

Food intake gesture monitoring system based-on depth sensor

Muhammad Fuad bin Kassim, Mohd Norzali Haji Mohd

Department of Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Malaysia

Article Info

Article history:

Received Dec 12, 2018

Revised Jan 1, 2019

Accepted Feb 4, 2019

Keywords:

Bite count

Depth camera

Food tracking

Gesture recognition

ABSTRACT

Food intake gesture technology is one of a new strategy for obesity people managing their health care while saving their time and money. This approach involves combining face and hand joint point for monitoring food intake of a user using Kinect Xbox One camera sensor. Rather than counting calories, scientists at Brigham Young University found dieters who eager to reduce their number of daily bites by 20 to 30 percent lost around two kilograms a month, regardless of what they ate [1]. Research studies showed that most of the methods used to count bite are worn type devices which has high false alarm ratio. Today trend is going toward the non-wearable device. This sensor is used to capture skeletal data of user while eating and train the data to capture the motion and movement while eating. There are specific joint to be capture such as Jaw face point and wrist roll joint. Overall accuracy is around 94%. Basically, this increase in the overall recognition rate of this system.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Muhammad Fuad bin Kassim,

Department of Electronic Engineering,

Universiti Tun Hussein Onn Malaysia (UTHM), 86400 Parit Raja, Batu Pahat Johor, Malaysia.

Email: fuadkassim93@gmail.com

1. INTRODUCTION

Unhealthy eating habits have come through to the obesity outbreak in Malaysia. You can uphold a fine weight and avoid certain health problems by establishing healthy eating habits. Basically, the person who is overweight and obese can deal with diet and exercise. Counting and try to lower calories intake can be a useful approach to weight loss. There are other approaches method to weight loss for those who don't want to count calories such as bite count. However, measuring and counting kilocalories in daily life is uninspiring and prone to error because not all food that buys at grocery or shop has label calories on the food. This paper describes a method for measuring bite count of food intake using a Kinect camera that tracking users with Kinect skeletal hand tracking of skeleton and face gesture expression. The system consists of a simple algorithm that can tell user and alert of how much food intake taken by detecting hand wrist roll rotation and face jaw movement while eating. Furthermore, the system can help people create lifelong normal eating patterns thus prevent obesity from increase quickly.

Most of the method used previously to track food intake was the wearable worn type. a Today's trend now going towards non-wearable equipment since worn type device has a high false alarm and difficult to use which is not applicable in real life. A study in [2] presented an approach which is based on body-worn sensors and mobile health technology. While [3] presented an estimate of an individual's kilocalorie intake using bite count and mean kilocalories per bite determined by a formula based on demographic and physical characteristics using wrist motion tracker. A novel method for measuring eating activity in free-living settings which also used wrist motion tracker to detect bite count while eating using non-wearable were presented in [4]. Next, a method to count bite with Participants by using Piezoelectric film strain sensor and throat microphone showed in [5, 6]. There also a wrist-worn device that can record 3-axis accelerometer and gyroscope [7].

Besides, detection accuracy by combining using a head-mount accelerometer which is Google Glass technology and using a wrist-worn accelerometer (Pebble watch) [8]. The depth sensing approach is an action to detect gesture of body skeleton or hand skeleton based on [9-13] they use Kinect version 1 which has limited body joint track and using additional sensor such as Myo band combine with RGB from Kinect sensor to achieve accurate result. This system can help people follow a proper way to overcome overweight or eating disorders by monitoring their meal intake and controlling eating rate. The bite counter is measured by using wrist joint and jaw motion detection when taking food. By detecting this pattern, the food intake can be identified when a bite of food has been taken thus it can monitor food intake in real time monitoring while showing bite count to the user. The system could tell the user to slow down or to stop eating after a bite count threshold reached.

These methods also help the user to track long-term eating patterns and can help people track their daily intake. Generally, fingers are aimed downwards to pick something up and roll the hand to place it into the mouth [14]. This pattern holds regardless of the type of food. The aim of this research is to keep the complexity minimum so that it could be potentially applied in real time.

2. RESEARCH METHOD

In this section, it is explained the research method and at the same time is given the comprehensive discussion to achieve the objectives of this paper.

2.1. Overview of the monitoring system

The proposed methodology is divided into 2 stages, which are hand skeletal recognition and face recognition. This project is developed by using Microsoft Visual Studio, this system uses Kinect sensor for capturing depth images. The depth camera can work both day and night [15, 16]. The system monitoring was placed in front of user table with 1.4 meters while taking user daily food intake. Kinect will track skeleton body of the user and recognize the eating pattern.

The gesture recognition of pitch, roll, and yaw are measured using Kinect sensor by detecting the joint rotation point. During food intake, the pitch and roll data arm can be used to trigger food intake detection. Thus, by capturing this data could be used for tracking and recognition. Therefore, the system can detect food intake without using wearable sensors such as Myo armband or gyroscope. Figure 1 shows all the system for food intake monitoring.

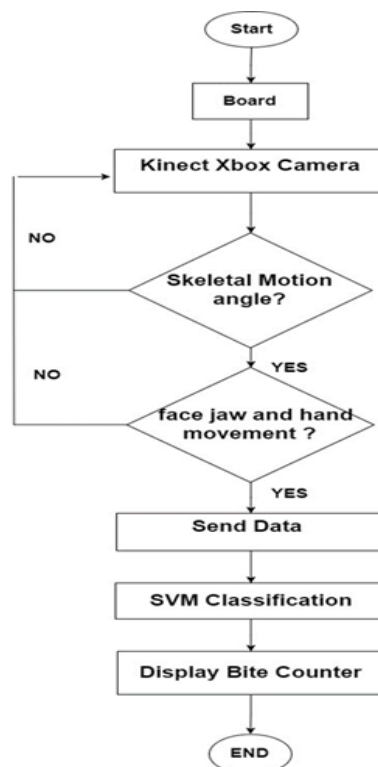


Figure 1. Flowchart of the food intake monitoring system

2.2. Gesture detection and distance measurement

As discussed, this system uses Kinect for capturing depth images and skeletal joint tracking. The depth images are generated by the IR sensor. Using skeletal tracking function, Kinect can detect movement of the arm and the angle position of the user.

The Kinect sensor used here can detect location of 25 joints on each body for six people using Kinect SDK. This function to track the skeleton image within the Kinect's field of view using the infrared (IR) camera. In default range mode, Kinect can track people standing between 0.8 meters to 4.0 meters away, therefore, user able to use their arm at that distance and allowing the recognition of body parts to be tracked.

The Kinect sensor was placed at a specific distance which is around 1.4 meters from the user eating table, at a height of 1.4 meters from the floor. The distance was made to avoid overlapping of the hands and face detection while eating and like the position for which the sensor was designed in front of the user eating place.

The lunch activity basically consisted of eating specific food and drinking water. The user will have instructed to eat only during the sessions, using foods given that are easy to grasp by using a hand. The shoulder joint can be tracked using a point from angle calculation from the depth camera. The skeletal tracking will detect the angle of upper limb joint and the data will be training with Support Vector Machine (SVM) to classify the posture. The upper limb joint is divided into the shoulder, arm elbow, wrist, and hand. Figure 2 shows system structure design of our project. SVM classification will be used to predict the eating posture of the user. The system will use internet dataset which available online to train the data.

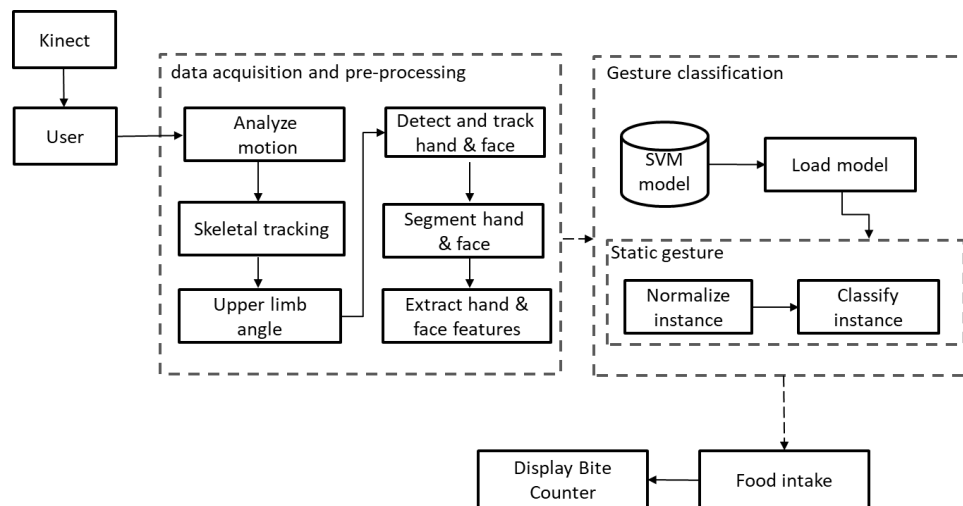


Figure 2. System structure design of food intake counting

2.3. Food intake gesture analysis

Food classification gesture analysis features can be detected by each person's eating patten style by monitoring and training the data. First, the user will be asked to seat in a chair while eating this is to obtain skeleton's joints from the Kinect sensor and it was observed that the joints can be detected and be used to train SVM for seated position recognition. The user's arm and shoulder joints, which represent the key-points of this data. The system will analyze and monitoring the situation of user's hand if it close to mouth and counting the food intake by displaying the bite counter in the GUI of the system. The system will be able to differentiate between drinking and eating.

To analyze the gestures, the system will focus on the head part of user which is the jaw movement and hand wrist joints gesture. This method should help the system differentiate whether the hand is near of far from mouth and monitor the jaw point movement while eating thus can use this point for tracking and monitoring for food intake. To avoid overlapping, the decision technique will be used as a stated in the flowchart. This distance allows us to better understand the variation of distance hand and the jaw movement.

Figure 3 shows the illustration of right arm angle calculation by Kinect. Kinect's 3-dimensional coordinate system defines the three-dimensional coordinate of the shoulder, elbow, and wrist. And uses the vector formula.

$$\overline{AB} = B(X_b, Y_b, Z_b) - A(X_a, Y_a, Z_a) \quad (1)$$

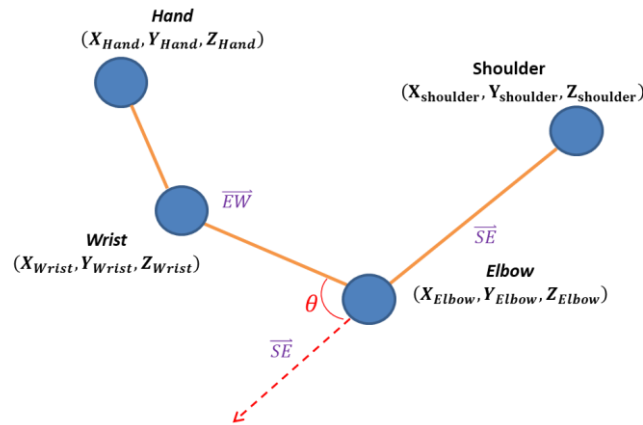


Figure 3. Illustration of right arm angle calculation by Kinect

To find the vector formula of shoulder to elbow and of the elbow to wrist. The vector formula is shown below.

$$\overline{SE} = Elbow(X_{Elbow}, Y_{Elbow}, Z_{Elbow}) - Shoulder(X_{Shoulder}, Y_{Shoulder}, Z_{Shoulder}) \quad (2)$$

$$\overline{EW} = Wrist(X_{Wrist}, Y_{Wrist}, Z_{Wrist}) - Elbow(X_{Elbow}, Y_{Elbow}, Z_{Elbow}) \quad (3)$$

After solving the vector formulas between joints, the elbow bending angle could be found by the law of cosines and anti-law of cosines of vector formula. The formula is shown below.

$$\cos \theta = \frac{\overline{SE} \cdot \overline{EW}}{|\overline{SE}| |\overline{EW}|} \quad (4)$$

$$\theta = \cos^{-1}(\cos \theta) \quad (5)$$

Data were also taken of the movements of right hands when they leave the plate in a direction to the mouth and from the body to the mouth. The bite counters can count the total number of bites the user has taken and provided the rate of bites taken (bites per minute) of the user.

The data can be stored to review and evaluate the device later. Kinect can classify the bite count and drink. The SVM will be used to train the bite pattern of user eating style. This system will have monitoring GUI system so that user can see their daily intake or can be using to alert the user. The user will not have to wear any sensor on their body. The goal of this system to provide the user with motivational information and identify bite behavior to help the user lose weight effectively.

2.4. Rotation 3D hand wrist joint recognition

To create a rotation in 3-dimension, the axis and position of rotation need to be analyzed. The X, Y, and Z of user's right-hand wrist point are reported based on coordinate system where the origin of the sensor and user are specified with Kinect's skeleton coordinate frame.

Translations are in meters. The user's hand wrist rotation is captured by three angles which is pitch, roll, and yaw. This can be used for tracking and monitoring the rotation of the hand wrist around the X, Y and Z axis. Euler angle are used to represent the rotation in 3-dimension space.

A quaternion is a set of four vector that are used to specify a rotation in 3D space, quaternion will be used to detect these 4 vectors in hand wrist rotation that is known as joint yaw, pitch and roll. Each quaternion is absolute to its parent bone. The frame-based approach is to provide a continuous real time hand wrist rotation by utilizing depth by Kinect sensor. Our approach built is to analyze the accuracy rotation of wrist hand since most of the method to track hand wrist using wearable gyroscope but now with Kinect the same function as gyroscope available by monitoring the roll, pitch and yaw of hand wrist. Figure 4 shows the illustration of rotation in 3 dimensions by Kinect sensor.

Figure 5 shows a hand gesture tracking using the Kinect sensor. Hand gesture system is widely used in today's trend since it only detecting hand, finger, and gesture making it easier to identify the target. Depth image from Kinect sensor is used to detect the Skelton's hand. The finger tracking should be able to track both hands. The first step in finger tracking algorithm is to detect the hand joints. Next, specify the search area and then find the contour of hand to specify only depth value of the hand and lastly, the convex hull which is the point of valid fingertips. The fingertips are the edges of a polyline that contains all the contour points thus the data can be used for implementing in gesture based food tracking while eating.

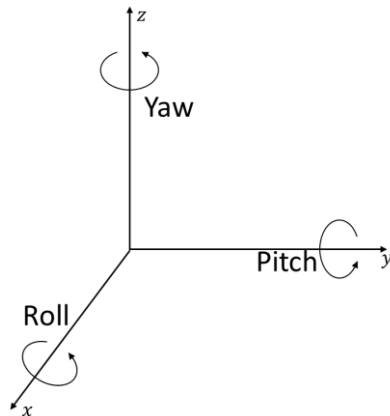


Figure 4. Illustration of rotation in 3 dimensions by Kinect sensor



Figure 5. Hand gesture tracking

3. RESULTS AND ANALYSIS

Bite detections were classified as true positive (TP), false negative (FN), or false positive (FP). Bites were considered as true positive, when the system recognize a food intake detection when the arm is near raised near to mouth when taking a food. Any detections falling outside these or duplicate detections within a single process were marked as false positive.

The food intake tracking is an event that occur at a point along with a continuous duration of timeline when taking a bite therefore this cannot be consider as discrete points which cannot classified as binary as bite taken or not thus there is no way to justify true negatives result. In gesture detection module the test will be performed for each type of meal. The TP is the eating gesture that system detects, and the user performed the gesture while the FP is the eating gesture that system detects but participant does not perform the gesture and lastly, FN is the gesture where the system does not detect but participant performed gesture. Table 1 present the bite evaluation detection system.

Table 1. Bite evaluation detection system

Accuracy=94%	Eating	Sit/Rest	Drink	Wrist Roll
Eating	48	0	2	0
Sit/Rest	0	50	0	0
Drink	10	0	40	0
Wrist Roll	0	0	0	50
Sensitivity	96%	100%	80%	100%

Table 1 shows the table of confusion matrix of food intake by Kinect. Confusion matrix for prediction performance on data from the system. Correct predictions are given on the diagonal, and the sensitivity is display below. The recognition rates obtained are shown in the form of confusion matrices for the ten studied gestures of "Eating", "Sit/rest", "Drink" and "Wrist Roll". Table above corresponds to the situation when using the Kinect system GUI. It is seen that the overall accuracy is around 94%. Basically, this increase in the overall recognition rate of this system.

Figure 6 shows the data obtained from 5 different opponents with age of 25, 26, 23, 24 and 44. The user is asked to be seated in a chair and the Kinect obtained their upper limb angle detection. The Kinect is positioned in front of the opponent, at 1.4 meters from the user. The Kinect will automatically detect user's

sitting skeleton and calculated upper limb joint. Although the angle detected is not accurately same with another opponent because there is some factor which not same of all user, for example, the body size, therefore, SVM classification can be used to train Kinect to detect angle when seated. Datasets such as Kinect Activity Recognition Dataset (KARD) and Cornell Activity Dataset (CAD-60) which are available online to train the SVM for classification. The data angle captured still in the range of 200 to 250 degree which can be used to recognize the automatic sitting position of the user when eating.

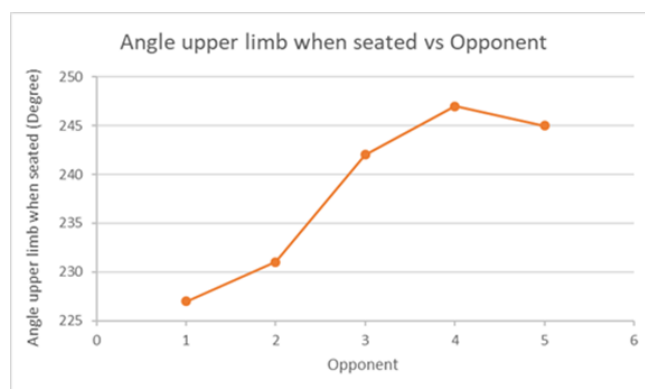


Figure 6. Data of angle upper limb when seated vs opponent

4. CONCLUSION AND FUTURE WORK

This research involves detecting the bite count of eating motion to control the over intake of food leading towards obesity. A bite-based measure of kilocalorie intake shows for individual use for self-monitoring to use for monitoring free-living kilocalorie intake. It is an easily collected and based on the motion that could be refined to more accurately estimate kilocalorie intake.

The recent development of new sensors that allow tracking important parts of the human body resulted in a proliferation of different approaches to gesture recognition and their practical applications. Therefore, any new improvement of previous solutions towards better recognition accuracy or shortening recognition time is better for this project to be develop.

In the future, some defect in a proposed study of food intake can be improved such as food intake algorithm there is still some high probability that false detection may occur. To reduce false detection and undetected food tracking there several things to consider. For example, instead of using SVM other machine learning can be used such as deep learning and more new technology.

The system GUI also can be improved such as adding new automatic counting calories with number bites based on depth sensor. The system can also have database 'cloud' system which can be accessed using phone camera that has depth camera specification making the system portable and can be carrying anywhere using smartphone.

ACKNOWLEDGEMENTS

This research is supported by a Fundamental Research Grant Scheme (FRGS) VOT 1580 sponsored by the Center for Graduate Studies UTHM and KPT. The research also received funding from the Office for Research, Innovation, Commercialization and Consultancy Management (ORICC), UTHM.

REFERENCES

- [1] "Bits & bites | research and creative discovery | Clemson University." [Online]. Available: <http://glimpse.clemson.edu/bits-bites/>. [Accessed: 13-Dec-2017].
- [2] Z. Huang, "An Assessment of the Accuracy of an Automated Bite Counting Method in a Cafeteria Setting," Dissertation, no. August, 2013.
- [3] J. N. Salley, A. W. Hoover, M. L. Wilson, and E. R. Muth, "Comparison between Human and Bite-Based Methods of Estimating Caloric Intake," *J. Acad. Nutr. Diet.*, vol. 116, no. 10, pp. 1568–1577, 2016.
- [4] J. L. Scisco, E. R. Muth, and A. W. Hoover, "Examining the utility of a bite-count-based measure of eating activity in free-living human beings," *J. Acad. Nutr. Diet.*, vol. 114, no. 3, pp. 464–469, 2014.
- [5] J. M. Fontana and E. S. Sazonov, "Evaluation of Chewing and Swallowing Sensors for Monitoring Ingestive Behavior," *Sens. Lett.*, vol. 11, no. 3, pp. 560–565, 2013.

- [6] S. Päßler and W. J. Fischer, "Food intake monitoring: Automated chew event detection in chewing sounds," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 1, pp. 278–289, 2014.
- [7] S. Sharma, P. Jasper, E. Muth and A. Hoover, "Automatic Detection of Periods of Eating Using Wrist Motion Tracking," *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, Washington, DC, 2016, pp. 362-363.
- [8] Xu Ye, Guanling Chen and Yu Cao, "Automatic Eating Detection using head-mount and wrist-worn accelerometers," *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, Boston, MA, 2015, pp. 578-581.
- [9] S. Manton, G. Magerowski, L. Patriarca, and M. Alonso-Alonso, "The 'Smart Dining Table': Automatic Behavioral Tracking of a Meal with a Multi-Touch-Computer," *Front. Psychol.*, vol. 7, 2016.
- [10] L. Costa, P. Trigueiros, and A. Cunha, "Automatic Meal Intake Monitoring Using Hidden Markov Models," *Procedia Comput. Sci.*, vol. 100, pp. 110–117, 2016.
- [11] S. Gaglio, G. L. Re and M. Morana, "Human Activity Recognition Process Using 3-D Posture Data," in *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 586-597, Oct. 2015.
- [12] S. Sen, V. Subbaraju, A. Misra, R. K. Balan and Y. Lee, "The case for smartwatch-based diet monitoring," *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, St. Louis, MO, 2015, pp. 585-590.
- [13] E. Cippitelli, S. Gasparini, E. Gambi and S. Spinsante, "Unobtrusive intake actions monitoring through RGB and depth information fusion," *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Cluj-Napoca, 2016, pp. 19-26.
- [14] M. F. Bin Kassim and M. N. Haji Mohd, "Tracking and Counting Motion for Monitoring Food Intake Based-On Depth Sensor and UDOO Board: A Comprehensive Review," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 226, no. 1.
- [15] Y. Nizam, M. N. H. Mohd, and M. M. A. Jamil, "Development of a user-adaptable human fall detection based on fall risk levels using depth sensor," *Sensors (Switzerland)*, vol. 18, no. 7, 2018.
- [16] Y. Nizam, M. N. H. Mohd, and M. M. A. Jamil, "Human Fall Detection from Depth Images using Position and Velocity of Subject," *Procedia Comput. Sci.*, vol. 105, no. Iris 2016, pp. 131–137, 2017.

BIOGRAPHIES OF AUTHORS



Muhammad Fuad bin kassim is a Master student at Universiti Tun Hussein Onn Malaysia (UTHM). His research interests include Computer Vision, Depth Camera, Image Recognition, and Patern Recognition. He received his bachelor's degree in Electronic Engineering from Universiti Tun Hussein Onn Malaysia (UTHM) in 2016.



Mohd Norzali Haji Mohd is currently working as Senior lecturer, Faculty Lab Manager and Former Industrial Training Coordinator at the Department of Computer Engineering, Faculty of Electrical and Electronic Engineering, Univeristi Tun Hussein Onn Malaysia (UTHM). He received Diploma in Computer Engineering from Toyama Maritime College, Japan in 2000 and B.Eng., M.Eng. from Fukui University, Japan in 2002 and 2004 respectively. In April 2015, he received Ph.D. from the Department of Information Sciences and Biomedical Engineering, Kagoshima University, Japan.