Detection of eating and drinking arm gestures using inertial body-worn sensors

Oliver Amft, Holger Junker, Gerhard Tröster
Wearable Computing Lab, IfE,
ETH Zurich, Switzerland
{amft, junker, troester}@ife.ee.ethz.ch

Abstract

We propose a two-stage recognition system for detecting arm gestures related to typical meal intake. Information retrieved from such a system can be used for automatic dietary monitoring in the domain of behavioural medicine. We demonstrate that arm gestures can be clustered and detected using inertial sensors. Such sensors can be integrated unobtrusively into normal clothing. To validate our method, experimental results including 384 gestures from two subjects are presented. Using an isolated discrimination based on HMMs an accuracy of 95% can be achieved. When spotting the gestures in continous movement data, an accuracy of of up to 87% is reached.

1 Introduction

Maintaining a healthy lifestyle is generally considered as the most important prerequisite in prevention of cardiovascular diseases. Today, these heart related diseases are the most prominent cause of death for large share of the population caused by three main risk factors: sedentary lifestyle, stress and wrong diet.

In Europe, a research program has been established which aims to reduce the heart disease risk by combining long term physiological and behavioural monitoring with personalised direct or professional-observed feedback. In the related project presented in this paper, our goal is to deploy wearable sensing technology to aid the individual and health professionals in monitoring the individual's eating habits.

Dietary information for health maintenance Complete dietary monitoring involves a variety of aspects, including composition, daily schedule, rate of intake, speed and timing. To date, approaches to monitor individual diet schedule are based on user questionnaires which are considered imprecise and which require large regular effort by the user in entering the data manually. Rate of intake and additional timing factors have been observed in laboratory settings only.

Health maintenance and disease prevention requires a continuous, quasi permanent, monitoring to be effective. Hence, a system intended for automatic acquisition of dietary information is needed to reduce the user's effort.

Automatic diet monitoring The unsupervised estimation of type and amount of all meal intake is clearly more a vision than a realistic approach to start. However, we believe that with a combination of wearable sensors and a degree of environment augmentation useful assistive systems are conceivable. We envision that such a system will have the benefits of 1) Making a rough estimation on food consumption, similar to today's physical activity monitors, 2) Providing the user with a best guess on the type, schedule and amount of what has been eaten and ask for corrections, 3) Indicating unhealthy eating habits, e.g. high speed of food intake or day schedule.

Appropriate wearable and non-invasive sensing domains which provide evidence for the recognition of food intake are 1) Using motion sensors to detect gestures related to food intake, 2) Detecting and analysing chewing sounds, 3) Using electrodes mounted at the lower throat, e.g. in a collar, to monitor swallowing.

Paper contributions In this work we concentrate on information derived from wearable motion sensors to detect gestures directly related to food intake, e.g. moving the arm towards the mouth and back. In particular, we present first results using a two-stage detection approach based on a segmentation of continuous sensor data and subsequent identification of relevant gestures on the pre-evaluated segments. The paper presents the following results:

1. We show that motion sensors at the user's arms can provide good quality information for the detection of eating and drinking gestures. Moreover, we show that a set of defined isolated eating and drinking gestures can be discriminated using Hidden Markov Models

(HMMs) from other usual movements with good accuracy.

2. We show that it is possible to spot relevant gestures individually in a continuous stream of movement data and present a segmentation procedure for the relevant gestures.

3. We present first results of the continuous gesture recognition and show that a discrimination of eating and drinking gesture categories from arbitrary movements and other intended gestures is possible.

Related work The recognition of gestures is studied broadly by the use of vision based systems, e.g. for computer interfaces [1]. Less work has been made to gather information form body-worn inertial sensors, e.g. accelerometers and gyroscopes [2, 3]. Besides the on-body information acquisition, large efforts have been made to determine user context from the instrumented environments. Realisation of such intelligent environments have been studied, e.g. in the context of smart homes [4]. Smart identification systems have been developed [5] which may provide additional information associated to nutrition phases, e.g. smart cups [6].

2 Methodology

Approach Gestures related to nutrition intake can be roughly discriminated into coarse preparation phase of the food or beverage items, e.g. unpacking, cooking, plate loading and the actual feeding, e.g. gestures which are intended to fine-cutting, loading and manoeuvring the prepared item to the mouth. The feeding phase can be supported by means of specific tools, e.g. fork or spoon or can be conducted directly with the hand. The used tool does heavily influence the gesture perceived.

In this work, we focus on the feeding phase, since it relates directly to the aspects of the envisioned automatic diet monitoring system. Therefore, we attempt to recognise intentional arm movements of the manoeuvring sub-phase to and from the mouth using inertial sensors. Obviously the sub-phases for fine-cutting and loading may not be available in certain gesture categories. Hence, they can be viewed as an additional discriminating information source.

The challenge for this recognition task is twofold: on the one hand, the robust recognition of relevant nutrition gestures and on the other hand, the spotting of such non-periodic, sporadically occurring movements in a continuous data stream containing long periods of non-relevant motions. Both problems are related to each other, although most existing work only address the problem of recognition on well-defined sequences of gestures, e.g. [7].

Methods for spotting activities out of non-relevant data using a threshold model have been proposed e.g. in [1]. In this paper, we use a similarity search as proposed in early work [2] to address the spotting task. This search requires an explicit segmentation of the data stream and relies on a natural partitioning of the motion data into characteristic motion segments. The similarity search then identifies subsequent motion segments potentially containing a relevant gesture segment. To provide robust recognition, the segments retrieved by the similarity search are then classified using HMMs to eliminate falsely retrieved gesture segments.

Experiments To evaluate our recognition approach, a variety of data sets were recorded using a commercially available motion sensor system [1]. The sensors were attached on the body as illustrated in Fig. 1. Two test persons (1 female, 1 male) were seated in front of a table with the nutrients and instructed to eat/drink normally. Individual sessions were recorded for the different nutrition categories with at least 20 minutes break in between. Food types were kept constant for each individual category. All meals were cold enough to allow normal intake. Table 1 shows the different nutrition categories recorded.

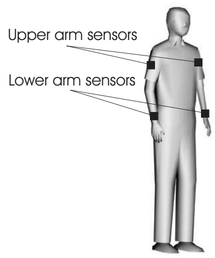

Figure 1. Sensor placement for capturing nutrition related gestures

Category	Tools	Food type and gesture
Cutlery	fork, knife and plate	Lasagne; fork tap and loading, involved knife cutting activity
Spoon	spoon and bowl	Cereals with milk; spoon loading and manoeuvring
Hand	-	Chocolate bar; held in hand, moved to mouth and back
Drink	glass	Water; glass was moved to mouth and back on table

Table 1. Details of the nutrition gesture categories

Table 2 summarises the acquired data which was inspected and annotated. The following hand gestures have additionally been recorded in the same setup for performance evaluation: 1) Scratching head (86x), 2) Touching

[1]Manufacturer: XSens, Model MT-9B

chin (89x), 3) Turning pages of newspaper (115x), 4) Arbitrary arm movements (93x).

Category	Total number of recorded gestures	Mean duration of gesture [s]
Cutlery	100	11.2
Spoon	151	6.8
Hand	62	6.3
Drink	71	9.7

Table 2. Statistics of acquired nutrition gestures

3 Gesture segmentation

The first stage of our recognition system is dedicated to the segmentation of continuous sensor data into motion segments and to derive possible gesture segments as a series of adjacent motion segments.

Motion segment estimation For the motion segmentation the SWAB (Sliding window and bottom up) algorithm proposed by Keogh et al. [8] was used due to its excellent performance. SWAB combines the advantages of a precise bottom segmentation scheme and a sliding window algorithm. This allows the algorithm to be used on-line. SWAB uses a simple cost metric on piecewise linear representations of the data to merge segments. We have chosen the squared sum of the linear regression as the primary cost function. To further reduce the number of segments derived, we merged two adjacent segments if their linear approximation appeared to have similar slopes.

We use the angle from lower arm rotation (pronation/supination) as motion parameter for the SWAB algorithm. This parameter showed the best explainable behaviour of the lower arm parameters in all relevant gestures.

Gesture segment estimation To identify potential gestures, a similarity search based on the Euclidean distance in a feature space is performed on the segmented data stream. The following features are used for this search: Gesture segment length, number of motion segments, begin and end of lower arm rotation as well as rate of turn in lower arm rotation and pitch (angle between lower arm and vertical plane) orientations. A gesture segment is considered for the HMM recognition, if the Euclidean distance calculated from the extracted features of this segment is smaller than pre-defined, gesture-specific threshold values.

Table 3 summarises the performance of the similarity analysis for the relevant gesture categories. The gesture-specific threshold-values were set to such values that the number of deletions (gestures that have occurred but not identified) are minimised , while also keeping the the number of insertions (falsely retrieved gesture segments) small.

	Category			
	Cutlery	**Drink**	**Hand**	**Spoon**
Relevant	100	71	62	151
Retrieved	90	61	39	99
Insertions	43	3	60	38
Deletions	10	10	24	55
Accuracy	0.90	0.86	0.63	0.66

Table 3. Gesture segmentation performance

The similarity analysis performs well on gesture classes Cutlery and Drinking, while introducing more deletions for the classes Hand and Spoon.

Figure 2 depicts the segmentation result on a sample of the sensor data from the lower arm.

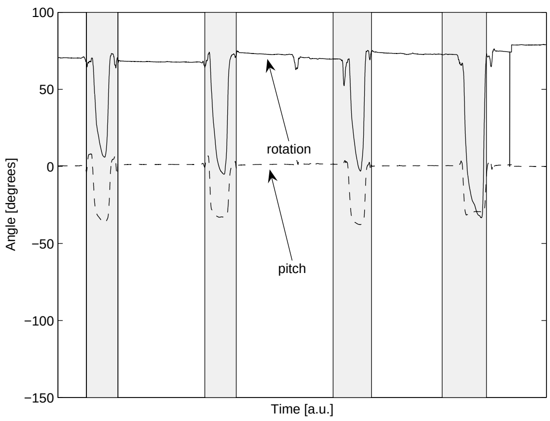

Figure 2. Sensor data (lower arm) segmentation obtained for class Drinking

4 Recognition of gestures

For the recognition of relevant nutrition gestures from the previous segmentation stage, HMMs were used. We deployed continuous features and left-right models. Individual HMMs were built for each nutrition gesture class using single Gaussians to model the states. The number of states for each model was varied between 3 and 10. More than 5 states only improved the recognition accuracy marginally, therefore the results presented reflect the performance of 5 state HMMs for all gesture classes. The features used for the HMMs include the pitch angles of lower and upper arms, rate of turn of lower arm, rotation of lower arm and upper arm change of angle to the horizontal plane.

To validate the models, classifications were made using a 10-fold cross-validation on 62 gestures per class from the database summarised in table 2. The number of training gestures was varied between 50% and 90% with marginal recognition improvements for higher number of training segments. This suggests that the gesture features are consistent and stable over multiple conducted recording sessions and the natural differences in the test persons. Results for the recognition on isolated gestures are shown in table 4.

Down: Truth	Cutlery	Drink	Hand	Spoon	Movements
Cutlery	310	0	0	0	0
Drink	0	297	0	0	13
Hand	9	0	301	0	0
Spoon	0	0	0	304	6
Movements	26	40	19	13	1142

Table 4. Confusion matrix of isolated HMM nutrition gesture recognition (Class 'Movements' includes 4 sub-classes)

The combined recognition performance of gesture segmentation search using motion segments and the HMMs trained with the same parameter set as above is shown in table 5. For training 50% of the 62 gestures were used.

	Cutlery	Drink	Hand	Spoon
Relevant	31	31	31	31
Recognised	27	25	22	16
Insertions	9	4	12	10
Deletions	4	6	9	15
Accuracy	0.87	0.81	0.71	0.52

Table 5. Performance evaluation for spotting nutrition gestures

5 Conclusion and Future Work

The results presented in this paper indicate, that a discrimination of eating and drinking gestures from other, intended gestures and arbitrary movements is possible with an accuracy of 95% on isolated gesture segments. The isolated HMM-based recognition showed that the gestures and the features extracted are consistent and stable over multiple conducted recording sessions and the natural difference in the test persons.

The performance of the continuous recognition is clearly bound to the results of the segmentation step. The proposed segmentation, consisting of a time series partitioning into motion segments and the estimation of nutrition gestures is an applicable concept, although additional investigations are needed to improve the detection performance. The aspect of moving the hand to and from the mouth may be a valuable segmentation information, that could be derived from arm and trunk orientation or by means of a proximity sensing between hand and a collar worn receiver.

While much still remains to be done, our work proves the feasibility of using gestures as one important component in a diet monitoring system. By discriminating different gesture categories even more detailed information can be derived, contributing to the detection of meal type.

References

[1] L. Hyeon-Kyu and H. Kim-J, "An HMM-based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Oct. 1999.

[2] H. Junker, P. Lukowitz, and G. Troester, "Continuous recognition of arm activities with body-worn inertial sensors," in *Proceedings of the Eighth International Symposium on Wearable Computers, ISWC*, pp. 188–189, 2004.

[3] J. C. Lementec and P. Bajcsy, "Recognition of arm gestures using multiple orientation sensors: Gesture classsification," in *7th International IEEE Conference on Intelligent Transportation Systems*, October 2004.

[4] E. Mynatt, A.-S. Melenhorst, A.-D. Fisk, and W. Rogers, "Aware technologies for aging in place: understanding user needs and attitudes," in *IEEE Pervasive Computing*, pp. 36– 41, April-June 2004.

[5] K. Romer, T. Schoch, F. Mattern, and T. Dubendorfer, "Smart identification frameworks for ubiquitous computing applications," in *Proceedings of the First IEEE International Conference Pervasive Computing and Communications, 2003 (PerCom 2003)*, 2003.

[6] M. Beigl, H.-W. Gellersen, and A. Schmidt, "MediaCups: Experience with design and use of computer-augmented everyday artefacts," *Computer Networks, Special Issue on Pervasive Computing*, vol. 35, no. 4, pp. 401–409, 2001.

[7] S. Chambers-G., S. Venkatesh, W. West-G.-A., and H. Bui-H., "Hierachical recognition of intentional human gestures for sports video annotion," in *Proceedings International Conference on Pattern Recognition 2002*, vol. 2, 2002.

[8] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *Proceedings 2001 IEEE International Conference on Data Mining*, 2001.