Technology Reports
Wearable Device | Motion Recognition | Meal Content Estimation

Meal Content Estimation Technology Focusing on Forearm Motion —Toward Simplified Meal Management—

Research Laboratories Takato Saito Satoshi Kawasaki† Daizo Ikeda Masaji Katagiri

Recent years have seen the social issues of lifestyle-related diseases caused by overeating and unbalanced diets emerge. While there are a number of balanced diet management services available, these require the user to constantly record information on their diet without any omissions, which takes time and is an impediment to receiving the appropriate support from the service. In this article, we propose a meal content estimation methodology that focuses on the forearm motion associated with eating. This technology uses sensor data from a wearable device on the user's dominant hand to enable recognition of meal content without the hassle of user input.

1. Introduction

It has often been said in recent years that dietary issues such as unbalanced eating habits or overeating can cause lifestyle-related illnesses manifesting as social issues. With the diversification of lifestyles and the expansion of meal choices, opportunities to eat out and consume processed foods have been increasing. These foods tend to have high salt and fat content and continued consumption of these foods raises concerns about increases of the number of people likely to get lifestyle-related illnesses. Therefore, it is important to consider a balanced diet in case of daily eating-out and consumption of processed foods.

As an initiative to promote the health of citizens, the Ministry of Agriculture, Forest and Fisheries has formulated dietary guidelines [1], and advocates review of eating habits. Also, the Ministry of Health, Labor and Welfare has formulated a balanced food

©2018 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

† Currently, Service Innovation Department

guide [2] to tie specific actions to these dietary guidelines. In spite of these initiatives, a survey by the Cabinet Office indicates that one in two people do not have an appropriate diet [3], the most prevalent reasons for which are (1) they don't exactly know what they should do regarding diet, or (2) they are too busy and don't have time to consider diet [4].

These circumstances have led to a wide range of companies and organizations providing services to support management of dietary balance. These services take many forms, and include a wide variety of smartphone-based dietary support services [5]. However, to receive proper assistance using these services, users should continuously record meal information every day without omissions. Therefore, many of the existing services require manual user entry of feeding details such as the time and content of meals, which takes time and hence can discourage continued use.

Technology for recognizing wrist action to record the time of meals with high precision has been developed [6], and wearable devices with these functions [7] are available for practical use. In addition, to record meal content, recognition methods using images of food captured with the camera [8] and methods of recognizing the upper limb movements associated with feeding*1 [9] etc. have been researched. The former has issues with users forgetting to capture images and privacy protection, while regular use of the latter is seriously hindered because multiple devices must be attached to the body in a number of locations.

Hence, NTT DOCOMO has developed technology to estimate meal content using only forearm motion with the aim of being able to continually and automatically grasp the details of meal content without any of these hindrances. This article describes the method of recognizing meal content details from sensor data acquired from a generic wristband-type wearable device worn on the forearm of the user's dominant hand.

2. Related Research

2.1 Amft, et al.'s Method

As the method for estimating meal content, Amft, et al. proposed a method using accelerometers*2 and gyro sensors*3 placed in four locations on both upper arms and forearms to collect upper limb motion data, as well as a throat microphone*4 and microphone earbuds*5 to collect data [9] [10]. Their method estimates meal content through detection via sensors picking up the sequence of actions that occur with eating, which include preparatory motions to cut food, motions to bring the food to the mouth, and chewing and swallowing. Hence, they showed the effectiveness of using data sources other than imaging. This research suggests the importance of a temporal sequence of motions for estimating meal content. However, this method requires wearing multiple devices for accurate detection using a state transition model*6 and thus presents a major impediment to regular use.

2.2 Zhang, et al.'s Method

In technological research to recognize human motion such as walking and running, Zhang, et al. proposed a high accuracy method of recognizing human motion that focuses on the frequency of states rather than the transition of states [11]. In

*1 Feeding: The act of carrying food to the mouth and consuming it.

*2 Accelerometer: A sensor that measures changes in speed. Equipping a mobile terminal with an accelerometer enables it to sense changing movement. Sensors fixed mutually and orthogonally to each other to measure acceleration in 3 directions are also referred to as "3-axis accelerometers."

*3 Gyro sensor: A sensor that measures rotational speed.

*4 Throat microphone: A microphone attached to the throat to capture low-level vocalizations or swallowing sounds etc.

*5 Microphone earbud: A microphone inserted into the inner ear to capture the sounds associated with chewing that travel through the jaw.

this method, actions such as walking and running are interpreted as collections of multiple characteristic primitive motions, where a Bag-of-Words (BoW) representation*7 is applied to distinguish actions. As this research shows, we believe the BoW representation is an effective way to focus on human motion, even for recognizing forearm motions associated with eating. Hence, we employ a method that includes BoW representations to recognize forearm motion as combinations of characteristic primitive motions.

3. The Proposed Method

3.1 Overview

Figure 1 shows an overview of our proposed method. This study focuses on differences in forearm motions when eating to estimate meal content. For example, when eating a hamburger, motions to carry the food to the mouth and motions to change the chewing location occur. In contrast, when eating ramen noodles, motions to lift up the noodles, motions to cool the noodles and motions to slurp the noodles occur in a time sequence. Meal content estimation is made possible by recognition based on the time sequence of these motions, because they are different depending on meal contents. This proposed method entails wearing a device on forearm of the dominant hand and acquiring sensor data accompanying forearm motion when eating (hereinafter referred to as "forearm motion data"). Features*8 are extracted from this

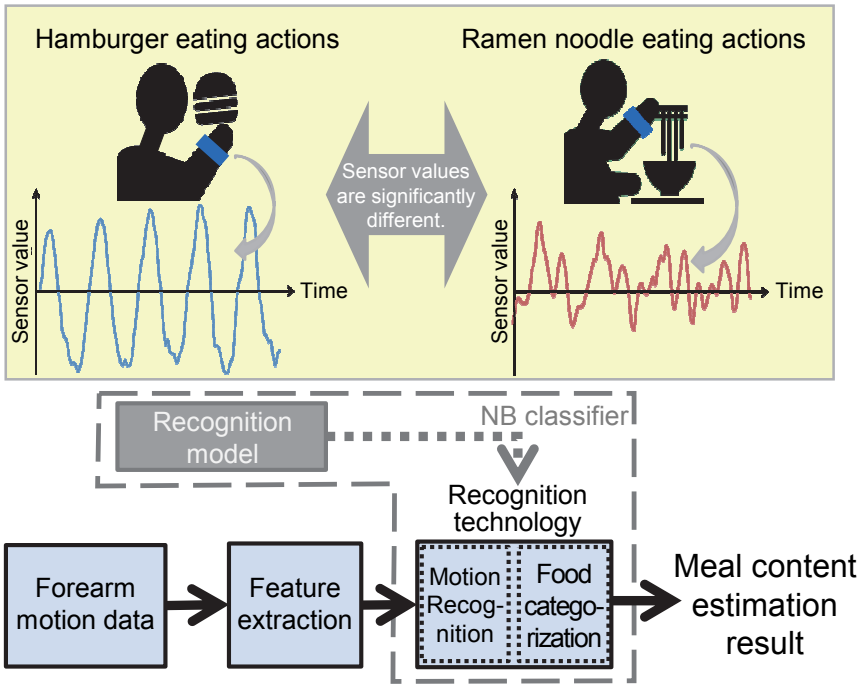


Figure 1 Overview of the proposed method

*6 State transition model: A model that expresses the flow of a procedure as transitions of states beginning with an initial state, and finishing with a final state. For example, in the process of eating, the initial state describes preparatory actions, followed by repeated transitions between carrying food to the mouth and chewing, with swallowing as the final state.

*7 BoW representation: Used with document categorization, a method of expressing a document in terms of the frequency that words appear in it.

*8 Feature: Values, or collections of values arbitrarily calculated to compare data. Calculating the features from source data is called "feature extraction." With data consisting of multiple columns such as that from a 3-axis accelerometer, or if multiple calculation methods are selected, features can also consist of multiple columns. In such cases, the number of columns that make up the feature is called the "dimension."

forearm motion data, and a Naive Bayes classifier*9 (hereinafter referred to as "NB classifier") recognizes motions to estimate meal content.

As previous methods to recognize meal content required multiple wearable devices, achieving a similar method with one device has been problematic. In this proposed method, we use BoW representations to recognize eating movements as combinations of characteristic primitive motions. Then, through an approach extending to N-gram*10 to describe time series information based on the order that motions occur, we have achieved meal content estimation with fewer wearable devices, which was problematic with the aforementioned conventional methods.

We have achieved a simpler method for continually and automatically recognizing meal content without hindering user participation, using only one commercially available wristband-type device worn on the dominant hand.

3.2 Flow of Estimation

Data from a 3-axis accelerometer is used as the forearm motion data, and the feature is extracted from this sensor data in fixed width moving windows*11 after the data has been filtered to reduce noise. If the dimensions of the feature are increased unnecessarily, the performance of the classifier will degrade, hence in this study we use a 12 dimension statistical value of the accelerometer for features, as shown in **Table 1**. With forearm motion when eating, a recognition method based on the order that motions occur is effective. Hence, firstly a BoW representation is applied to the 12 dimension feature, and expressed as a combination of elemental characteristic movements (hereinafter referred to as "primitive motions") (**Figure 2** (1)). Then, considering the order that the primitive motions occur and summarizing them into collections (hereinafter referred to as "primitive motion sequence") (Fig. 2 (2)), so that forearm movement data can be expressed as the frequency of occurrence of primitive motion sequences (Fig. 2 (3)) which is then applied to the NB classifier to estimate meal content. As shown in Fig. 2, it's possible to represent a set of primitive motion sequences by extending to N-gram after expressing forearm motion data as a set of primitive motions. Here, the primitive motions are expressed as "words," hence, movements such as slurping noodles or eating a donburi (rice bowl with topping) can be expressed as primitive motion sequences. For instance, the case of slurping noodles consists of movement to grab the noodles, movement to lift them up, and movement to carry them to the mouth. These movements can be expressed as primitive motions on a time axis. However, since eating actions such as carrying food to the mouth or chewing in which the forearm does not move occur with every meal, these words are not effective for estimating meal content. This means lowering importance placed on frequent words for various common eating actions,

Table 1 Extracted features and their dimensions

Features	Number of dimensions
Average value	3
Standard deviation	3
Variance	3
Median	3

*9 Naive Bayes classifier: A classifier based on a probability theory called "Bayesian probability."

*10 N-gram: A method of viewing N number of words in a series, among the words included in a document, as a new word.

*11 Moving window: In processing sensor data, the range of the target data is referred to as a "window." One of these windows moved according to certain rules is referred to as a "moving window."

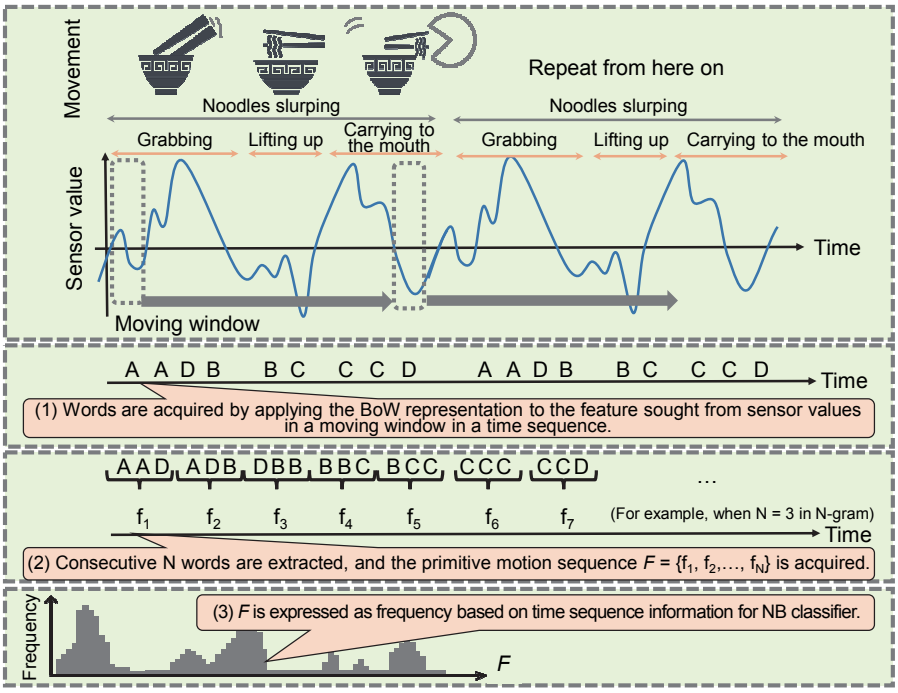


Figure 2 BoW representation using N-gram

while magnifying the importance of words for unique eating actions for specific meal content to raise the accuracy of meal content estimation.

When eating, primitive motions are repeated or performed alternately. In contrast to the aforementioned method of Amft, et al., research based on the continuity of all state transitions included in a single meal, this method can capture local continuity because it focuses on the frequency of actions using an NB classifier. For this reason, our method holds the promise of better meal content estimation accuracy compared to conventional methods, even with sudden noise or changes in the order of eating due to meal contents.

4. Experiments and Evaluations

4.1 Forearm Motion Data Collection and Experimental Conditions

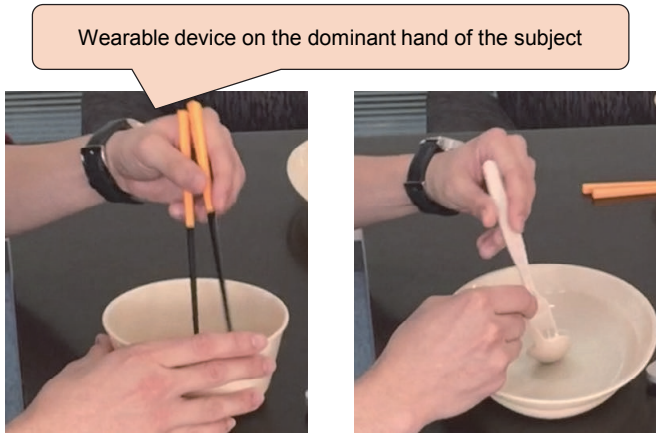
1) Experimental Conditions

We chose five items for meal content recognition as shown in **Table 2**, and also tabulated utensils used by experimental subjects at mealtimes, and the number of data collected. Forearm data was collected using an Android Wear™*12 -based wristband-type device (HUAWEI WATCH W1®*13) to collect forearm motion data. As shown in **Figure 3**, subjects ate with utensils in their dominant hand, and forearm movement data was collected. Forearm motion data covers approximately seven minutes

*12 Android Wear™: A trademark of Google Inc.
*13 HUAWEI WATCH W1®: A registered trademark of HUAWEI Technologies Co.

Table 2 Breakdown of meal contents, utensils and number of data

Meal content	Utensils	Requirements
Donburi, (with beef, pork cutlet, chicken and egg, tempura (fritter) or seafood topping)	Chopsticks	29
Curry and rice	Spoon	32
Breads (pasties, sandwiches, pizza and burgers)	Bare hands	26
Pasta (spaghetti)	Fork	27
Noodles (ramen)	Chopsticks	33



(a) Grabbing noodles with chopsticks

(b) Coiling pasta with a fork

* Data collected while actually eating meals

Figure 3 Forearm motion data collection experiments

per meal. Of the nine subjects, whose ages ranged from 20 to 40 (seven males and two females), one male was left-handed, while the remainder were right-handed. Referring to the number of data used in research done by Amft, et al. [9], we collected forearm motion data for approximately 30 meals per meal content item. However, because forearm motion data for the meal contents we chose for this research was collected during the subjects' daily lives, there were differences in the amount of forearm motion data collected depending on the subjects and the content of the meals they ate. In addition, the forearm motion data was collected in company canteens, residences and restaurants etc., and hence was not exactly the same, even for the same meal contents.

2) Collection of Experimental Data

It is said that in general, 99% of bodily movement takes place at or below 15 Hz [12]. For this reason, recognizing actions such as walking or

running with accelerometers often entails measuring frequencies approximately 100 Hz to provide a margin [11]. Also, it has been shown that frequencies of eating actions obtained from wearable devices worn on the wrist in real life are between 0.2 and 0.6 Hz [13]. Since it is preferable to perform sampling at least twice the frequency of the motion to be sampled, features were extracted in this research using forearm motion data acquired at 20.0 Hz, which leaves plenty of margin for the 0.6 Hz maximum frequency of eating movements.

3) Applying the BoW Representation

A codebook*14 is created to record the primitive motions (words) from the extracted features. The total amount of words in the codebook is referred to as a "Vocabulary". When applying a BoW representation to motion recognition, a Vocabulary with comparatively low values is effective [11]. Hence, we applied a Vocabulary = 20 BoW representation to the feature extracted from the forearm motion data.

4) Application of N-gram

After creating the BoW representation, by applying the N-gram focusing on N words that occurred, we attempted to express actions such as slurping noodles or coiling pasta with a fork, which make sense with time width. Because the combinations of words increases explosively with the application of N-gram, we adopted the N=3 Trigram*15 for this research. This value was set based on the duration of actions such as slurping noodles or coiling pasta with a fork, and used to express a 2 second sample by moving a 1 second moving window in 0.5 second increments for the forearm motion data subsampled*16 at 20.0 Hz.

4.2 Evaluations

In this experiment, the NB classifier learned from the collected forearm motion data and the estimation performance was measured. We used the Leave-One-Out Cross Validation (LOOCV)*17 method for evaluating estimation performance. With this method, the forearm motion data collected for one meal is used for evaluation, while the remaining data is used for classifier learning. Evaluation indicators are a general f1 measure*18, the harmonic mean value between Recall*19 and Precision*20, and Accuracy, the ratio of the total number of true positives*21 and true negatives*22 for each meal content to the total number of forearm motion data.

Figure 4 shows a confusion matrix of LOOCV. In this table, the true values of meal contents to which forearm motion data belong are shown in the vertical direction, while the results of classifier estimation are shown in the horizontal direction. The closer the background color is to black, the higher the percentage of estimated result accuracy. It can be seen from the confusion matrix that the results of estimation for noodles are higher than other meal contents. In contrast, there were a significant number of cases of curry rice mistakenly estimated to be pasta or noodles, because in this experiment, we did not use a gyro sensor that could directly measure rotation of the wrist so it was difficult to reflect wrist rotation. Specifically, the action of eating curry rice entails gathering the rice and curry roux and then carrying it to the mouth. With the features adopted in Table 1, the actions of gathering the curry roux and rice and carrying them to the mouth seem to be seen as the same action, which means distinguishing those

*14 Codebook: An index of all words used for BoW representations. Also called a "dictionary."

*15 Trigram: The name given to an N-gram where N=3. N-grams with N=2 and N=1 are called bigrams and unigrams, respectively.

*16 Subsampling: Extracting certain portions of sensor data according to certain rules.

*17 LOOCV: A method of splitting data used for classifier evaluation. Because all the data is used for evaluation, this method is often used when the total amount of data is small.

*18 f1 measure: A combination index derived by finding the harmonic mean of both the Recall and Precision indexes. While classifiers with both high Recall and Precision are ideal, there is a trade-off between the indexes. For this reason, the f1 measure is often used because it enables combined evaluation rather than evaluation using the individual indexes.

actions from other meal contents can be problematic, hence there were many estimates mistaken for noodles, which prior probability[*23] is the highest.

As meal content estimation results, **Figure 5** shows f1 measures, and **Figure 6** shows Accuracy. The average f1 measure for meal contents is 63%, and the average Accuracy is 72%. Since Recall and Precision are remarkably low for curry rice, both the f1 measure and Accuracy have the lowest estimation performance. Other meal contents have f1 measures of 65% or more, and Accuracy of 70%

		Estimation result				
		Donburi	Curry	Breads	Pasta	Noodles
True value	Donburi	0.55	0.00	0.00	0.17	0.28
	Curry	0.16	0.09	0.00	0.34	0.41
	Breads	0.04	0.04	0.68	0.12	0.16
	Pasta	0.19	0.04	0.04	0.62	0.08
	Noodles	0.03	0.00	0.00	0.06	0.91

Figure 4 Meal content estimation: Confusion matrix

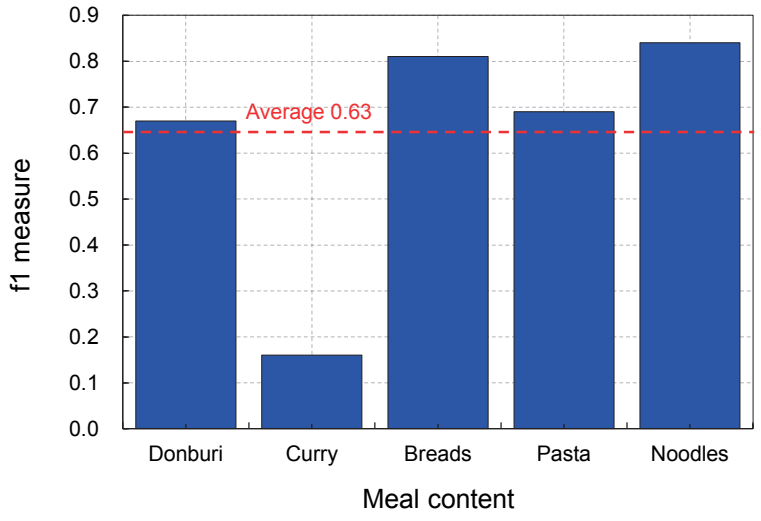


Figure 5 Meal content estimation results (f1 measure)

*19 Recall: Expresses comprehensiveness as a lack of leakage with estimation results, but cannot express precision of estimation results.

*20 Precision: An index that can express the accuracy of estimation results, but cannot express the comprehensiveness of estimation results.

*21 True positive: Estimation results that match actual actions taken, from among the estimation results matching true values. For example, when actually eating ramen noodles, the number of events that are estimated to be ramen noodle events.

*22 True negative: Estimation results that match actual actions not taken, from among the estimation matching true values. For example, if meal content is not ramen noodles, the number of events that are estimated not to be ramen noodles.

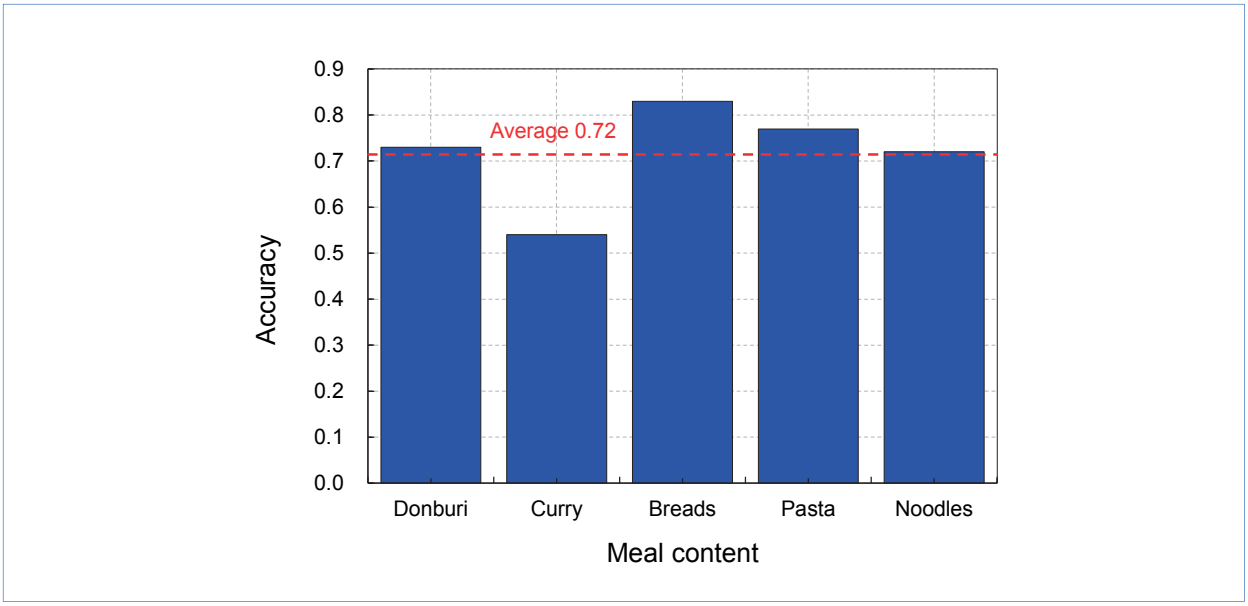


Figure 6 Meal content estimation results (Accuracy)

or more, which indicates that it's possible to estimate meal contents from forearm motion data in situations where food items are limited to some degree.

5. Conclusion

Aiming to produce technology that can automatically and continually grasp meal content without any hassle to the user, we have studied a meal content estimation method that focuses on forearm motion. In experimental conditions, the recognition performance of this proposed method achieved an average f1 measure of 63%, and an average Accuracy of 72%. These results indicate that it's possible to estimate meal content in situations where food items are relatively limited such as workplace canteens. We intend to work towards commercializing services with this technology by improving its recognition accuracy so that it can be used with a wider range of food items such as those in restaurants or at home. Features need to be added that can express forearm motion in more detail, because, with the general statistics we adopted for features, the estimation accuracy was low for some meal contents. We also believe formulating recognition models and adjusting features for individuals will be effective as we found differences in the forearm motions of experimental subjects even with the same meal contents.

REFERENCES

[1] Ministry of Agriculture, Forestry and Fisheries: "Dietary guidelines," (In Japanese). http://www.maff.go.jp/j/syokuiku/shishinn.html
[2] Ministry of Health, Labor, and Welfare: "Balanced meal guide," (In Japanese). http://www.mhlw.go.jp/bunya/kenkou/eiyou-syokuji.html

*23 Prior probability: In Bayesian probability, a theory of probability that an event will occur based on changes in the amount of knowledge related to that event. Before acquiring the knowledge, the assumed probability that an event will occur is called prior probability, while the probability after knowledge is acquired is called posterior probability.

[3] Cabinet Office: "Survey Report on nutrition education (March 2016)," (In Japanese).
http://www.maff.go.jp/j/syokuiku/ishiki/h28/

[4] Ministry of Health, Labour, and Welfare: "2014 Health, Labor and Welfare White Paper - Toward the realization of a healthy, long-living society - first year of health and prevention," (In Japanese).
http://www.mhlw.go.jp/wp/hakusyo/kousei/14/

[5] NTT DOCOMO: "d healthcare pack," (In Japanese).
https://www.nttdocomo.co.jp/service/dmarket/healthcare/

[6] Y. Dong, J. Scisco, M. Wilson, E. Muth and A. Hoover: "Detecting Periods of Eating During Free-Living by Tracking Wrist Motion," IEEE Journal of Biomedical and Health Informatics, Vol.18, No.4, pp.1253-1260, Sep. 2013.

[7] TDK Corporation: "SilmeeW20," (In Japanese).
https://product.tdk.com/info/ja/products/biosensor/biosensor/silmee_w20/index.html

[8] K. Aizawa, Y. Maruyama, L. He and C. Morikawa: "Food Balance Estimation by Using Personal Dietary Tendencies in a Multimedia Food Log," IEEE Transaction on Multimedia, Vol.15, No 8, pp.2176-2185, Dec. 2013.

[9] O. Amft, M. Kusserow and G. Tröster: "Probabilistic parsing of dietary activity events," International Workshop on Wearable and Implantable Body Sensor Networks, Vol.13, pp.242-247, 2007.

[10] O. Amft and G. Tröster: "On-Body Sensing Solutions for Automatic Dietary Monitoring," IEEE Pervasive Computing, Vol.8, No.2, pp.62-70, 2009.

[11] M. Zhang and A. A. Sawchuk: "Motion Primitive-Based Human Activity Recognition Using a Bag-of-Features Approach," Proc. of the 2nd ACM SIGHIT International Health Informatics Symposium, pp.631-640, Jan. 2012.

[12] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell and B. G. Celler: "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," IEEE Trans. on Information Technology in Biomedicine, Vol.10, No.1, pp.156-167, Jan. 2006.

[13] K. Yano and H. Kuriyama: "Humans × sensors" - sensor information changing people, organizations and society," Hitachi Review, Vol.89, No.07, pp.62-67, Jul. 2007 (In Japanese).