. First, it adapts to the nature of the signal, since both moving averages $\bar{y}_1[t]$ and $\bar{y}_2[t]$ react in consonance with the original signal $y[t]$. Second, it deals with the different length of gestures successfully. For instance, in a long drinking gesture, the decrease on the fast moving average $\bar{y}_1[t]$ after the glass has been taken to the mouth is slower than in a short gesture, since the hand movement that causes the decrease on $y[t]$ and therefore on $\bar{y}_1[t]$ is deferred. Third, CAST can be used real-time since it adapts to new incoming data adjusting the moving averages accordingly. Fourth, given that the signal utilised for the segmentation is the accelerometer y-axis, this technique avoids the undesired impact of the gyroscope long-term drift, since the crossings between the moving averages can act as a gyroscope trigger.

3.4. Gesture Recognition

Once the potential segments containing an eating or drinking gesture are identified, gesture recognition is tackled as a classification problem. For the four proposed computational solutions (CS1, CS2, CS3, CS4), four different feature sets are employed as follows:

FS1:- Dynamic Time Warping

FS2:- Feature Vector

FS3:- Gesture Discrepancy

FS4:- Feature Vector and Gesture Discrepancy

More detail about these approaches is provided in the following sections.

3.4.1. Dynamic Time Warping

Let $q[t] = [q_1, q_2, ...q_n]$ and $s[t] = [s_1, s_2, ...s_n]$ be two temporal sequences with values at every time instant t=[1,...,n]. The distance $d(q, s)$ is typically measured as their Euclidean distance:

$$d(q,s) = \sqrt{\sum_{t=1}^{n}(q[t]-s[t])^2} \tag{2}$$

Two major constraints are found on the use of the Euclidean distance on time-dependent sequences: 1) the length of the sequences must be equal i.e. $|q| = |s|$, 2) it does not consider the temporal distortion that may exist between q and s, since it measures the vertical distance between pairs of points according to their indexes at their respective sequences.

To overcome the above constraints, the optimal alignment between time-dependent sequences is calculated with the use of DTW (Sakoe and Chiba, 1978). The alignment can be explained as follows: Considering the two temporal sequences q and s of respective lengths $|q|$ and $|s|$, DTW finds a mapping path $\{(p_1, r_1), ..., (p_j, r_j)\}$ such that the distance on the mapping path $\sum_{i=1}^{j} |x(p_i) - y(r_i)|$ is minimised with the following two constraints:

$$\begin{cases} \text{Anchored beginning: } (p_1, r_1) = (1, 1) \\ \text{Anchored end: } (p_j, r_j) = (|q|, |s|) \end{cases} \tag{3}$$

The DTW distance between q and s is then calculated as the cost of the optimal alignment as follows:

$$D_{i,j} := D(q(i), s(j)) + min \left\{ \begin{array}{l} D(i-1, j) \\ D(i-1, j-1) \\ D(i, j-1) \end{array} \right\} \tag{4}$$

where $D(q(i) - s(j))$ is calculated as the Euclidean distance.

Figure 3 illustrates the use of the Euclidean distance and DTW to measure the similarity between temporal sequences. It can be seen that DTW overcomes the drawbacks encountered when using the Euclidean distance. First, it can measure the distance between signals with different lengths, since one point of the sequence q can be aligned to more than one point of the sequence s and vice versa. Second, the alignment performed is able to capture the temporal distortion between the signals.

Ultimately, the DTW distance is used for gesture recognition. To do so, a K-Nearest Neighbors (KNN) classification model is employed, whereby unseen segments are assigned to the most common class among its k-nearest neighbours, with DTW being the distance measure between the different segments using the y-axis accelerometer signal.

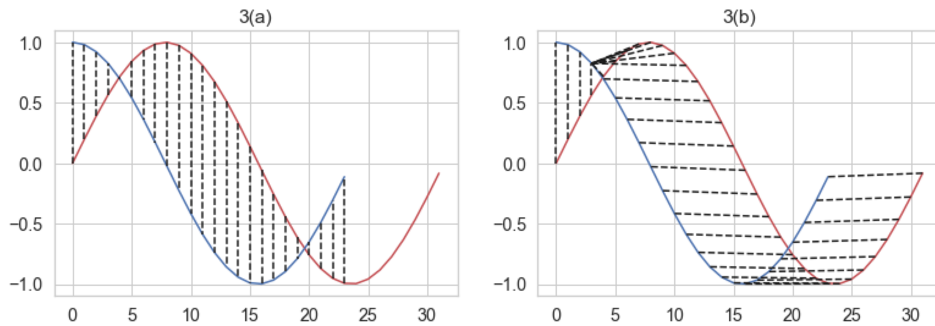

Figure 3: Difference between the Euclidean distance and the DTW distance of two signals; a) Euclidean distance, b) DTW distance: the distance between two points is calculated as their Euclidean distance (vertical distance) after alignment.

3.4.2. Feature Vector

This computational solution makes use of long-established set of features used within the field of HAR (Ortega-Anderez et al., 2018b; Ravi et al., 2005; Casale et al., 2011; Bayat et al., 2014). The feature vector has been conscientiously culled to provide a knowledgeable description of the data regarding a wide array of signal characteristics. These include measures of central tendency, periodicity, dispersion, changes in direction, frequency distribution and magnitude area. The range of features proposed was calculated over the medio-lateral a_x, antero-posterior a_y and vertical a_z acceleration corresponding to the tri-axial accelerometer readings, as well as on the yaw g_x, roll g_y and pitch g_z corresponding to the tri-axial gyroscope readings across the potential segments. On top of the above, the duration of each segment is also incorporated into the feature set. The resultant dimensionality of the feature vector proposed is $n = 85$.

3.4.3. Gesture Discrepancy

This computational solution introduces a gesture discrepancy measure as a mean of a signal descriptor. To do so, the Soft-DTW differentiable loss function proposed by Cuturi and Blondel (2017) is employed to calculate a gesture barycenter for each of the gestures within the different proposed gesture sets through a minimisation problem. Further, the DTW distances to each of the calculated barycenters are used to build the feature set.

Let's consider multivariate time series of varying length taking values in $\Omega \subset \mathbb{R}^p$, whereby they are represented as a matrix of p rows. Soft-DTW unifies the original DTW discrepancy (Sakoe and Chiba, 1978) and the Global Alignment Kernel (GAK) proposed by Cuturi et al. (2007), both used to compare two time series $x[t] = [x_1, x_2, ..., x_n] \in \mathrm{I\!R}^{\mathrm{p \times n}}$ and $y[t] = [y_1, y_2, ..., y_m] \in \mathrm{I\!R}^{\mathrm{p \times m}}$.

Given the cost matrix $\Delta(x, y) := [\delta(x_i, y_j]_{ij} \in \mathrm{I\!R}^{\mathrm{n \times m}}$ and the set of binary alignments matrices $A_{n,m} \subset \{0, 1\}$, the inner product $\langle A, \Delta(x, y)\rangle$ of the cost matrix with an alignment matrix A in $A_{n,m}$ gives the score of A. DTW and GAK consider respectively the cost of all possible alignment matrices as follows:

$$\begin{aligned} DTW(x,y) &:= \min_{A \in A_{n,m}} \langle A, \Delta(x,y)\rangle, & (5)\\ \kappa^{\gamma}_{GA}(x,y) &:= \sum_{A \epsilon A_{n,m}} e^{-\langle A, \Delta(x,y)\rangle/\gamma} & (6)\end{aligned}$$

From the equations above, a unified algorithm can be formulated as:

$$min^{\gamma}\{a_1, ..., a_n\} := \begin{cases} min_i \leq_n a_i, & \gamma = 0, \\ -\gamma \ \log \sum_{i=1}^{n} e^{-a_i/\gamma}, & \gamma > 0. \end{cases} \qquad (7)$$

where γ is a smoothing parameter taking values in $\mathrm{I\!R}_{\geq 0}$. Given the above, γ-soft-DTW can be defined as:

$$dtw_{\gamma}(x,y) := min^{\gamma}\{\langle A, \Delta(x,y)\rangle, A \in A_{n,m}.\} \qquad (8)$$

Therefore, the original DTW score is recovered when γ is set to 0 and $dtw_{\gamma} = -\gamma log k^{\gamma}_{GA}$ when $\gamma > 0$.

Ultimately, given a group of N time series y_1, ..., y_N, that is, N matrices of p rows and varying number of columns, m_1, ..., m_N, the interest is to define a single barycenter time series x for that group under a set of normalised weights $\lambda_1, ..., \lambda_N \in \mathrm{I\!R}_+$ such that $\sum_{i=1}^{N} \lambda_i = 1$. Thus, the barycenter is calculated by approximately solving the following problem:

$$\min_{x \in \mathrm{I\!R}^{\mathrm{pxn}}} \sum_{i=1}^{N} \frac{\lambda_i}{m_i} dtw_{\gamma}(x, y_i) \qquad (9)$$

where it is assumed that x has fixed length n. Given the proposed gesture sets G_1, G_2, G_3 of respective lengths $|G_1|$, $|G_2|$, $|G_3|$, a barycenter was calculated for each of the gestures different from the 'Null' class $g_1, ..., g_{|G_i|-1}$ within $G_1, ...G_3$, for each of the experiment participants $P_1, ..., P_6$, for each of the time series in $a_x, a_y, a_z, g_x, g_y, g_z$, corresponding to the tri-axial accelerometer and the tri-axial gyroscope readings. A posteriori, the DTW

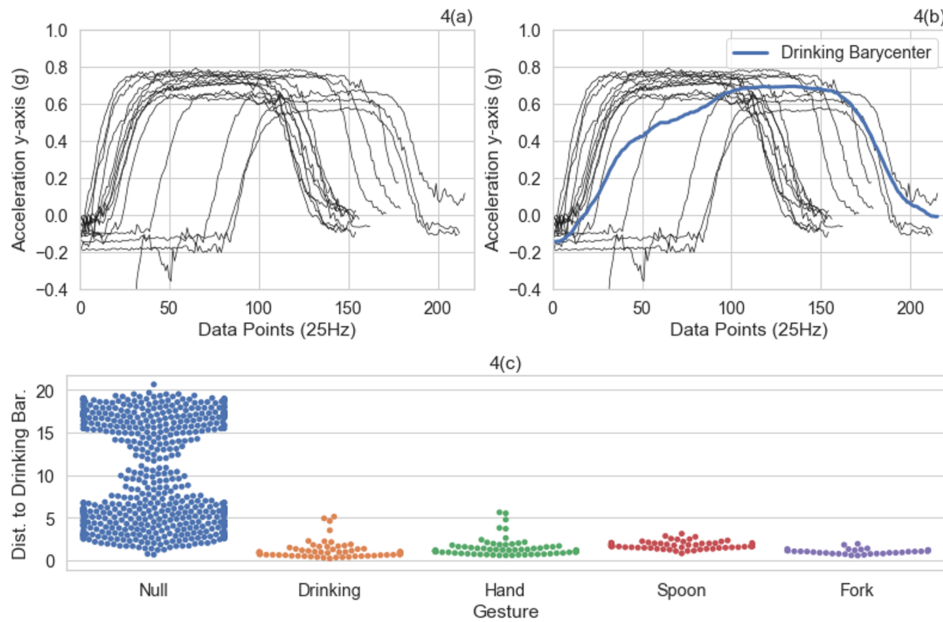

Figure 4: Distance to the drinking barycenter (accelerometer y-axis) of one of the experiment participants; a) Different drinking gestures from the participant, b) Calculation of the participant's drinking barycenter, c) Distribution of distances to the barycenter in (b) across the gestures from the rest of the participants.

distances to the set of calculated barycenters were computed and further used as feature descriptors.

Two pictorial examples of the calculation of a gesture barycenter and the distribution of the DTW distances to the calculated gesture barycenter across the different gestures are shown in Figure 4 and 5. Further, the bi-dimensional distribution of the DTW distances to the barycenters exposed in Figure 4 and Figure 5 across the different gestures is shown in Figure 6 for illustration purposes. As a result of the above distance computations, the resultant dimensionality of the feature vector was $n = 36$ for the 2-class classification problem, $n = 72$ for the 3-class classification problem and $n = 144$ for the 5-class classification problem.

3.4.4. Feature Vector and Gesture Discrepancy

Feature set FS4 is a combination of the features introduced in FS2 and FS3 to evaluate whether the addition of a gesture discrepancy measure to long-established feature vectors improves the recognition rate of the system. The combination of both feature sets gives a resultant dimensionality

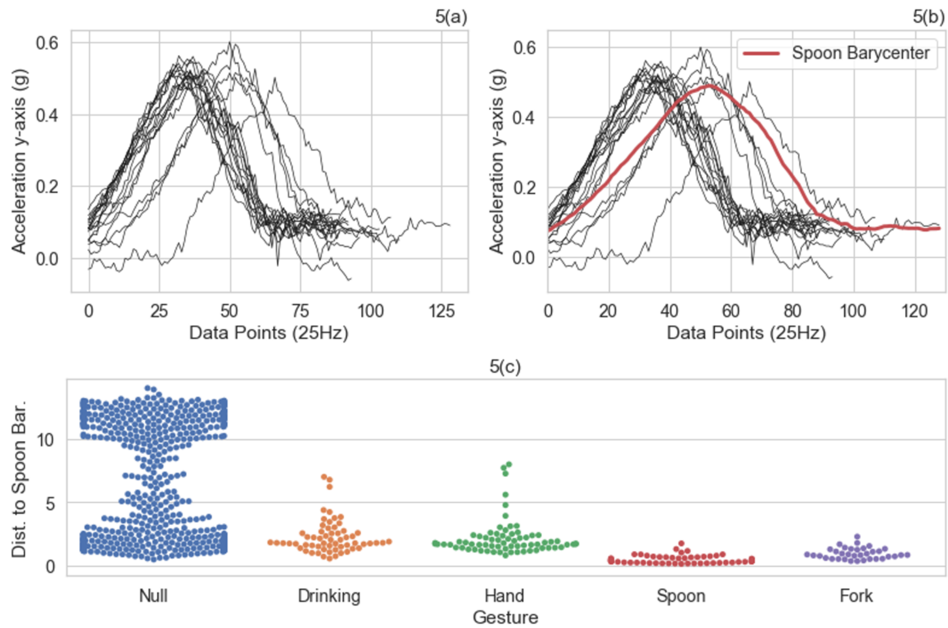

Figure 5: Distance to the spoon barycenter (accelerometer y-axis) of one of the experiment participants; a) Different spoon gestures from the participant, b) Calculation of the participant's spoon barycenter, c) Distribution of distances to the barycenter in (b) across the gestures from the rest of the participants.

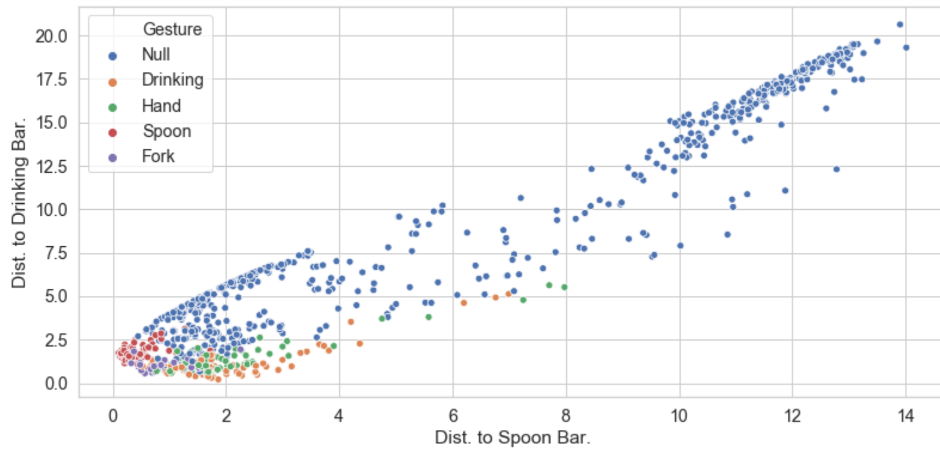

Figure 6: Bi-dimensional distribution of the DTW distances to the drinking and spoon barycenters of one of the participants across the gestures from the rest of the participants.

of $n = 121$ for the 2-class classification problem, $n = 157$ for the 3-class classification problem and $n = 229$ for the 5-class classification problem.

3.5. Evaluation

A leave-one-out cross-validation strategy was employed for evaluation purposes. That is, a different participant was used as the test set on each of the cross-validation steps. For the feature sets including the gesture discrepancy measure (FS3 and FS4), the distances to the barycenters of the participant used as the test set on each cross-validation cycle were removed from the feature set.

Given the special structure of the feature set FS1 proposed in CS1, its performance was evaluated by the employment of a KNN classifier. The rest of the computational solutions were evaluated across a range of state-of-the-art classification models, including Artificial Neural Networks (ANN), Support Vector Machines, Random Forest (RF) as well as KNN.

4. Results

This section presents the results obtained by the implementation of the presented methodology. Section 4.1 shows the performance of the proposed CAST segmentation technique at spotting potential eating and drinking gestures. Section 4.2 presents the results achieved by the different computational solutions proposed for gesture recognition. A discussion upon the results obtained is given in Section 4.3.

4.1. Gesture Spotting

As explained in Section 3, the first step on the development of the proposed fluid and food intake recognition system is to spot potential segments containing an eating or a drinking gesture. This was tackled by the implementation of CAST, which uses the crosses between two moving averages to spot those potential segments. A pictorial example for one of the experiment participants is shown in Figure 7.

Given that more computational intensive tools are to be applied at the gesture recognition step, the aim at this preliminary spotting step was to optimise the classification recall, that is, minimising the number of 'False Negatives', in this case eating or drinking gestures classified as pertain to the 'Null' class. The achieved spotting results shown in Figure 8 outline an average precision of 29% and an average recall of 100%, showing that this is successfully achieved by the segmentation technique proposed.

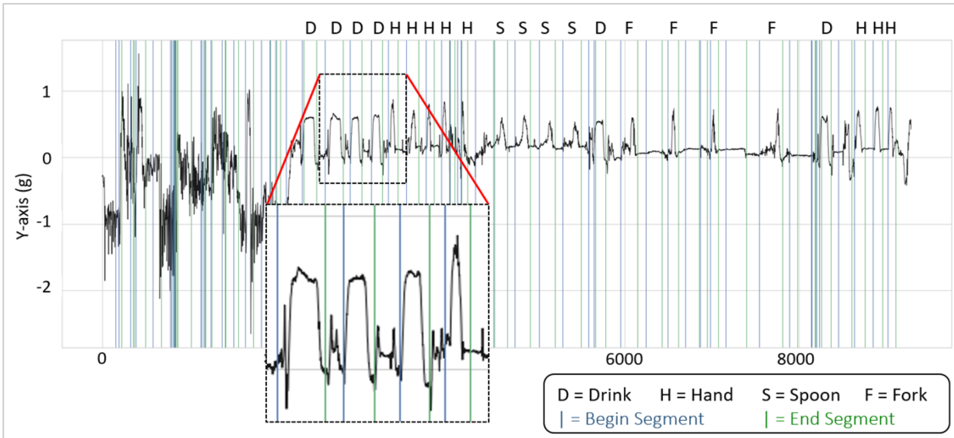

Figure 7: Performance of the Crossings-based Adaptive Segmentation Technique for one of the experiment participants.

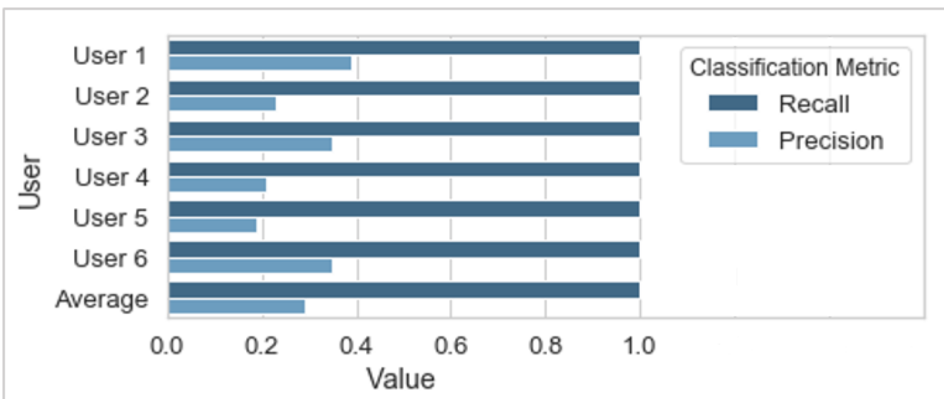

Figure 8: Spotting performance of CAST.

4.2. Gesture Recognition

After the segments potentially containing an eating or a drinking gesture are identified, gesture recognition comes into place. Four different computational solutions were proposed across three different experiments. A comprehensive study upon the performance of the implemented computational solutions was performed and the best results obtained in each of the three experiments are presented below:

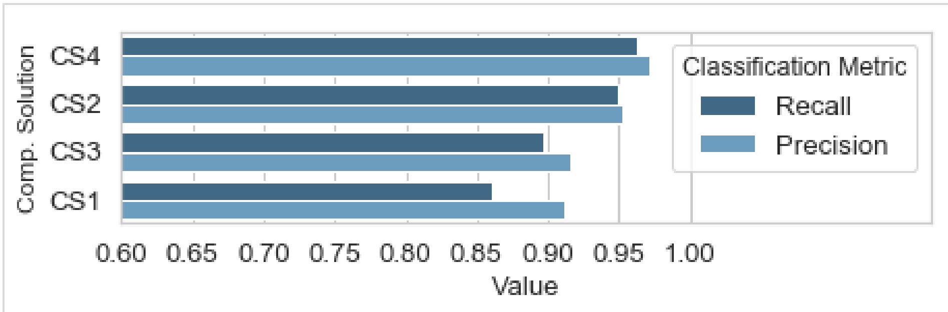

Figure 9: Classification performance of the four computational solutions proposed on the 2-class classification problem.

4.2.1. Experiment 1: 2-Class Classification Problem

In this experiment eating and drinking gestures are grouped together and classified against the 'Null' class. The results presented in Table 1 outline an average per-class classification accuracy of 97.4%, a precision of 97.2% and a recall of 96.3% using a Random Forest Classifier on the feature set composed of the proposed range of features alongside the gesture discrepancy measure (FS4). Figure 9 shows the performance of the four computational solutions proposed.

4.2.2. Experiment 2: 3-Class Classification Problem

This experiment aims at the recognition of eating and drinking gestures separately. This was tackled as a 3-Class classification problem, with the classes being 'Null', 'Drinking' and 'Eating'. The classification metrics shown in Table 2 report an average per-class classification accuracy of 98.2%, a precision of 95.7% and a recall of 95.0%. The reported results were achieved using an Artificial Neural Network (ANN) on the feature set (FS4). The classification performance achieved by each of the computational solutions proposed are shown in Figure 10.

Table 1: Classification metrics for the 2-class classification problem using CS4 with RF

	Accuracy (%)	Precision (%)	Recall (%)
Null	97.4	97.6	98.8
Eating or Drinking	97.4	96.8	93.8
Average per-class	97.4	97.2	96.3

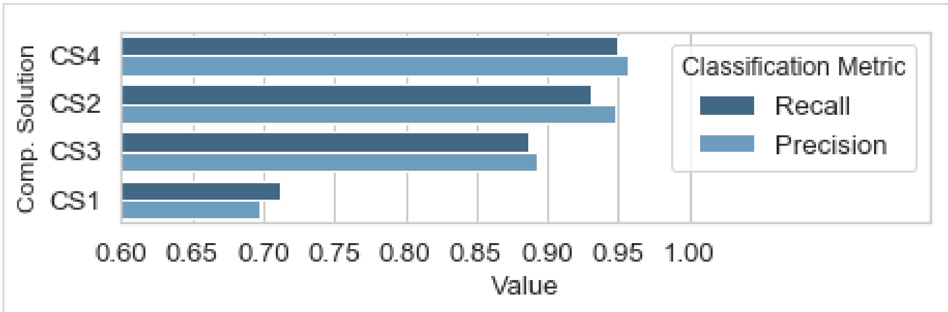

Figure 10: Classification performance of the four computational solutions proposed on the 3-class classification problem.

4.2.3. Experiment 3: 5-Class Classification Problem

In this experiment, the 'Eating' class is further divided into 3 different classes ('Spoon', 'Fork' and 'Hand'), leading to a 5-class classification problem, with the classes being 'Null', 'Drinking', 'Spoon', 'Fork' and 'Hand'. The classification metrics in Table 3 report an average per-class classification accuracy of 97.8%, a precision of 88.7% and a recall of 85.8%, using an ANN on the feature set (FS4). The classification performance of the four computational solutions are shown in Figure 11.

4.3. Discussion

The methodology proposed addressed the problem of spotting and recognising fluid and food intake gestures with the use of a single wrist-worn inertial unit. At the spotting step, the aim was to minimise the number of false negatives. This was based on the fact that more computational intensive tools, namely classification models, were to be applied at the recognition step. The novel adaptive segmentation technique proposed (CAST) cor-

Table 2: Classification metrics for the 3-class classification problem using CS4 with an ANN

	Accuracy (%)	Precision (%)	Recall (%)
Null	97.9	98.1	99.0
Drinking	99.0	93.3	93.3
Eating	97.7	95.7	92.8
Average per-class	98.2	95.7	95.0

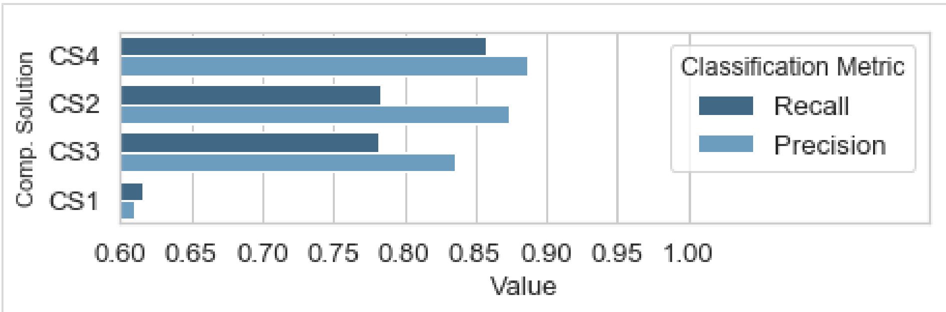

Figure 11: Classification performance of the four computational solutions proposed on the 5-class classification problem.

rectly identified all the drinking and eating gestures. Although the average precision was considerably low (29%), a 100% recall was achieved, indicating the aim proposed was successfully accomplished. Further, a range of four different feature sets was proposed for gesture recognition. As expected, the addition of the gesture discrepancy measure as a feature descriptor consistently improved the classification performance of the system across the three experiments proposed. This can be explained by the fact that the signal alignment performed through the use of DTW accounts for the gestures intra-person and inter-person temporal distortion, thus adding crucial information to long-established feature sets used in previous activity or gesture recognition problems.

Given the great variety of gestures involved in an eating activity, previous research has varied the way of tackling its recognition. To fairly evaluate the proposed methodology against previous similar work, the performance

Table 3: Classification metrics for the 5-class classification problem using CS4 with an ANN

	Accuracy (%)	Precision (%)	Recall (%)
Null	97.0	97.2	98.8
Drinking	98.6	90.2	91.7
Spoon	99.0	96.5	90.2
Fork	97.6	75.0	68.6
Hand	97.0	84.4	80.0
Average per-class	97.8	88.7	85.8

of the recognition of drinking gestures is considered. As shown in Table 4, the proposed methodology in this paper out-performs previous gesture recognition work including both the spotting and recognition stages. Only the classification precision achieved in (Chen et al., 2017) shows a higher value. Although, Chen et al. (2017) did not include the spotting step. That is, the data set lacked a 'Null' class as well as other gestures different from drinking gestures. As a result of this, the precision and recall metrics were clearly boosted, since the experiment proposed was evidently biased towards the recognition of drinking gestures.

5. Conclusions and Future Work

This paper presented a novel intelligent system for the spotting and recognition of eating and drinking gestures in a free-living scenario using a single wrist-worn inertial unit as a mean of data collection.

From the methodological view-point, two major contributions are made. 1) The novel adaptive segmentation technique proposed (CAST) overcomes the two major drawbacks observed in previous similar work. On the one hand, as contrary to previous adaptive segmentation techniques in the field (Junker et al., 2008), CAST achieves a 100% spotting recall, thus preventing the system from having false negatives at the preliminary spotting phase. This is crucial since the errors at the spotting phase will propagate to the recognition phase, therefore limiting the performance of the whole intelligent system. Other systems have opted for the employment of sliding windows (Feng and Duarte, 2019; Ronao and Cho, 2016; Serrano et al., 2017), where the window length is typically estimated as a hyper-parameter of the clas-

Table 4: Comparison of the proposed approach with previous work on the recognition of drinking gestures.

Method	Sensor Units	Spot.	Recog.	Accuracy	Precision (%)	Recall (%)
Junker et al. (2008)	5	✓	✓	-	88.0	83.0
Amft et al. (2010)	1	✓	✓	-	84.0	90.0
Chen et al. (2017)	1	X	✓	-	96.5	91.3
Schiboni and Amft (2018)	1	✓	✓	-	75.2	76.1
Serrano et al. (2017)	4	✓	✓	-	82.28	84.42
Ramos-Garcia et al. (2014)	1	✓	✓	86.5	-	-
Proposed Approach (3-class)	1	✓	✓	99.0	93.3	93.3
Proposed Approach (5-class)	1	✓	✓	98.6	90.2	91.7

sification problem. This approach fails to adapt to the duration variability present in drinking and eating gestures, therefore biasing the system towards the most common duration found in the corresponding experimental dataset. 2) While long-established feature sets incorporate shallow (normally statistical) characteristics of the signals, the Soft-DTW based gesture discrepancy measure proposed accounts for the intra and inter-personal temporal distortion at performing eating and drinking gestures. As shown by the results obtained, this clearly offers an advantage to our system, which has seen a consistent improvement across the three experiments proposed. In terms of the average per-class classification recall, the addition of the gesture discrepancy measure improves the performance of the system from 0.950 to 0.963, from 0.931 to 0.95 and from 0.783 to 0.858 for the 2-class, 3-class and 5-class classifications problems respectively. Regarding the average per-class classification precision, the performance improvements seen are from 0.952 to 0.972, from 0.948 to 0.957 and from 0.874 to 0.887.

From a technical perspective, the intelligent system proposed shows two major advantages. 1) It overcomes the occlusion issues and privacy concerns of systems employing video cameras (Chen and Shen, 2017) and depth sensors (Kim et al., 2019) in a home environment, while providing more intrinsic information about the subject than systems employing ambient sensors. Although ambient systems have shown good results at detecting simple activities such as sleeping or toileting, those results are significantly worsened when attempting the recognition of complex activities like eating (Wen and Zhong, 2015). 2) It significantly out-performs previous intelligent systems on the recognition of eating and drinking gestures with wearable devices.

This study has potential limitations. The way eating and drinking gestures are performed does not vary significantly between different subjects, however, the performance of the system on participants with functional limitations such as patients suffering from Parkinson's disease or stroke patients, could potentially be affected. In addition, this study assumes eating and drinking are performed from a sitting position. Ultimately, it should be stated subjects sometimes use their non-dominant hand to eat or drink, however, these actions are normally performed with the dominant hand. Our priority here is fostering the usability of the system by avoiding the undesired obtrusiveness found in systems employing more than one sensing unit.

Despite these limitations, the aforementioned contributions not only imply a great step forward towards the development of an intelligent dietary tracking system, but a valuable input in the form of an adaptable and flexible novel segmentation technique (CAST) and the introduction of a feature descriptor based on gesture discrepancy for their employment in future work on intelligent systems for activity and gesture recognition as well.

Future work will be focused on four different aspects: 1) Although the proposed system shows positive results, short-term future efforts will be focused on the search of possibilities to further improve the performance of the system. To do so, deep learning models, namely Long Short-Term Memory (LSTMs) and Convolutional Neural Networks (CNNs) will be explored alongside the CAST segmentation technique and compared to the performance of the current computational solution. Transfer learning in the form of inductive learning will be proposed for future intelligent systems on activity and gesture recognition. 2) Given the local dependency characteristics of temporal sequences and the translation invariant nature of human gestures, these will be explored by the use of multi-input deep neural networks, whereby the current future set will be complemented with the patterns encountered through the different convolutional layers of a CNN. 3) Further to the search for computational solutions to improve the current gesture recognition rate, the development of an intelligent system for the detection of meal periods based on the occurrence of the gestures across time will be the next step of this work. 4) The ultimate effort of the work will be directed to the investigation and implementation of trend analysis techniques to develop an intelligent system for the identification of anomalies on the dietary behaviour of individuals.

References

Amft, O., Bannach, D., Pirkl, G., Kreil, M., Lukowicz, P., 2010. Towards wearable sensing-based assessment of fluid intake, in: 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM), pp. 298–303.

Amft, O., Kusserow, M., Tröster, G., 2007. Probabilistic parsing of dietary activity events, in: 4th International Workshop on Wearable and Implantable Body Sensor Networks, Springer. pp. 242–247.

Anderez, D.O., Lotfi, A., Langensiepen, C., 2018. A novel crossings-based segmentation approach for gesture recognition, in: Advances in computational intelligence systems: contributions presented at the 18th UK Workshop on Computational Intelligence, Springer. pp. 383–391.

Bayat, A., Pomplun, M., Tran, D.A., 2014. A study on human activity recognition using accelerometer data from smartphones. Procedia Computer Science 34, 450–457.

Billiet, L., Swinnen, T.W., Westhovens, R., de Vlam, K., Van Huffel, S., 2016. Accelerometry-based activity recognition and assessment in rheumatic and musculoskeletal diseases. Sensors 16, 2151.

C. rush, E., Valencia, M.E., Plank, L.D., 2008. Validation of a 7-day physical activity diary against doubly-labelled water. Annals of Human Biology 35, 416–421.

Casale, P., Pujol, O., Radeva, P., 2011. Human activity recognition from accelerometer data using a wearable device. Pattern Recognition and Image Analysis , 289–296.

Chen, L.H., Liu, K.C., Hsieh, C.Y., Chan, C.T., 2017. Drinking gesture spotting and identification using single wrist-worn inertial sensor, in: IEEE International Conference on Applied System Innovation (ICASI), pp. 299–302.

Chen, Y., Shen, C., 2017. Performance analysis of smartphone-sensor behavior for human activity recognition. Ieee Access 5, 3095–3110.

Chernbumroong, S., Cang, S., Atkins, A., Yu, H., 2013. Elderly activities recognition and classification for applications in assisted living. Expert Systems with Applications 40, 1662–1674.

Cuturi, M., Blondel, M., 2017. Soft-DTW: a differentiable loss function for time-series, in: Precup, D., Teh, Y.W. (Eds.), Proceedings of the 34th International Conference on Machine Learning, PMLR, International Convention Centre, Sydney, Australia. pp. 894–903.

Cuturi, M., Vert, J.P., Birkenes, O., Matsui, T., 2007. A kernel for time series based on global alignments, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, pp. 413–416.

Dong, Y., Scisco, J., Wilson, M., Muth, E., Hoover, A., 2014. Detecting periods of eating during free-living by tracking wrist motion. IEEE journal of biomedical and health informatics 18, 1253–1260.

Feng, S., Duarte, M.F., 2019. Few-shot learning-based human activity recognition. arXiv preprint arXiv:1903.10416 .

Junker, H., Amft, O., Lukowicz, P., Tröster, G., 2008. Gesture spotting with body-worn inertial sensors to detect user activities. Pattern Recognition 41, 2010–2024.

Kenney, W.L., Chiu, P., 2001. Influence of age on thirst and fluid intake. Medicine and science in sports and exercise 33, 1524–1532.

Keogh, E., Chu, S., Hart, D., Pazzani, M., 2004. Segmenting time series: A survey and novel approach, in: Data mining in time series databases. World Scientific, pp. 1–21.

Kim, J.M., Jeon, M.J., Park, H.K., Bae, S.H., Bang, S.H., Park, Y.T., 2019. An approach for recognition of human's daily living patterns using intention ontology and event calculus. Expert Systems with Applications 132, 256–270.

Kwon, Y., Kang, K., Bae, C., 2014. Unsupervised learning for human activity recognition using smartphone sensors. Expert Systems with Applications 41, 6067–6074.

Lawrence, E.S., Coshall, C., Dundas, R., Stewart, J., Rudd, A.G., Howard, R., Wolfe, C.D., 2001. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. Stroke 32, 1279–1284.

Liu, L., Peng, Y., Wang, S., Liu, M., Huang, Z., 2016. Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors. Information Sciences 340, 41–57.

Lohrmann, C., Dijkstra, A., Dassen, T., 2003. The care dependency scale: an assessment instrument for elderly patients in german hospitals. Geriatric Nursing 24, 40–43.

Lovrić, M., Milanović, M., Stamenković, M., 2014. Algoritmic methods for segmentation of time series: An overview. Journal of Contemporary Economic and Business Issues 1, 31–53.

Mackay, J., 2004. The atlas of heart disease and stroke. volume 5. World Health Organization.

Mahler, M.A., Li, Q., Li, A., 2017. Securehouse: A home security system based on smartphone sensors, in: IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 11–20.

Mannini, A., Sabatini, A.M., 2010. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors 10, 1154–1175.

Mbientlab, 2018. Mbientlab. URL: `https://tinyurl.com/Mbientlab`. accessed: 2018-08-18.

Noor, M.H.M., Salcic, Z., Kevin, I., Wang, K., 2017. Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer. Pervasive and Mobile Computing 38, 41–59.

Ortega-Anderez, D., Lotfi, A., Langensiepen, C., 2018a. A hierarchical approach in food and drink intake recognition using wearable inertial sensors, in: Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference, ACM. pp. 552–557.

Ortega-Anderez, D., Lotfi, A., Langensiepen, C., Appiah, K., 2018b. A multi-level refinement approach towards the classification of quotidian activities using accelerometer data. Journal of Ambient Intelligence and Humanized Computing , 1–12.

Parate, A., Chiu, M.C., Chadowitz, C., Ganesan, D., Kalogerakis, E., 2014. Risq: Recognizing smoking gestures with inertial sensors on a wristband, in: Proceedings of the 12th annual international conference on Mobile systems, applications, and services, ACM. pp. 149–161.

Ramos-Garcia, R.I., Muth, E.R., Gowdy, J.N., Hoover, A.W., 2014. Improving the recognition of eating gestures using intergesture sequential dependencies. IEEE journal of biomedical and health informatics 19, 825–831.

Ravi, N., Dandekar, N., Mysore, P., Littman, M.L., 2005. Activity recognition from accelerometer data, in: Proceedings of the Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, Pittsburgh, Pennsylvania, USA. pp. 1541–1546.

Ronao, C.A., Cho, S.B., 2016. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications 59, 235–244.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing 26, 43–49.

Sawka, M.N., Cheuvront, S.N., Carter III, R., 2005. Human water needs. Nutrition reviews 63, S30–S39.

Schiboni, G., Amft, O., 2018. Sparse natural gesture spotting in free living to monitor drinking with wrist-worn inertial sensors, in: Proceedings of the International Symposium on Wearable Computers, ACM. pp. 140–147.

Schoeller, D.A., Thomas, D., Archer, E., Heymsfield, S.B., Blair, S.N., Goran, M.I., Hill, J.O., Atkinson, R.L., Corkey, B.E., Foreyt, J., et al., 2013. Self-report–based estimates of energy intake offer an inadequate basis for scientific conclusions. The American journal of clinical nutrition 97, 1413–1415.

Serrano, J.I., Lambrecht, S., del Castillo, M.D., Romero, J.P., Benito-León, J., Rocon, E., 2017. Identification of activities of daily living in tremorous patients using inertial sensors. Expert Systems with Applications 83, 40–48.

Smith, B.J., Marshall, A.L., Huang, N., 2005. Screening for physical activity in family practice: evaluation of two brief assessment tools. American journal of preventive medicine 29, 256–264.

Tai, T.M., Jhang, Y.J., Liao, Z.W., Teng, K.C., Hwang, W.J., 2018. Sensor-based continuous hand gesture recognition by long short-term memory. IEEE Sensors Letters 2, 1–4.

Tseng, M.C., Liu, K.C., Hsieh, C.Y., Hsu, S.J., Chan, C.T., 2018. Gesture spotting algorithm for door opening using single wearable sensor, in: International Conference on Applied System Invention (ICASI), IEEE. pp. 854–856.

Wellman, N.S., Friedberg, B., 2002. Causes and consequences of adult obesity: health, social and economic impacts in the united states. Asia Pacific journal of clinical nutrition 11, S705–S709.

Wen, J., Wang, Z., 2017. Learning general model for activity recognition with limited labelled data. Expert Systems with Applications 74, 19–28.

Wen, J., Zhong, M., 2015. Activity discovering and modelling with labelled and unlabelled data in smart environments. Expert Systems with Applications 42, 5800–5810.

Westergren, A., Unosson, M., Ohlsson, O., Lorefält, B., Hallberg, I.R., 2002. Eating difficulties, assisted eating and nutritional status in elderly (65 years) patients in hospital rehabilitation. International journal of nursing studies 39, 341–351.

Wundersitz, D.W., Josman, C., Gupta, R., Netto, K.J., Gastin, P.B., Robertson, S., 2015. Classification of team sport activities using a single wearable tracking device. Journal of biomechanics 48, 3975–3981.

Xu, R., Zhou, S., Li, W.J., 2012. Mems accelerometer based nonspecific-user hand gesture recognition. IEEE sensors journal 12, 1166–1173.

Zhang, S., Ang, M.H., Xiao, W., Tham, C.K., 2009. Detection of activities by wireless sensors for daily life surveillance: eating and drinking. Sensors 9, 1499–1517.