Automated Analysis of in Meal Eating Behavior using a Commercial Wristband IMU Sensor

Konstantinos Kyritsis[1], Christina Lefkothea Tatli[2], Christos Diou[1] and Anastasios Delopoulos[1]

***Abstract*— Automatic objective monitoring of eating behavior using inertial sensors is a research problem that has received a lot of attention recently, mainly due to the mass availability of IMUs and the evidence on the importance of quantifying and monitoring eating patterns. In this paper we propose a method for detecting food intake cycles during the course of a meal using a commercially available wristband. We first model micro-movements that are part of the intake cycle and then use HMMs to model the sequences of micro-movements leading to mouthfuls. Evaluation is carried out on an annotated dataset of 8 subjects where the proposed method achieves 0.78 precision and 0.77 recall. The evaluation dataset is publicly available at `http://mug.ee.auth.gr/intake-cycle-detection/`.**

I. INTRODUCTION

Eating behavior is a key factor affecting the development of obesity, a disease that has reached epidemic proportions and is currently threatening 1.9 billion overweight and 600 million obese adults worldwide [1]. In particular, the analysis of eating behavior during the course of a meal has been associated with the total food intake and the overall obesity risk. For example, analysis of eating rate deceleration during the course of a meal has been correlated with disordered eating [2] while, on the other hand, restraining eating behavior using feedback has been shown to have a significant impact towards the treatment of obese children [3]. Such findings indicate that there is a lot to be gained by detailed objective measures of in-meal eating behaviour, both for individuals who are already overweight or obese and for normal-BMI people who wish to maintain a healthy lifestyle.

Identifying eating occurrences and the objective quantification of eating behavior with the help of inertial sensors is a challenge that has received significant research attention during the past decade. In most cases, accelerometers or gyroscopes are used in conjunction with other sensors to measure eating-related indicators (e.g., [4]–[6]). There are, however, several works which aim at measuring eating behavior parameters using inertial sensors alone. In [7] the authors explored the potential of using a set of inertial sensors to identify isolated gestures associated with eating, such as “knife cutting”, “spoon loading” and several others. Results highlighted the potential of using inertial sensors for measuring eating behavior.

More recently, Dong et al. [8] proposed a bite detection and counting system based on gyroscopes worn at the wrists. The idea is to detect wrist rotations which precede bites in order to count the number of bites during the course of a meal, and correlate this information with energy intake.

A model of dependence between the different gestures during a meal is presented in [9]. The authors use Hidden Markov Models (HMM) combined with Gaussian Mixture Models (GMM) to model each gesture as well as the dependence between gestures. Although the method uses a small number of gestures and relies on manually segmented sequences, it shows the temporal dependence of gestures, as well as the temporal structure of each gesture.

Motivated by the finding of [9], we propose to model the *food intake cycle* as a sequence of micro-movements associated with a single mouthful. Based on this model, our method uses a commercially available wristband to explicitly detect hand movements such as picking up the food, or moving the food from the plate to the mouth, using an array of SVM classifiers. It then uses the detected movements in conjunction with discrete HMM models to identify intake cycles and characterize eating behavior. The method is described in detail in Section II.

Evaluation was carried out on a dataset of 8 meal sessions performed by 8 subjects recorded at the Aristotle University of Thessaloniki (Section III) and led to encouraging results (Section IV). The dataset is publicly available at the Multimedia Understanding Group website[1].

II. PROPOSED APPROACH

A *food intake cycle* is defined as a period that typically starts by manipulating a utensil for picking up food from the plate. This period then continues with an upwards motion of the hand towards the mouth area, progresses with a motion that enables the placement of food inside the mouth and finishes with a downwards motion of the hand away from the mouth area. Of course, the previous definition of the food intake cycle refers to an ideal case, while in a real life scenario the intake cycle includes repetitions of hand motions as well as times where no hand movement is exhibited (or where the hand movements are not the ones mentioned above).

The term *micro-movement* is used to describe a simple, short duration motion of the hand, as measured by the acceleration and gyroscope sensors. In the context of this study, the term micro-movement refers to the fundamental

*The work leading to these results has received funding from the European Community’s Health, demographic change and well-being Programme under Grant Agreement No. 727688, 01/12/2016 - 30/11/2020 (http://bigoprogram.eu/).

[1]Multimedia Understanding Group, Information Processing Laboratory, Aristotle University of Thessaloniki, Greece

[2]Imperial College Business School, London, UK

[1]`http://mug.ee.auth.gr/intake-cycle-detection`

TABLE I: Table listing all identified micro-movements

Micro-movement	Description
Pick food	Hand manipulates a fork to pick food from the plate
Upwards	Hand moves upwards, towards the mouth area
Downwards	Hand moves downwards, away from the mouth area
Mouth	Hand inserts food in mouth
No movement	Hand exhibits no movement
Other movement	Every other hand movement

periodic movements that are performed during a typical meal session. Table I lists all micro-movements that were identified, along with a short description.

The method presented at this paper aims at detecting intake cycles. This is achieved by initially using a Support Vector Machine to handle the recognition of the individual micro-movements. Two discrete HMMs are then employed for classifying the sequence of the recognized micro-movements as food intake or not.

An overview of the proposed method is presented in Figure 1. The rest of this Section describes each processing step in more detail.

A. Pre-processing and Feature Extraction

Since the accelerometer sensor captures both the acceleration caused due to the subject's movements, as well as the acceleration caused by the effect of the earth's gravity field, the first step is to remove the gravitational component. For this purpose, we employed the algorithm proposed by [10], allowing the usage of gyroscope data as a means of transforming the accelerometer samples to the initial frame. By assuming that the wristband is initially still, the gravitational effect can effectively be removed, by subtracting the first accelerometer measurement. Prior to removing the gravitational component, we smoothed the accelerometer and gyroscope signals with a 5'th order median filter.

Following the pre-processing step, a sliding window approach was adopted for extracting features from the accelerometer and gyroscope streams. The length W_l and the step W_s of the rectangular sliding window W were selected to be 0.2 and 0.1 seconds respectively. For each window a number of features were extracted both from the frequency and the time domain. In more detail, the set of extracted features for each window include, for each axis of both the accelerometer and the gyroscope: i) the zero crossing rate, ii) the mean value, iii) the standard deviation, iv) the maximum value, v) the minimum value, vi) the range of values, vii) the variance, viii) the normalized energy and ix) the first l Discrete Fourier Transform coefficients, where $l = \frac{W_l}{2} + 1$. Finally, for each sensor the SMA (simple moving average) was calculated by $\frac{1}{W_l}\sum_{i=1}^{W_l}|x(i)| + |y(i)| + |z(i)|$, where $x(i)$, $y(i)$ and $z(i)$ correspond to the $i'th$ sample of the x, y and z sensor streams in a given window.

B. Learning the Micro-movements

For the micro-movement recognition task, a multiclass Support Vector Machine classification scheme with one-versus-one classifiers was adopted. In addition, the radial basis function (RBF) was selected as the kernel of choice.

Given the start and end moments provided by annotating the video sequences, the features belonging to each micro-movement except "other movement" were selected for SVM training. Essentially, the five classes of interest were the first five micro-movements of Table I, resulting in ten one-versus-one SVMs. The reason behind the exclusion of the "other movement" class is the large inner-class variance, since it was used to describe every micro-movement not related with the classes of interest and thus, very difficult to model.

Moreover, it was noticed that the "mouth" and "downwards" classes contained significantly less samples compared to the "pick food" class. This was caused due to the first two movements being much shorter in duration than the latter. To this end, each class was proportionally weighted based on it's prior probability. Finally, each feature was linearly scaled in $[0, 1]$.

C. Modeling time evolution

Two discrete 5-state Hidden Markov Models (HMMs) were trained to determine whether or not a sequence of micro-movements constitutes a food intake cycle. The first HMM (HMM_p) was trained using positive sequences, i.e. sequences that correspond to a valid food intake cycle, while on the other hand the second HMM (HMM_n) was trained using negative sequences, i.e. sequences not corresponding to a valid cycle. By using the true labels corresponding in each window W, the ideal food intake cycles were computed as follows. By considering the sequence of SVM labels as symbols, a sequence of symbols was considered as an ideal food intake cycle if it begun with the corresponding "pick food" symbol and ended with a "downwards" symbol. In practice, the start and end points of a valid sequence were the positions of the first "pick food" symbol in sequence of "pick food" symbols and the position of last "downwards" symbol in a sequence of "downwards" symbols.

The Jaccard Index (Eq. 1) was employed in order to enrich the collection of positive sequences as well as generating realistic negative sequences.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

More specifically, given the sequence of labels of a complete meal session of a subject with known start and end positions of the positive sequences, two random integer numbers were generated. The first random number corresponded to the start position of a candidate sequence, and the second number corresponded to the length of the sequence. The Jaccard Index was calculated against all positive sequences of the session, if the result versus a single positive sequence was greater or equal than T_u then the candidate sequence was added to the positive sequences pool. On the other hand, if the Jaccard Index was lower than T_l, then the candidate

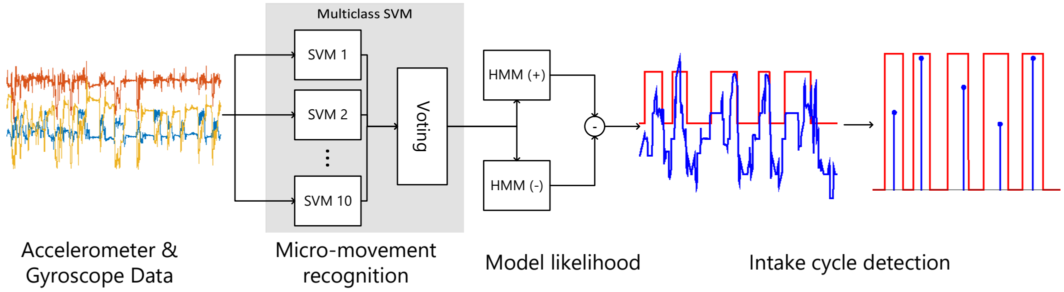

Fig. 1: Overview of the proposed method

sequence was added to the negative sequences pool. In any other case, the candidate sequence was simply discarded. As a result of the previous procedure, 200 positive and 200 negative sequences were generated for each subject.

Finally, all samples belonging in the positive sequences pool, were used for prediction for the given trained SVM models. It should be emphasized that sequences both in the positive as well as the negative pool contain samples belonging in the "other movement" class, that did not participate in the SVM training procedure. The SVM predictions of the samples corresponding to the positive sequences were used as the training sequences for HMM_p. HMM_n was also trained in a similar fashion.

D. Detection

Given the sequence of SVM predictions of a meal, food intake cycles can be detected by adopting a sliding window approach and the two HMMs. Given this approach, a rectangular window W' traverses the predictions sequence with a length $W'_l = 3$ seconds and a step $W'_s = 0.2$ seconds. For each prediction sub-sequence in the series, two log likelihoods λ_p and λ_n are calculated. The first likelihood, λ_p, is a result of modeling the sub-sequence selected by W' using HMM_p, whereas the latter by using HMM_n. Both λ_p and λ_n are normalized by dividing with W'_l. Consequently, a new series $diff$, is created by applying $diff(i) = \lambda_p(i) - \lambda_n(i)$, where i is the position of window W' on the given predictions series. Eq. 2, was used to filter the $diff$ series, by discarding samples where the difference was below a certain threshold T_d.

$$diff'(i) = \begin{cases} diff(i), & \text{if } diff(i) \geq T_d \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

Subsequently, a local maximum search was performed in the filtered $diff'$ series. The minimum distance between two successive peaks was set at 3 seconds. The timestamp associated with each peak, corresponds to the middle of the respective W' responsible for producing the difference $diff(i)$. The resulting peaks correspond to the detected food intake cycles.

III. Dataset

A. Data Collection Protocol

For the purposes of this study, a total of 8 subjects were recorded while eating their launch at the university's cafeteria. The total duration of the recording sums up to 109 minutes, with a mean duration of 13.6 minutes.

The wristband used for the study was an off-the-shelf Microsoft Band 2, containing a triaxial accelerometer providing measurements in g units and gyroscope providing measurements in degrees per second (°/sec). Moreover, an Android application was developed for the purpose of acting as an interface between the wristband's sensors and a mobile phone. The wristband was able to acquire and transmit accelerometer and gyroscope data at approximately 62 Hz. In order to handle inconsistencies in the sampling rate, we estimated the sampling frequency of each sensor, followed up by linear interpolation and then re-sampling the interpolated series to the target sampling frequency of 100 Hz.

Furthermore, a Go Pro Hero 5 action camera was also used for generating ground truth annotations. It was mounted on a short tripod (approximately 23 cm height) placed on top of the table, while facing the subject's upper torso and food tray.

All participants were free to select the food type of their choice. A typical recorded meal consists of a starter, a salad and a main course. Each participant was asked to wear the wristband encapsulating the sensors to the hand responsible for handling the fork. In order to synchronize the accelerometer and gyroscope streams with the camera feed, prior to starting his meal the subject was asked to clap his hands. This procedure enabled synchronization of the two streams with different timestamps, since the quick hand motion required for clapping exhibits an easily identifiable sharp peak in the acceleration signal. No other instructions were given to the subjects on how to eat their meals. As a result, the participants were free to engage in conversations with other individuals seated near them, use their cellphones and perform any other non meal-related activities. After the subject finished their meal, they were asked to perform a second clap marking the end of the collection procedure as well as verifying the initial synchronization offset.

B. Data Annotation

For all video recordings, the start and end points of all six micro-movements of interest were manually labeled. The annotation process was performed in such a way that the start and end times of each micro-movement span the whole meal session, without overlapping each other. More specifically, at

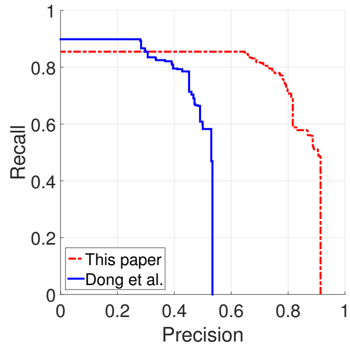

Fig. 2: Precision-recall curve, for the method presented at [8] and the proposed approach

each time t each sensor generated sample $s(t)$ is associated by a specific label.

IV. Experiments & Results

Both the classifiers for micro-movements as well as for the food intake cycles were trained and tested using the Leave One Subject Out (LOSO) cross validation scheme. Meaning that for each subject j in the dataset, one SVM model SVM^j and one pair of HMM_p^j and HMM_n^j were produced. In more detail, all data samples from all subjects, belonging in the "other movement" class were excluded from the SVM training process. However, every occurrence of the "other movement" class existed in the test sets. As a consequence, for each subject, the HMMs were trained by using the SVM output (i.e. the SVM interpretation of the samples belonging in the "other movement" class).

Moreover, the SVM parameters were set to $\gamma = 0.1$ and $C = 100$ for all subjects, after experimenting with a small subset of the micro-movement data. In a similar fashion, thresholds T_u and T_l were set to 0.8 and 0.6 respectively. The threshold parameter T_d used in Eq. 2 was selected by picking the one that achieved the highest F1 score and was set to 0.016, however the complete precision-recall curve is also provided (Figure 2).

Given, t_s^k and t_e^k, as the true start and end points of the k'th food intake cycle respectively and t_p^l as the timestamp of the l'th detected peak resulting by the detection methodology, we were able to evaluate the performance of our approach by using the following evaluation scheme.

If for l'th detected peak, the inequality $t_s^k \leq t_p^l \leq t_e^k$ holds, then it counts as a true positive. However, every other occurrence of a peak in the same food intake cycle k, counts as a false positive. Moreover, if for a specific detected peak the above inequality doesn't hold for any food intake cycles of the session, then it counts as a false positive as well. Finally, if for a specific food intake cycle there is no detected peak that can fulfill the inequality above, then it counts as a false negative.

Given this evaluation scheme, the performance of the proposed method is measured by calculating the precision and recall metrics. For comparative evaluation purposes, the method proposed in [8] was also implemented and was evaluated on the same dataset. Figure 2 depicts the relation between the precision and recall pair for both methods, given different threshold values. Furthermore, the evaluation results are listed in detail in Table II. For the method of Dong *et al.* the threshold was selected as described in [8].

TABLE II: Evaluation results

Method	**TP**	**FP**	**FN**	**Precision**	**Recall**
Proposed approach	308	109	115	0.7806	0.7713
Dong *et al.*	395	482	108	0.4504	0.7853

Based on these results, the proposed method of explicitly modeling the micro-movements during meal cycles leads to satisfactory results regarding the automatic and objective quantification of eating behavior.

V. Conclusions

We have presented an approach for the detection of intake cycles during the course of a meal using a commercially available IMU. Results indicate that the combination of SVM-based micro-movement detectors with discrete HMMs leads to effective detection of intake cycles. Evaluation on an annotated dataset of 8 meals shows that the proposed method achieves high effectiveness and improves over a known method proposed in the bibliography.

References

[1] WHO, "Fact sheet: obesity and overweight," *Internet: http://www.who.int/mediacentre/factsheets/fs311/en/ (accessed 16 Feb 2017)*, 2016.

[2] M. Zandian, I. Ioakimidis, C. Bergh, and P. Södersten, "Linear eaters turned decelerated: Reduction of a risk for disordered eating?" *Physiology & behavior*, vol. 96, no. 4, pp. 518–521, 2009.

[3] A. L. Ford, C. Bergh, P. Södersten, M. A. Sabin, S. Hollinghurst, L. P. Hunt, and J. P. Shield, "Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial," *Bmj*, vol. 340, p. b5388, 2010.

[4] V. Papapanagiotou, C. Diou, L. Zhou, J. van den Boer, M. Mars, and A. Delopoulos, "A novel chewing detection system based on ppg, audio and accelerometry," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, 2016.

[5] 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*. ACM, 2016, pp. 451–462.

[6] J. M. Fontana, M. Farooq, and E. Sazonov, "Estimation of feature importance for food intake detection based on random forests classification," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 6756–6759.

[7] O. Amft, H. Junker, and G. Troster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," in *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*. IEEE, 2005, pp. 160–163.

[8] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Applied psychophysiology and biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.

[9] R. I. Ramos-Garcia, E. R. Muth, J. N. Gowdy, and A. W. Hoover, "Improving the recognition of eating gestures using intergesture sequential dependencies," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 825–831, 2015.

[10] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," in *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–7.