M-019 Food intake monitoring based on acoustic and acceleration information

Ahmad Yeaseen Khan[1] Masafumi Nishimura[1]

Abstract— The health condition of a person is closely related to his /her food intaking behavior. From this perspective, we tried to develop a system to monitor a person's food intaking behavior. By monitoring food intaking behavior we mean detection of chewing strokes, swallowing events, hand movements and using this information to estimate the number of food picking events and water container picking events. The detection of food picking events is useful to measure the speed of food intake behavior. And the detection of water picking events can help to monitor the water drinking habit of the subjects. The acoustic information was gathered by a skin contact microphone is placed on the subjects under the ear position. Acceleration data was captured by an IMU (Inertial Measurement Unit) sensor placed on the subject's wrist position. We used Convolutional Recurrent Neural Network (CRNN) model for classification and detection of the chewing strokes, swallowing events and hand movements. By utilizing the acceleration and acoustic information we are able to detect number of food picking and water picking events.

I. Introduction

Nowadays, monitoring of food intake behavior has great importance. Obesity has become a problem worldwide. That is why much research is conducted in this field. In [1] acoustic information was used to train Long Short-Term Memory (LSTM) network. In [2] a frame-based method was proposed to train a combination of GMM (Gaussian mixture model) and LSTM for classifying speaking and food intaking events using limited amount of training data. In [3] and [4], LSTM-CTC based approach was also used to detect chewing strokes and swallowing events. An early approach was made in [5] to detect food intaking courses by Support Vector Machine (SVM) using acceleration information acquired by a sensor positioned in wrist position. CRNN based approach was also used to classify six types of wrist movement during food intake in [6]. In this paper, we propose to utilize the information of hand movements as well as the chewing and swallowing. We used frame-by-frame approach to the acceleration information gathered from the IMU sensor in the wrist position. And after the frame-by-frame classification of chewing, swallowing and hand movements with using the CRNN we tried to utilize the information to detect water picking and food picking events. Food picking and water picking events are closely related to our eating habit and this method will help to detect them.

II. Propose Method

A. Data Acquisition

Total 10 subjects and 10 food intake sessions were included on the experiment. By each session we mean each period of food intake. Each subject participated in 1 food intake session. In each session the subject had to intake the food item and drink water. As food items apples and cookies were used. We also included water during food intake sessions. An IMU(WT901BLECL) sensor was installed in each subject's wrist position and a skin contact condenser microphone was installed under the ear position. The IMU sensor was used to acquire acceleration information and the microphone was used for acoustic information. The Figure I illustrates the installation of IMU sensor and microphone. We recorded of total 96 minutes of both acceleration and acoustic information. The acceleration information acquired from the wrist position is used to detect wrist up and down movements. The acoustic information is used to detect chewing and swallowing events. The sampling rate for acceleration data is 100Hz. The acceleration data is 9 dimensional contains the angular movement, acceleration, and angular acceleration in the X, Y, and Z-axis. The acceleration information is classified into three classes as wrist moving from food item/water container to mouth, wrist moving from mouth to food item/water container and others. The others label contains small movements of wrist and wrists in stable position. The frame size for acceleration information is 3sec and the frame shift is 1s. The acceleration data contains 398 Wrist moving from food item/water container to mouth and 398 Wrist moving from mouth to food item/water container events. Condenser microphone positioned under the ear position was used to gather acoustic information. The sampling rate of audio data is 8000Hz. The acoustic data is classified into three classes as chewing, swallowing, and silence. The frame size for acoustic information is 300ms and the frame shift is 100ms. The dataset contains a total of 552 swallowing and 2167 chewing events. We used total 7 sessions of 7 subjects both acoustic and acceleration information for training purpose. For testing purpose, we have used 3 sessions of 3 different subjects both acoustic and acceleration information. The training dataset of acceleration information contains 270 wrists moving from food item/water container to mouth and 270 wrists moving from mouth to food item/water container events. The testing dataset for acceleration information contains 128 events of both labels. The training dataset form acoustic information contains 1516 chewing and 386 swallowing events. And the testing dataset of acoustic information contains 651 chewing and 166 swallowing events. The target number food container picking event in the test dataset is 46 and the target number of water container picking event in the test dataset is 34. We have manually labeled the dataset of both acoustic and acceleration information. As feature extraction technique for audio information, we used a 26-dimensional log filter bank. In case of acceleration information, we used the IMU sensor data without any feature extraction technique. For detection of food picking and water container picking events we utilized both acoustic and acceleration information.

Figure I: Installation of IMU sensor on wrist position (left) and microphone on under the Ear Position (Right)

B. Frame Based Evaluation of Acoustic and Acceleration Information

The CRNN model is used for training and evaluation of acoustic and acceleration information. Two layers of convolution are followed by max-pooling. After the max pooling, two layers of LSTM units are used. The convolution layers are used to extract the features from the dataset and the LSTM units will summarize the data. In Figure II the structure of the model is shown in detail.

[1] Shizuoka University

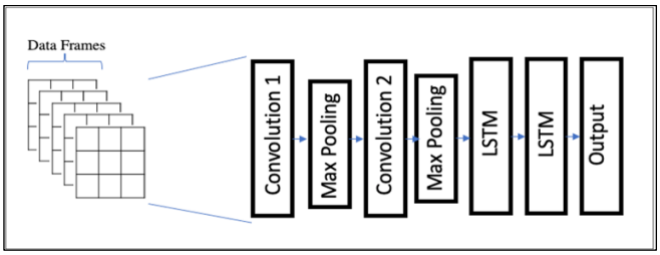

Figure II: The detail of the CRNN model.

In Table I the frame-based evaluation result for acoustic information is presented with precision, recall and F-1 score.

TABLE I. FRAME BASED DETECTION RESULT CHEWING AND SWALLOWING EVENTS

Feature Type	Data Type	Precision	Recall	F-1 score
Log filter bank	Chewing	0.99	0.96	0.98
	Swallowing	0.85	0.99	0.91

We used another CRNN model of different input shape for the classification of acceleration information. This CRNN model can classify wrist moving to mouth events and wrist moving to food items from mouth events. The frame-based result for IMU data is presented in Table II.

TABLE II. FRAME BASED DETECTION RESULT WRIST MOVEMENT

Event	Precision	Recall	F-1 score
Wrist moving from food item/water container to mouth	0.91	0.92	0.91
Wrist moving from mouth to food item/water container	0.90	0.93	0.91

C. *Event Based Evaluation of Food and Water Container Picking Events*

We apply the CRNN models to the test datasets of both acceleration and acoustic information. The models detect the chewing and swallowing events form acoustic information and wrist movements from acceleration information. By food picking events we mean picking up the food to the mouth and intake of the food item. By water container picking event we mean picking up the water container and drinking water from it. Both acoustic and acceleration events are used to detect and classify the number of water container picking event and food item picking event. On a food picking event a wrist moving to mouth event following bites, chewing, swallowing and wrist moving to food events occur. On a water container picking event a wrist moving to mouth event following several swallowing and wrist moving to container event occur. The chewing and swallowing events can only be detected from acoustic information. And wrist movements can be detected from acceleration information. That's why we have used both the acoustic and acceleration both information to detect the food and water container picking events.

We wrote a program on Python programming language to detect the food picking events and water container picking events. After a wrist moving from food item/water container to mouth if we observe a chewing event first and then there are several chewing and swallowing events before wrist moving to the food item/ water container the program detects this total sequence of events as a food picking event. And after a wrist moving from food item/water container to mouth if we observe several swallowing events and no chewing event before wrist moving to the food item/ water container the program detects this total sequence of events as a water container picking event. The result for food picking events and water container picking events are presented in Table III.

TABLE III. EVENT BASED DETECTION RESULT FOR FOOD AND WATER CONTAINER PICKING EVENTS

Event	Precision	Recall	F-1 score
Food picking events	0.92	1.00	0.95
Water container picking events	1.00	1.00	1.00

From the result we can see that the proposed method can detect and classify food and water container picking events. Though the dataset size was limited, the results was not bad.

III. CONCLUSION

We are also planning to collect more data from many other subjects. The data size is very limited but feasible results were obtained for understanding the food intaking behavior. In future we can update the method to distinguish water swallowing events from other food item swallows and we can get an idea about water swallowing during food intake sessions.

REFERENCES

[1] Dzung Tri Nguyen, Eli Cohen, Mohammad Pourhomayoun, Nabil Alshurafa, "SwallowNet: Recurrent Neural Network Detects and Characterizes Eating Pattern," Second IEEE PerCom Workshop on Pervasive Health Technologies, 2017.

[2] Jumpei Ando, Takato Saito, Satoshi Kawasaki, Masaji Katagiri, Daizo Ikeda, Hiroshi Mineno, Takashi Tsunakawa, Masafumi Nishida and Masafumi Nishimura, "Dietary and Conversational Behavior Monitoring by Using Sound Information," NCSP 2018, 2018, pp.675-678.

[3] Akihiro Nakamura, Takato Saito, Daizo Ikeda and Ken Ohta, Hiroshi Mineno and Masafumi Nishimura, "Automatic Detection of the Chewing Side Using Two channel Recordings under the Ear," IEEE LifeTech 2020, March 2020 pp. 82-83.

[4] M. M. Billah et al., "Estimation of Number of Chewing Strokes and Swallowing Events by Using LSTM-CTC and Throat Microphone," 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 2019, pp. 920-921.

[5] Christos Maramis, Vasilis Kilintzis, and Nicos Maglaveras. "Real-time bite detection from smartwatch orien- tation sensor data". Proceedings of the 9th Hellenic Conference on Artificial Intelligence. 2016, pp. 1–4.

[6] Konstantinos Kyritsis et al. "Automated analysis of in meal eating behavior using a commercial wristbßßand IMU sensor". 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017, pp. 2843–2846