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

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

Abstract

We propose a two-stage recognition system for detecting arm gestures related to human 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. To validate our method, experimental results including 384 gestures from two subjects are presented. Using isolated discrimination based on HMMs an accuracy of 94% can be achieved. When spotting the gestures in continous movement data, an accuracy 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 this regard obesity can be an indication for unhealthy lifestyle.

It is envision that a combination of long term physiological and behavioural monitoring with personalised direct or professional-observed feedback may reduce the heart disease and obesity risk. 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.

1.1 Dietary information for health maintenance

Complete dietary monitoring involves a variety of aspects, including meal composition, daily schedule and rate of intake. To date, approaches to monitor individual diet schedule are based on user questionnaires which are considered imprecise and require large regular effort of 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 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.

1.2 Automatic diet monitoring

Obviously, the unsupervised estimation of type and amount of all meal intake is currently more a vision. However, we believe that with a combination of wearable sensors and a degree of environment augmentation, useful assistive systems are conceivable. 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 and 3) Indicating unhealthy eating habits, e.g. high speed of food intake or inadequate daily schedule.

Appropriate wearable and non-invasive sensing domains which provide evidence for the recognition of food intake include 1) Gestures related to food intake, 2) Detection and analysing chewing sounds and 3) Detection of swallowing.

1.3 Paper contributions

In this paper 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 very good accuracy.

Proceedings of the 2005 Ninth IEEE International Symposium on Wearable Computers (ISWC'05)
0-7695-2419-2/05 $20.00 © 2005 IEEE

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.

1.4 Related work

The recognition of gestures has been studied broadly using vision based systems, e.g. for computer interfaces [1]. Less work has been made to gather information from body-worn inertial sensors, e.g. accelerometers and gyroscopes [2]. 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 [3]. Smart identification systems have been developed [4] which may provide additional information associated to nutrition phases, e.g. smart cups [5].

2 Methodology

2.1 Approach

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

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, it is 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. [6].

Methods for spotting activities out of non-relevant data using a HMM threshold model have been proposed, e.g. in [1]. To address the spotting task, we develop a similarity search based on earlier work of our group [7]. 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.

2.2 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 wrist and upper arm as illustrated in Fig. 1. In this setting the movements of the lower and upper arms were tracked with a sampling rate of 100 Hz. Two test persons (1 female, 1 male, mean age 29 years) were seated in front of a table with the nutrients and instructed to eat and 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 recorded nutrition categories.

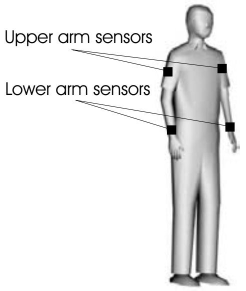

Figure 1. Sensor placement for capturing nutrition related gestures

Table 1. Nutrition gesture categories

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 2 summarises the acquired data which was inspected and annotated. The following additional arm gestures have been recorded in the same setup to evaluate

[1] Manufacturer: XSens Technologies B.V., model MT-9B

the recognition performance: 1) Scratching head (86x), 2) Touching chin (89x), 3) Turning pages of newspaper (115x) and 4) Arbitrary arm movements (93x).

Table 2. Statistics of acquired gestures

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

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 subsequent motion segments.

3.1 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-up segmentation scheme and a sliding window algorithm, that allows the algorithm to be used on-line. To obtain segmentation boundaries, the algorithm uses a simple cost metric on piecewise linear representations of a selected segmentation feature. When the cost for two adjacent initial segments is below a threshold the segments are merged.

The following implementation was chosen: We use the angle from the lower arm rotation (pronation/supination) as segmentation feature for the SWAB algorithm. This feature is provided directly by the motion sensors and showed the best explainable behaviour of all lower arm parameters in the relevant gestures. We have chosen the squared sum of the linear regression as the primary cost function on the segmentation feature. To further reduce the number of segments derived, we merged two adjacent segments if their linear approximations appeared to have similar slopes.

3.2 Gesture segment estimation

To identify potential gestures, a similarity search based on the Euclidean distance in a feature space is performed on multiple adjacent motion segments derived by the SWAB algorithm. The following features were used for this search: Segment length, number of motion segments, begin and end of lower arm rotation as well as rate of turn of 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 a pre-defined, gesture-specific threshold value.

Table 3 summarises the performance of the similarity analysis all the relevant gesture categories. The gesture-specific threshold values were found by minimising the number of deletions (gestures that have occurred but not identified), while keeping the number of insertions (falsely retrieved gesture segments) small.

As Table 3 indicates, the similarity analysis performs well on the gesture classes Cutlery and Drink, 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.

Table 3. Gesture segmentation performance

	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

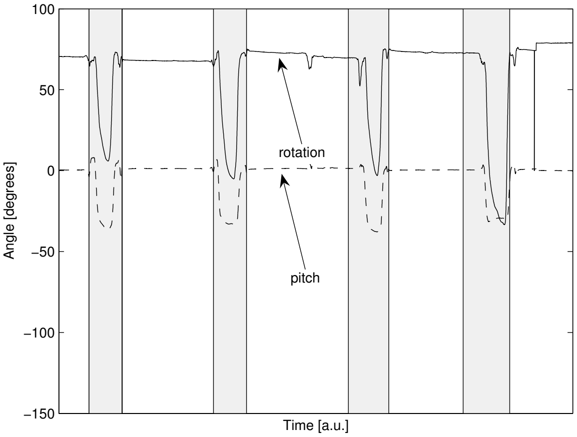

Figure 2. Example of sensor data (lower arm) and segmentation obtained for class Drink

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 improved the recognition accuracy only 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 and upper arm.

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. Results for the recognition on isolated gestures are shown in Table 4.

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. From each class 31 gestures were used for training and testing respectively.

Table 4. Confusion matrix of isolated HMM nutrition gesture recognition, accuracy: 94% (Class 'Movements' includes 4 sub-classes)

Down: Truth	Cutlery	Drink	Hand	Spoon	Movements
Cutlery	310	0	0	0	0
Drink	0	299	0	0	11
Hand	8	0	302	0	0
Spoon	0	0	0	305	5
Movements	29	52	17	13	1129

Table 5. Performance evaluation for spotting nutrition gestures

	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

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 94% on isolated gesture segments. Moreover the isolated HMM-based recognition showed that the gestures and the features extracted are consistent and stable over multiple recording sessions and the natural differences 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 additional information, that could be derived from arm and trunk motion sensor orientation or by means of a proximity sensing between hand and a collar worn receiver.

The initial results obtained in this work will be verified in a next step by including additional test persons and extending the set of non-relevant gestures used for the evaluation by real-life data acquisition. Furthermore the influence of the test person age will be evaluated.

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[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.