Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 16 (2014) 1383 – 1390

Procedia Technology

CENTERIS 2014 - Conference on ENTERprise Information Systems / ProjMAN 2014 - International Conference on Project MANagement / HCIST 2014 - International Conference on Health and Social Care Information Systems and Technologies

Evaluation of MS Kinect for elderly meal intake monitoring

António Cunha[a,*], Luís Pádua[b], Luís Costa[b], Paula Trigueiros[c]

[a]*University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal and INESC TEC - INESC Technology and Science (formely INESC Porto)*
[b]*University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal*
[c] *UNIDCOM/IADE – Unidade de Investigação em Design e Comunicação, and FEUP - Faculdade de Engenharia da Universidade do Porto*

Abstract

Any form of eating disorder is detrimental for health. Having an eating disorder increases the risks for chronic diseases and general morbidity, leading to several health problems such as obesity, hypertension and cardio-vascular diseases. The risk is greater for elderly people, as ageing submits the body to several functional changes that affect health and nutrition conditions. Automatic monitoring systems can help to prevent these risks by supporting people to maintain appropriate eating behaviours. Ageing services based on ICT assistive services are increasing as a result of the awareness of the growing socio-economic relevance of this issue, especially when we consider the rural and very sparsely-populated areas. In order to assess these requirements, systems should be automatic, non-intrusive and low cost. This paper presents an evaluation test of the Microsoft Kinect sensor for monitoring older people's meal intake, with the aim of contributing to the development of an automatic diet monitoring system.

Keywords: Intake gestures; Microsoft Kinect; eating monitoring; assistive technology; aging

* Corresponding author. Tel.: +35-125-935-0300; fax: +35-125-935-0300.
E-mail address: acunha@utad.pt

doi:10.1016/j.protcy.2014.10.156

1. Introduction

Healthy eating and maintaining energy levels are key factors for a healthy life. Any form of eating disorder is detrimental to health, since they contribute to risks for chronic diseases and general morbidity. This is becoming especially apparent with the current obesity epidemic, which is leading to a rise in several health problems such as hypertension and cardio-vascular diseases.

Ageing is a natural process that submits the body to several functional changes, which affect people's health and nutrition conditions. Some of those changes are progressive, causing a reduction in functional capacity [1].

One form of prevention is to support individuals in maintaining appropriate eating and drinking behaviours. So far, most prevention programmes for diet coaching use techniques for self-reporting eating behaviour, which are an additional burden for participants [2]. Long-term behavioural monitoring and coaching can contribute to achieving or maintaining a healthy lifestyle, as well as reducing the risks of eating disorders [3].

Developed countries have seen an increase in life expectancy, due to factors such as improvements in the quality of healthcare along with better living and working conditions [4]. Subsequently, there is a growing number of elderly people living alone, who depend on the support of a carer (a family member or an institutional service). These personalised services can be expensive, particularly in rural areas where people live far from the care providers.

The northern interior region of Portugal presents a worrying case of an ageing population with very low income, which makes this an important socio-economic issue.

Today, several Information and Communications Technologies (ICT) devices can help provide personalised health and care services that meet individual needs at reasonable prices [5].

However, there are several known barriers to the use of new technologies, such as the fact that many people, and especially the elderly, are still not familiar or comfortable with their use. Consequently, assistive systems that monitor food intake using non-invasive sensors could provide a valuable tool for monitoring these people's diets [6].

Among existing non-invasive sensors, the Microsoft Kinect is an RGB-D sensor that provides synchronised colour and depth images, at a much lower cost than traditional 3-D cameras. The sensor automatically captures the coordinates of standing or seated skeletons of people on scene [7].

In this paper we present an evaluation of the Microsoft Kinect sensor for meal intake monitoring, with a view to contributing to the development of an automatic system for monitoring older people's diets. This work is the first stage of development of a low cost and non-invasive system.

It builds upon related work on monitoring of feeding activities, outlined in section two. The methodology used to obtain data and its sampling is described in section three. In section four, we present an analysis of our results, followed by their discussion in section five. Finally, section six contains our conclusions and directions for future work.

2. Related work

The detection of feeding activity is a hard challenge, given the complexity of eating behaviour. To our knowledge, there is no sensor or device that automatically generates reports on diet monitoring in natural environments. For this reason, researchers and business solutions skip the step of detection and use classic approaches of self-report instead. Feeding monitoring solutions have focused on self-report techniques in electronic diaries, such as mobile devices [8]. However, this approach might not always show the truth and this kind of solution might present difficulties to unexperienced mobile device users [9], [10].

Research into alternative input data methods, such as voice logs and barcode scanning, resulted in estimates with similar errors [11, 12]. Some researchers made attempts for automatic diet monitoring using RFIDs, computer vision and combination of several sensors:

- Patterson et al. [13] used Radio-Frequency IDentification (RFID) tags for identification in 60 household objects and a tag reader placed in the user's hand to trace his morning activities, including the preparation and consumption of breakfast. Chang et al. [14] used a table with RFID sensors to identify food containers and weight sensors to monitor the transport of food from the containers to the plate. While the first approach has the potential to evaluate the feeding time and the type of food, the latter does additionally recording of the food's weight for better meal data. Both approaches require specific object and food labelling for proper identification.

- Gao et al. [15] used a computer vision approach to identify hand to the head movements (bringing cutlery to the mouth) of patients in a nursing home, which required the installation of cameras in the room. The development of approaches with wearable sensors is likely to eliminate these problems, but they are intrusive and affect user routine.
- Hondori el al. [16] used the Microsoft Kinect sensor alongside inertial sensors (in the cutlery and in the mug) to monitor the feeding activity of post-stroke patients. Amft el al. [17] used a combination of body sensors, including a motion sensor jacket, to track intake gestures, a mini microphone to record chewing sounds, and collar-based sensors on the throat to detect swallowing movements.

These systems are intrusive and require that the user wears sensors while eating, so they do not suit our purpose of monitoring elderly people. The system should be completely non-intrusive. As a solution we decided to evaluate a similar approach to Hondori el al. [16], using the Microsoft Kinect sensor without the inertial sensors.

3. Methodology

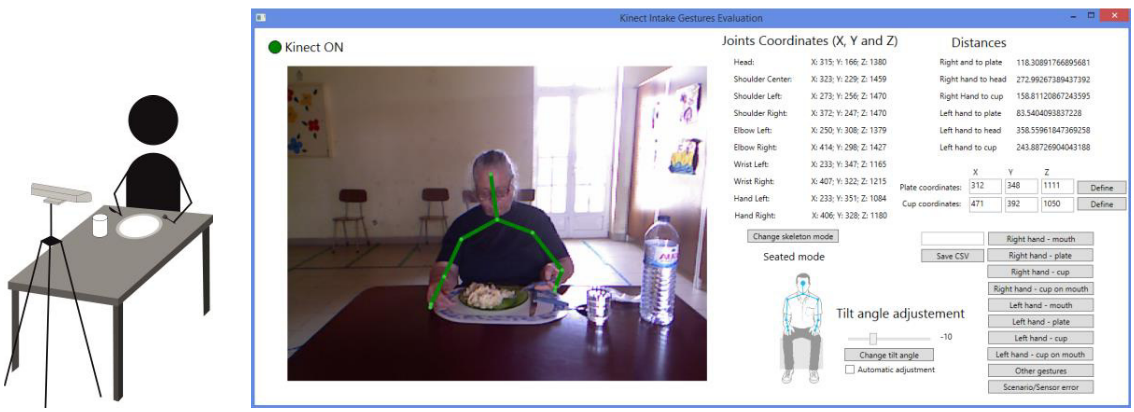
3.1. Approach

The meal intake gestures require upper body movements, using the arms, trunk and head. In the intake phase people use tools, such as knives, forks, spoons, and glasses (or mugs), to aid their movements, the preparation of food, and taking it to the mouth [17]. These gestures can be divided into several movement types: a) eating gestures related to movements done to take food to the mouth, b) drinking gestures related to gestures made to take fluids to the mouth, c) eating-related gestures that are connected with food preparation movements such as cutting food, pouring water, stirring soup, or spreading, and d) other miscellaneous gestures such as wiping the mouth with a napkin or touching a body part [18].

This work is focused on the evaluation of Microsoft Kinect seated skeleton coordinates of a) eating and b) drinking gestures, specifically the intake gestures of the distances “right hand from the plate to the mouth”, “left hand from the plate to the mouth”, “right hand from the glass to the mouth”. We set up a scenario and manually analysed user’s intake gestures and corresponding seated skeleton gestures.

3.2. Experiments

The scenario comprised a table, a plate, a glass, a fork and a knife, and the MS Kinect sensor was positioned in front, as shown in Fig. 1. (a).



(a) (b)

Fig. 1. (a) System scenario; (b) Kinect application to evaluate intake gestures.

The MS Kinect sensor was at an approximate distance of 1.2 meters from the user, at a height of 1.2 meters from the floor, and with a tilt of -10 degrees. This positioning was made in order to avoid occultation of the hands during eating and correspond to the position for which the sensor was designed - in front of the user.

An application was developed, Fig. 1 (b), that used Microsoft Software Development Kit to extract images of the skeleton in seated mode and record its joint XYZ coordinates at a rate of 30 frames per second from the Kinect sensor. The 10 different joints from the upper body are: head (h), shoulder centre (sc), left and right shoulders (sl, sr), left and right hands (hl, hr), left and right wrists (wl, wr), and left and right elbows (el, er).

Three participants ate lunch in this setting. These participants included two elderly women, Maria who was 83 years old and Alice who was 85 years old, and as a reference a young man, Pedro who was 25 years (fictitious names). The lunch activity consisted of eating vegetable soup, a main course (Alice and Maria ate cod with rice, and Pedro ate mashed potatoes with roast beef) and drinking water. The participants had to be alone and were asked to eat normally, not to stare at the camera and not to leave their seat until the meal was finished.

Table 1 presents the frequency of intake gestures for the three experiments.

Table 1. Intake gestures

	Experiment duration (min.:sec.)	Nº right hand plate-mouth	Nº left hand plate-mouth	Nº right hand glass-mouth	Other Gestures	Nº total
Maria	05:32	44	0	4	9	57
Alice	18:37	150	3	2	5	160
Pedro	06:29	31	27	3	2	63
Total	30:38	225	30	9	16	280

4. Intake gestures analysis

Intake gestures can be detected and classified through an analysis of spatial-temporal features; they are affected by each person's individual eating style [2].

First, the seated skeleton's joints obtained from the MS Kinect sensor were analysed and it was observed that the joints are not precisely detected, in relation to the subject. The participant's hand and head joints, which represent the key-points of feeding, were followed in a more precise and stable way than the elbow and wrist joints, which showed greater deviation and unstable behaviour. This difference is due to the fact that the latter suffer an occultation during the movements. It was also observed that in the situations where the participant's hands were still and close to the table, the detected joints showed greater deviations; this ceases to happen when the hands move away from the table.

To analyse the gestures, we chose to focus on the head and hand joints and we defined as discriminating parameters the distance from each hand to the head (Dist_left_hand_head, Dist_righ_hand_head), to the glass (Dist_left_hand_glass, Dist_right_hand_glass), to the plate (Dist_left_hand_plate, Dist_righ_hand_plate) and from the head to the plate (Dist_head_plate). The experiment parameters were represented in charts such as the ones presented in Fig. 2 (a) and (b). (Note that the position of the plate and the glass were always marked manually).

This approach allows us to observe when the hands move closer or further away from the head (and therefore from the mouth), from the plate, and from the glass. In order to eliminate the high frequency peaks, we applied a low-pass-filter (average of eight consecutive values).

Both charts also contain the distance between the head and the plate (Dist_head_plate), because during the analyses we verified that the participants frequently moved their head closer to the plate when they consumed food or fluids. This distance allows us to better understand the variation of the distance between the hands and the head (Dist_left_hand_head, Dist_righ_hand_head).

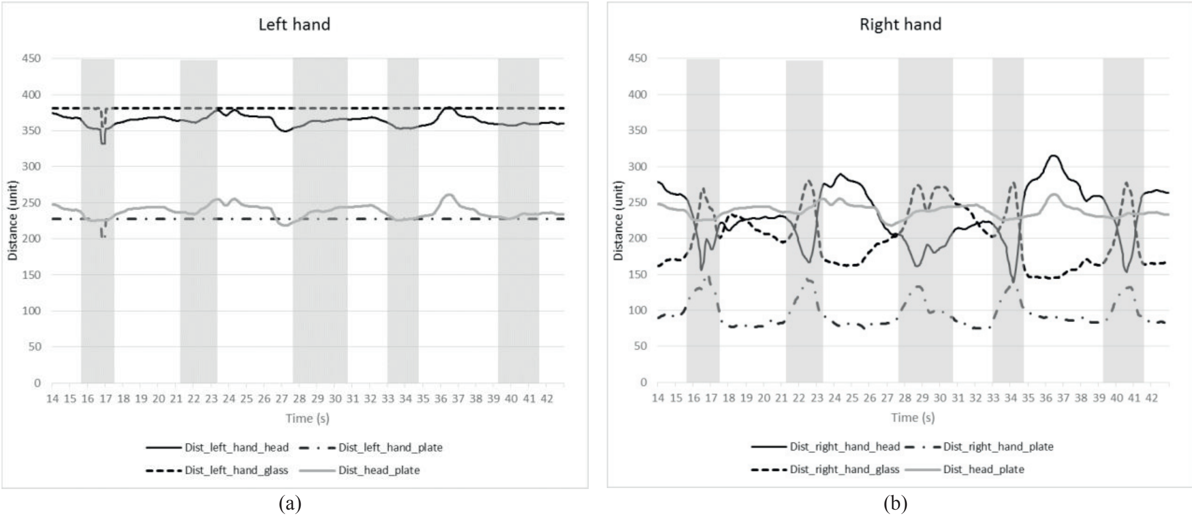


Fig. 2. Maria's intake gestures for eating soup (a) Left hand; (b) Right hand

From observing the distances from the left hand to the plate and from the left hand to the glass (Dist_left_hand_plate, Dist_left_hand_glass) in Fig. 2 (a), we can conclude that Maria kept her left hand still, closer to the plate than to the glass during the 28 seconds shown. On the other hand, the head variation was proportionally reflected in the distance from the left hand to the head (Dist_left_hand_head). From the distances from the right hand to the plate and from the right hand to the glass (Dist_right_hand_plate, Dist_right_hand_glass) in Fig. 2 (a), we can see that Maria proportionally moved her right hand five times closer and further from the plate and the glass, and her right hand was closer to the plate than to the glass; Maria moved her hand smoothly (see smooth curve's slope), which corresponds to eating five spoonfuls of soup (the feeding movements are marked with a grey rectangle in both figures). From the distance from the right hand to the head (Dist_right_hand_head), we can see that the movements were made in the direction of the head and result in an inverse curve, slightly distorted by the variation of the distance from the head to the plate (Dist_head_plate).

The analysis of the three experiments was based on the chart parameters, on the meal videos, and on the 3D charts obtained joints' coordinates (see Fig. 3).

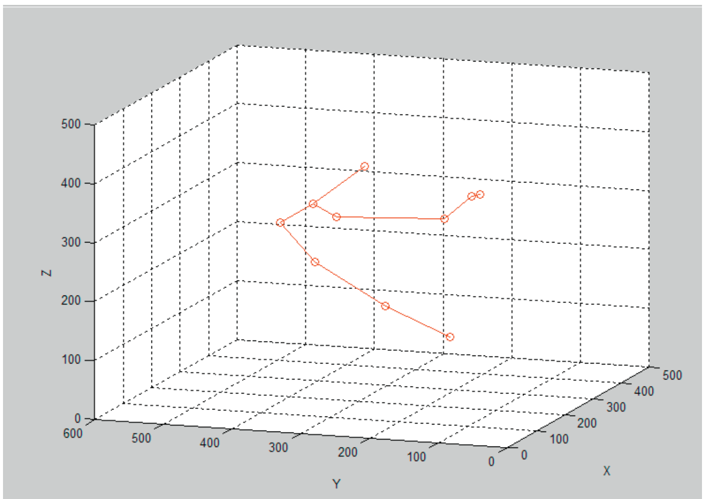


Fig. 3. 3D seated skeleton joints - Maria eating soup in second 23.

Notes were also taken of the movements of both hands when they leave the plate in direction to the head and from the body to the mouth.

During the analysis, we noticed several instances that did not correspond to the predicted scenario setup; for example, at the end of eating the soup, Maria and Alice grabbed the plate with their left hands. The person who served the meals came close to the women, and Maria served water and put the bottle back in a way that occulted her right hand. The following periods were excluded from the analysis: Maria's soup eating for 1 minute and 15 seconds; Maria's main course for 2 minutes and 23 seconds; Alice's soup for 1 minute and 11 seconds; and Alice's main course for 28 seconds.

4.1. Gesture characterisation

From the observation of the three participants, we identified several important characteristics in the hands and head movements that can be used to automatically recognise intake gestures.

To eat the soup, all three people held the spoon in their right hand. When the right hand moves from the soup dish to the mouth, it moves smoothly and the subjects often lean their heads forward. Pedro and Maria, for instance, lean forward when eating several spoonfuls, then take a break and repeat the procedure. Alice leans her head forward and backwards for each spoonful. All subjects mainly keep their left hand still, near the plate. Sometimes they used it to scratch their head, their chest and to wipe their mouth with a napkin.

During the main course, Maria and Alice ate with the fork in the right hand. Pedro ate with the fork in the left hand and the knife in the right hand. Generally, the hands and the head had the same behaviour when moving the hands in the direction of the mouth, however all participants' frequency of gestures during the main course was lower than for the soup. The gestures of cutting and arranging the food were sometimes possible to recognise, but they had many bad estimations.

Finally, to drink from the glass, all of them grabbed the glass with their right hand and moved it smoothly to the mouth, they leaned their head slightly backwards to drink, and then put the glass back down on the table. We noticed that they took longer to drink than to eat the soup or main course, which makes this one distinguishing fact for recognition of eating types. Usually the head leaned backwards when the subjects were drinking. Most of the time, the subjects kept their left hand still during drinking activities, but Maria once used both hands to grab the glass.

4.2. Gesture evaluations and Meal monitoring

In Table 2 we can see the values obtained from the analysis of the valid periods of the three experiments.

Table 2. Intake gestures

Intake gestures	Maria			Alice			Pedro		
	Nº	Success	%	Nº	Success	%	Nº	Success	%
Soup	Duration: 3 min. 3 sec.			Duration: 2 min. 21 sec.			Duration: 2 min. 56 sec.		
Right hand plate-mouth	29	29	100%	34	30	88%	31	24	77%
Other gestures	4	-	-	2	-	-	0	-	-
Main Course	Duration: 2 min. 29 sec.			Duration: 16 min. 18 sec.			Duration: 3 min. 33 sec.		
Right hand plate-mouth	15	12	80%	116	108	93%	0	0	-
Left hand plate-mouth	0	0	-	3	3	100%	27	20	74%
Right hand glass –mouth	4	4	100%	2	2	100%	3	3	100%
Other gestures	5	-	-	3	-	-	2	-	-

During the soup eating activity, most of the gestures were satisfactorily captured: for Maria, all 29 gestures were recognisable; for Alice, 80% were recognisable – the remaining gestures were not recognisable due to a total of 18 seconds of missed joint estimation; and for Pedro, 77% were recognisable and the remaining were not due to 37 seconds having the same problem as in Alice's case.

For the main course, gestures were also satisfactorily captured: for Maria, 80% of right hand joints were recognisable and the remaining were not due to 40 seconds missed joint estimation; for Alice, 93% were recognisable and the remaining were not due to 1 minute and 23 seconds missed joint estimation, and 100% of left hand gestures were recognisable; and for Pedro, 74% of left hand gestures were recognisable and the remaining were not due to 28 seconds missed joint estimation.

There were few drinking gestures (only ten) and they only occurred during the main course. All of them were recognisable.

We counted “other gestures” that subjects made during the experiments, which were six during the soup course and ten during the main course. Although they do not contribute for feeding monitoring, they were recognisable.

Looking at the sensor’s missed joint estimations, we observe that they were not significant and that they can be automatically identified by coherence analysis of seated skeleton, as sometimes the skeleton gets deformed and human gestures are always smooth.

After the analysis, we were confident that it is possible to monitor elderly people with the seated skeleton automatically detected by the MS Kinect sensor – a non-intrusive and cheap sensor.

5. Conclusions and future work

After analysing all of the three participants eating lunch in the given scenario, we concluded that the seated skeleton of the MS Kinect is capable of reasonably following the intake gestures of the people monitored. It had a success rate of at least 74% for an isolated distance (observed in Pedro’s left hand plate to mouth distance in the main course) and with an average of 89% success rate for all distances. Also, the sensor’s missed joint estimations were not significant and can be automatically identified by coherence analysis of the seated skeleton. We are therefore confident that these results are sufficient and it is possible to use this approach to develop a system to automatically detect intake gestures and do meal monitoring.

For the next step, we would like to reduce occlusion problems that might occur. We can do this, for instance, by using two Kinect sensors instead of only one and by creating a skeleton model, using coherence analysis, to validate skeleton positions. Accordingly, we will develop a system using a statistic classification model called Hidden Markov Model. The system should be capable of checking if the user is eating or not, and making an estimation of the amount of food and fluids ingested.

Acknowledgements

The authors thank the FEUP – Faculdade de Engenharia da Universidade do Porto through the Project I-City for Future Mobility: NORTE-07-0124-FEDER-000064, and European Project FP7 - Future Cities: FP7-REGPOT-2012-2013-1.

References

[1] Campos MTF de S, Monteiro JBR, Ornelas APR de C. Fatores que afetam o consumo alimentar e a nutrição do idoso. Rev Nutr. 2000 Dec;13[3]:157–65.
[2] Amft O. Automatic Dietary Monitoring Using on-Body Sensors: Detection of Eating and Drinking Behaviour in Healthy Individuals. ETH Zurich; 2008. 244 p.
[3] Amft O, Troster G. Methods for Detection and Classification of Normal Swallowing from Muscle Activation and Sound. Pervasive Health Conference and Workshops, 2006. 2006. p. 1–10.
[4] Rowland DT. Population Aging: The Transformation of Societies. Springer; 2012. 300 p.
[5] Gaßner KS, Conrad M. ICT Enabled Independent Living for the Elderly: A Status-quo Analysis on Products and the Research Landscape in the Field of Ambient Assisted Living (AAL) in EU-27. IIT; 2010. 91 p.
[6] Amft O, Kusserow M, Troster G. Automatic Identification of Temporal Sequences in Chewing Sounds. IEEE International Conference on Bioinformatics and Biomedicine, 2007 BIBM 2007. 2007. p. 194–201.
[7] Han J, Shao L, Xu D, Shotton J. Enhanced Computer Vision With Microsoft Kinect Sensor: A Review. IEEE Trans Cybern. 2013 Oct;43[5]:1318–34.
[8] Sevick MA, Piraino B, Sereika S, Starrett T, Bender C, Bernardini J, et al. A Preliminary Study of PDA-Based Dietary Self-Monitoring in Hemodialysis Patients. J Ren Nutr. 2005 Jul;15[3]:304–11.

[9] Yon BA, Johnson RK, Harvey-Berino J, Gold BC. The Use of a Personal Digital Assistant for Dietary Self-Monitoring Does Not Improve the Validity of Self-Reports of Energy Intake. J Am Diet Assoc. 2006 Aug;106[8]:1256–9.

[10] Yon BA, Johnson RK, Harvey-Berino J, Gold BC, Howard AB. Personal Digital Assistants are Comparable to Traditional Diaries for Dietary Self-Monitoring During a Weight Loss Program. J Behav Med. 2007 Apr 1;30[2]:165–75.

[11] Mankoff J, Hsieh G, Hung HC, Lee S, Nitao E. Using Low-Cost Sensing to Support Nutritional Awareness. In: Borriello G, Holmquist LE, editors. UbiComp 2002: Ubiquitous Computing. Springer Berlin Heidelberg; 2002. p. 371–8.

[12] Siek KA, Connelly KH, Rogers Y, Rohwer P, Lambert D, Welch JL. When Do We Eat? An Evaluation of Food Items Input into an Electronic Food Monitoring Application. Pervasive Health Conference and Workshops, 2006. 2006. p. 1–10.

[13] Patterson DJ, Fox D, Kautz H, Philipose M. Fine-grained activity recognition by aggregating abstract object usage. Ninth IEEE International Symposium on Wearable Computers, 2005 Proceedings. 2005. p. 44–51.

[14] Chang K, Liu S, Chu H, Hsu JY, Chen C, Lin T, et al. The Diet-Aware Dining Table: Observing Dietary Behaviors over a Tabletop Surface. In: Fishkin KP, Schiele B, Nixon P, Quigley A, editors. Pervasive Computing. Springer Berlin Heidelberg; 2006. p. 366–82.

[15] Gao J, Hauptmann AG, Bharucha A, Wactlar HD. Dining activity analysis using a hidden Markov model. Proceedings of the 17th International Conference on Pattern Recognition, 2004 ICPR 2004. 2004. p. 915–918 Vol.2.

[16] Hossein Mousavi Hondori M khademi. Monitoring Intake Gestures using Sensor Fusion (Microsoft Kinect and Inertial Sensors) for Smart Home Tele-Rehab Setting. 2012.

[17] Amft O, Troster G. On-Body Sensing Solutions for Automatic Dietary Monitoring. IEEE Pervasive Comput. 2009 Apr;8[2]:62–70.

[18] Troost PS. Vision and on-body sensor fusion for monitoring food intake behaviour in a restaurant environment. Technische Universiteit Eindhoven; 2013.