International Journal of Healthcare Information Systems and Informatics
Volume 16 • Issue 4

Food Intake Vision-Based Recognition System via Histogram of Oriented Gradients and Support Vector Machine for Persons With Alzheimer's Disease

Haitham Asaad Al-Anssari, Western Michigan University, USA*
Ikhlas Abdel-Qader, Western Michigan University, USA
Maureen Mickus, Western Michigan University, USA

ABSTRACT

Oftentimes, individuals with Alzheimer's suffer from malnutrition. Sadly, as these individuals experience cognitive decline, they simply forget to eat and must be assisted to consume even if food is presented. Here, a vision-based system to monitor eating patterns is discussed. The Viola-Jones method, a histogram of oriented gradients being generated for feature extraction, is applied to detect the upper body region (UBR) while a support vector machine is used to distinguish eating vs. non-eating. Within the identified UBR, Haar-like features are used to detect hands while moving between food being served and the mouth to reduce false positive results. In this work, a combined template image (CTI) method is proposed to eliminate false positive hand detections; specifically, 30 hand eating posture images have been selected and combined into a single template image. Results show that implementing a CTI to match subjects to images is 2.86 times faster than matching each subject separately to the 30 images. Moreover, an experimental simulation using 33 videos of 163,840 frames indicates that the proposed method achieves a high accuracy of 90.65%.

KEYWORDS

Combined Template Image (CTI), Histogram of Oriented Gradients (HOG), Support Vector Machine (SVM), Template Matching, Upper Body Region (UBR)

1. INTRODUCTION

Dementia is a neurodegenerative disease that entails significant cognitive decline, thereby seriously affecting activities in one's daily life. The Alzheimer's disease, ranked as the sixth cause of death in the US (Alzheimer's Disease Facts & Figures, 2017), is the most common form of dementia. Globally, about 44.4 million people are diagnosed with the disease. By 2050, this number is expected to rise to 131.5 million. In 2016, the Alzheimer's Disease International (ADI) reports that dementia cases worldwide is to rise by about 9.9 million annually (Alzheimer's Disease: Facts & Figures, 2018). During the period 2017 to 2050, it is purported that the number of inflicted US individuals will grow from 5 to 16 million (Alzheimer's Statistics, 2018), thereby leading to staggering public health care costs. The US health care expenditure for individuals inflicted with the Alzheimer's (and/or dementia)

DOI: 10.4018/IJHISI.295817 *Corresponding Author

in 2018 reached $277 billion, and has been touted to inflate to $1.1 trillion by 2050. Clearly, as this disease progresses, substantial caregiving will be needed by both family members and formal care providers (Alzheimer's Disease: Facts & Figures, 2018).

A major challenge for individuals needing dementia care is nutritional intake. In contrast to 26.7% of non-impaired older adults, about 81.4% of Alzheimer's patients are reportedly malnourished (Kai, 2015). This is because individuals with cognitive impairments often forget to eat and/or may not even recognize how to eat. A family member or a caregiver assistant is often required to help these individuals, for example, even with the use of simple utensils; hence, effective ways are needed to support the nutritional intake of these individuals, which is essential in maintaining their health and wellbeing (Rasheed & Woods, 2013).

A 2011 qualitative study (Johansson, Christensson & Sidenvall, 2011) illustrated that persons with Alzheimer's, particularly in the early stages, prefer to be independent in their eating as long as possible. Caregivers who provide unnecessary assistance to older individuals with the disease may create a premature dependency. Interventions are needed for both the person with dementia and care providers to extend independence for eating as long as is advisable to do so. Assistive technologies may be useful to encourage independent eating and reduce the burden on the part of caregivers (Marasinghe, 2016).

We organize this paper as follows. Following the introduction, Section 2 summarizes the background on food intake monitoring strategies for Alzheimer's patients. Section 3 then highlights a novel approach, encompassing three (3) implementation stages, to advance a food intake vision-based monitoring system via a computer tablet to track independent eating with Alzheimer's patients. Importantly, the role of verbal cueing to encourage eating within the targeted population is emphasized (e.g. Chard, Liu & Mulholland, 2008). The developmental framework lays the foundation for future research when incorporating audio and/or video prompts as needed. Next, in Section 4, the experimental work designed to test the proposed eating tracking framework and results are presented. Finally, Section 5 concludes our discussion with insights into research limitations, practical implications, and potential future work.

2. BACKGROUND

Broadly, past approaches for food intake monitoring have involved sensors attached to the subject's body such as the work presented in (Mendi, Ozyavuz, Pekesen & Bayrak, 2013; Dong & Biswas, 2016; Kalantarian, Alshurafa, Le & Sarrafzadeh, 2015; Päßler & Fischer, 2011; Shuzo, 2009; Rosa, Anastasova-Ivanova, Lo & Yang, 2019). Such systems are not suited for individuals with Alzheimer's as they may simply remove these sensors (or refuse to wear them in the first place) given that they may perceive these items as obtrusive.

Using a combination of YCbCr and YIQ color spaces to detect skin regions and creating bounding boxes around these regions, researchers have proposed a food intake monitoring system for Alzheimer's patients (Al-Anssari & Qader, 2016). Tracking implemented via such a technique of controlled bounding boxes has reported having an accuracy of 90.82%. Similarly, the work of Al-Anssari, Qader & Mickus (2019) to monitor eating for those with Alzheimer's has an even higher accuracy of 94.29%. Here, the researchers used two skin detection methods for the face and hands detection. Once the face and hands are located, a tracking algorithm is triggered to track the movement of the detected objects with a decision made to ascertain whether the subject is eating or not.

Both of the two aforementioned methods suited for Alzheimer's patients have been based on the skin color as the main attribute to detect the face and the hands of their study subjects. In some cases, when the background color is similar to the skin color, and/or owing to poor lighting conditions, these algorithms may not perform as expected. Accordingly, a novel method is proposed in this paper, emphasizing other features for the face and hands rather than relying on detecting skin color (Al-Anssari, et al., 2019). Gao, Hauptmann, Bharucha & Wactlar (2004) also proposed a food

intake monitoring system using hidden Markov model (HMM) to track certain features such as the physical distance between moving objects; however, their method only reported 77% accuracy. Qiu, Lo & Lo (2019) used Mask R-CNN or convolutional neural network for person and food detection. These researchers used real-time hand tracking via SSD neural network on "tensorflow" (an object detection API) to detect hands and associated movements. Following the detection, bounding boxes are created for the detected objects; then, for each frame, a hand-face distance is measured by finding the distance between the centers of their bounding boxes. The system has been pilot-tested via a 360 camera with two healthy subjects. To ensure the robustness of the system, more tests need to be conducted with more subjects under different lighting conditions.

Other research has focused on detecting the size of the food and the number of bites with a platform for calculating the number of calories in the intake food to reduce obesity (Villalobos, Almaghrabi, Hariri & Shirmohammadi, 2011). To compute caloric intake, the size of food portions is often determined by the calibration between pictures already taken by the users for their food before eating via a smart phone and their thumbs as a reference. Similarly, Sun, Liu, Schmidt, Yang, Yao, Fernstrom J., Fernstorm M., DeLany & Sclabassi (2008) proposed a method for food intake monitoring where a card is used to determine food quantity vis-à-vis the portion size for food size calibration. Cunha, Pádua, Costa & Trigueiros (2014) used the Microsoft Kinect (MSK) sensor for food intake monitoring for older persons. A skeleton for the participant is detected and imported via the MSKinect based on the participant's detected head and hands. To avoid the hands occlusion while eating, the MSKinect sensor is placed 1.2 meter higher from the floor with a tilt of -10 degrees and at an approximate distance of 1.2 meters from the user. Here, the food intake detection success rate was found to be at least 74% for an isolated distance but with an average of 89% success rate for all distances.

Today, in many object detection and recognition methods, the Histogram of Oriented Gradients (HOG) has been used. Chowdhury, Kowsar & Deb (2017) proposed a framework for human detection via adaptive background mixture and improved HOG. Upon detecting the region of interest (ROI), improved HOG is used for features extraction and Support Vector Machine (SVM) trained and implemented for human/non-human detection by these researchers. The framework achieves a precision of 93.7%. Wei, Tian & Guo (2013) combined Haar-like features, AdaBoost algorithm, HOG, and SVM for pedestrian detection. Haar is used for pedestrian detection while the HOG and SVM are used to differentiate a pedestrian from a non-pedestrian. Déniz, Bueno, Salido & Torre (2014) presented a method for face recognition via HOG. To eliminate the errors resulting in facial features extraction based on occlusion and illumination variation, they proposed to extract the HOG features via a regular grid. These researchers also found that results were improved when using HOG descriptors extracted from image patches with different sizes over selecting the best single image patch. Lin & Ding (2013) proposed a method for hand gesture recognition using HOG and motion trajectory. HOG and SVM are used for hand localization over video frames whereas the Mahalanobis distance between input gesture and generated database is implemented for gesture recognition.

Additionally, Haar-like features method has been widely used in object detection. Hand gesture recognition is one of the applications that could use of Haar features. Huang, Chao & Kao (2012) proposed a system for tracking, recognition, and distance detection of hand gesture for a 3-D interactive display. These researchers implemented hand detection via Haar-like features, then performed hand tracking via mean-shift and Kalman filter. Depth information is collected using Hough-transform algorithm. Similarly, Hsieh, Liou & Lee (2010) presented a real time hand gesture recognition via motion history images. They used Haar-like features for dynamic hand gesture classification. Their system has been trained to detect four hand movement recognition, up, down, right and left in addition to the hand fist and waving hand gestures.

Finally, Chen, Georganas & Petriu (2008) proposed a method for hand gesture recognition based on Haar-like features and stochastic context-free grammar. AdaBoost learning method has been used to speed the classification process and to build a strong classifier, whereas stochastic context-free

grammar (SCFG) is used for the hand gesture recognition. Bilal, Akmeliawati, Salami, Shafie & Bouhabba (2010) combined Haar-like features and skin-color detection in hand posture recognition. The hand is first detected over several frames, then the skin color distribution would be extracted from the hand and then being deployed for skin regions detection for the next frames. Both the hand and face blobs are being tracked via Kalman filter. Tribaldos, Serrano-Cuerda, López, Fernández-Caballero & López-Sastre (2013) proposed a method for human detection in color and infrared videos via HOG and SVM. Takahashi, Fujii, Shibata & Satoh (2010) presented a system for human behaviors recognition such as running, meeting, or object placed in crowded surveillance videos via HOG and SVM. Tracking of human region was done using Kalman filter. Motion cues are also involved to ensure robust behavior recognition. Additionally, background tracking is used to detect fast motion while optical flow is used for small motion detection.

3. PROPOSED FRAMEWORK

In the current work, we propose a novel framework with the goal of developing an automated alert system to detect food intake activity for people with Alzheimer's disease. The proposed method has three (3) stages of implementation, namely, upper body region (UBR) detection, hand detection, and eating recognition as detailed below. The system will use HOG for features detection and SVM to assess eating behavior.

Altogether, the proposed system focuses on the subject while eating. For this purpose, the UBR detection is first applied, followed by eating recognition within this region. To reduce the number of false positive eating detections, the upper body region will be checked for eating detection only when the hand is presented inside this region. Therefore, three methods for hand detection have been tested and the results are compared, HOG, Harr, and Local Binary Patterns (LBP). Our experiments have shown that Haar-like features detection method outperformed the HOG and LBP methods. Moreover, we further enhance the hand detection method with a template matching where each of the hand or hand-like detected region is matched with 30 different eating hand postures. **Figure 1** shows the schematic for our proposed framework.

When the UBR is detected and the hand is presented within this region, we then apply HOG to detect the features of eating gesture following with using the SVM to decide whether this is or is not an eating gesture.

3.1 Upper Body Detection

The first step in the proposed method is implementing the UBR detection via the Viola-Jones algorithm (Viola & Jones, 2001). The object detection procedure classifies images based on the value of simple features where integral images have been used. AdaBoost is then implemented to select a small set of features and to train the classifier. Finally, a cascade classifier is used to classify the detected object. Once the UBR is detected, hand detection and the monitoring of eating will be implemented on the UBR.

3.2 Hand Detection

Next is hand detection, which is applied to detect if the hand is presented inside the UBR. Three methods for hand features extraction are being tested here, namely, HOG, Haar, and LBP. These methods have been used successfully in object detection with each method having its own feature extraction procedure as noted below.

3.2.1. Histogram of Oriented Gradient or HOG

The HOG feature has been introduced for pedestrian detection by Dalal & Triggs (2005). This technique counts the occurrences of gradients used to describe the local object shape (Waghmare, 2012). The

Figure 1. Proposed Food intake monitoring system flowchart

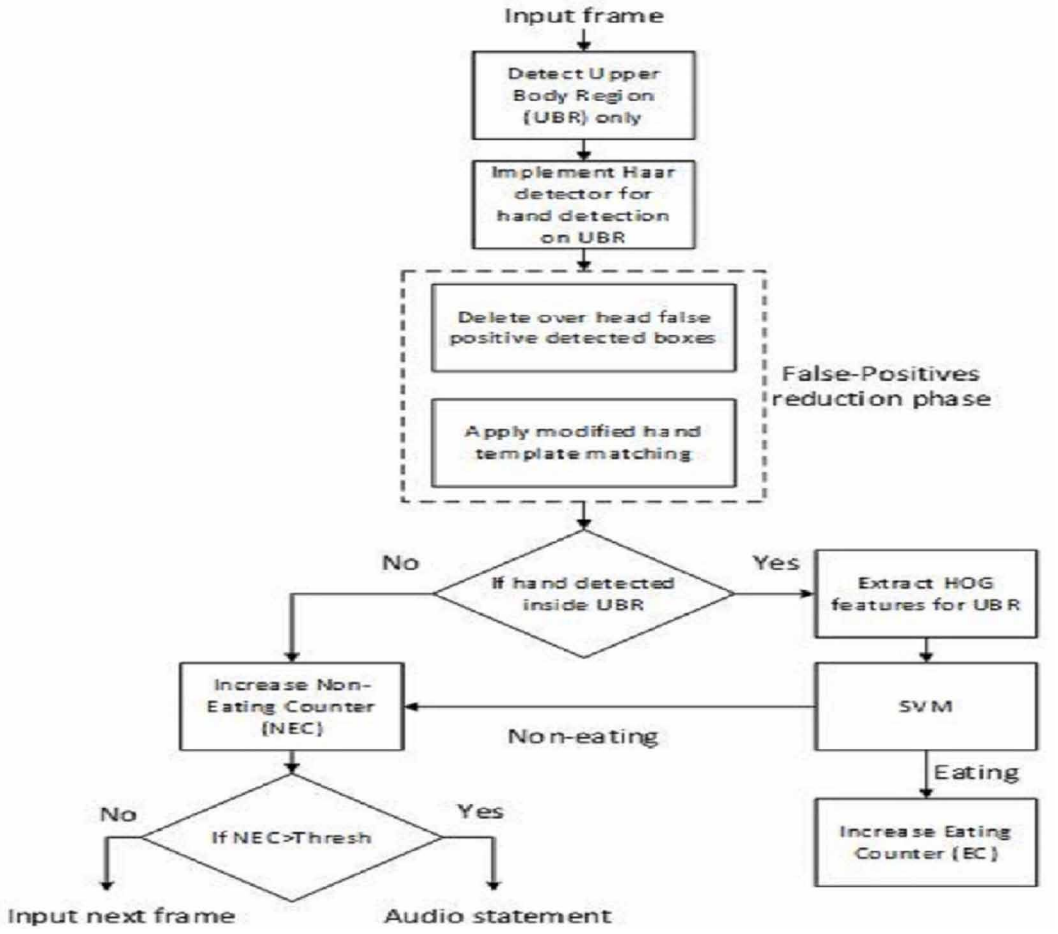


image is divided into blocks; moreover, each block is further subdivided into a set of small connected regions called cells (Barngrover, 2014). The orientation for the maximum gradient and the magnitude are computed at each location with a simple one-dimensional centered mask [-1,0,1]. Across each cell, and for each orientation bin, the magnitudes are summed. Orientations are either signed or unsigned; if signed, the gradient will be arrow-like feature and the direction of the gradient across the full 360þ whereas, with the unsigned orientation, the direction will be across the 180 þ range (Newell, 2011).

3.2.2. Haar-Like Features

Computationally fast for detection procedures, the Haar Classifier is a supervised learning technique. It was firstly implemented successfully in face detection, but it has also proven to perform highly in object detection and is touted to be invariant to changes in illumination, color and scale (Johnson, 2012). The sliding window is the technique used to detect objects in an image. Here, the image is scanned with the sliding window and thousands of sub-windows being generated. A classifier is then used to decide whether each of these sub-windows is either positive or negative (Saberian, 2012).

Haar-like features are digital images comprising two to three white and black rectangular regions as shown in **Figure 2.** The white rectangular will represent the bright area in the image while the black rectangular represents the dark area. Haar-like feature is found by subtracting the summation of

the pixels under the dark region from the summation of pixels under the white rectangular. Integral image is implemented with the Haar features extraction to speed up the process. Computing of the integral image is done in a recursive way as shown in Eq. (2). Each Haar-like descriptor is a weak classifier with high false-positive detections (Zhang, 2015). Weak classifiers are used and combined to result in a powerful discriminative classifier (Johnson, 2012).

Figure 2. The extended set of Haar-like features

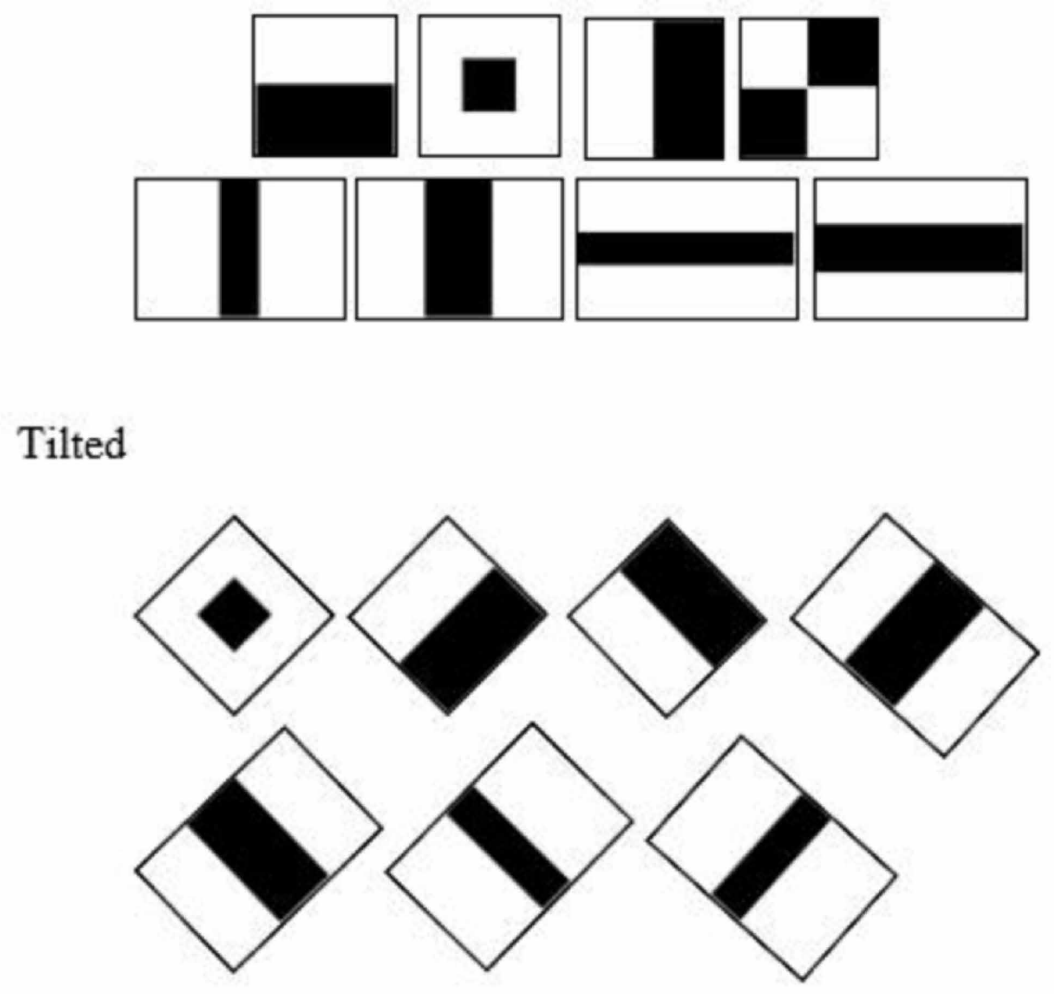
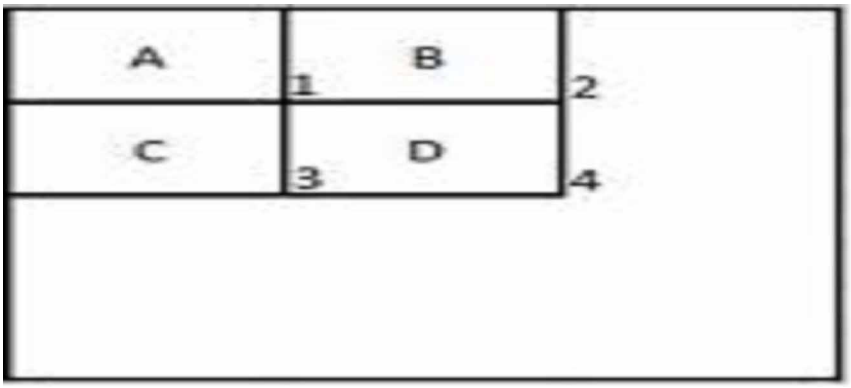


Figure 3 is an example of the implementation of the integral image. The value of the integral image at point labeled as 1 is the summation of pixels' intensities at region A. The summation of pixels in A + B regions is at point 2, whereas at point 3, it is A + C regions. Finally, at point 4 is A + B + C + D regions. Also, the sum of rectangular D is found by 4 + 1 – 2 – 3 (Johnson, 2012).

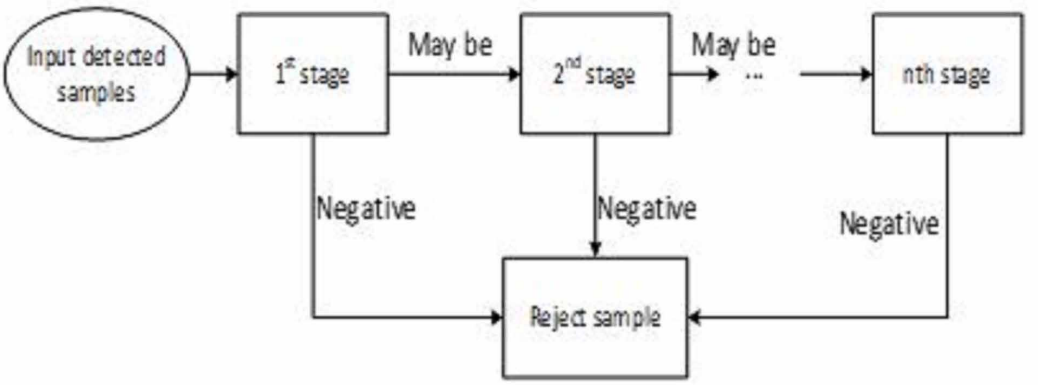
The cascade of classifiers increases positive detection rate with less computation time. Smaller boosted classifiers are used to reject many negative samples and detect the positives. At each classification stage, the classifier is trained to detect positive samples and reject the negatives. The

Figure 3. Integral image



positive detection from the first classifier triggers the second stage of classification. The third stage of classification is triggered by the positive detection from the 2nd stage, and so on as shown in **Figure 4**. At any stage, if any negative outcome results from the classification, it will be rejected directly. The task for the succeeding classifier is harder than the proceeding one, whereby computation time will be longer or larger (Johnson, 2012).

Figure 4. Cascade of classifiers with n stages



3.2.3. Local Binary Patterns or LBP

The LBP feature descriptor is a binary descriptor used to describe the texture of an image. The LBP feature detection method thresholds the neighborhood pixels' intensities located around the center focus pixel to produce a binary representation. It converts each pixel into a binary representation by thresholding the neighborhood pixels around the pixel under focus. The image is divided into blocks; for each of these blocks, the center pixel intensity is considered as a threshold value to threshold neighborhood pixels which will be assigned either 0 or 1. Then, these values are multiplied by powers of two and added together to generate the label for the center pixel. Accordingly, for 3x3 block size, there will be 8 neighbors; then, 2 to the power of 8 will be 256 labels.

Figure 5 shows the resulting LBP implementation, in which label is 1+4+16+64=85 (Sharma, 2014; Pietikäinen, Hadid, Zha & Ahonen, 2011). An example of LBP feature extraction is given in **Figure 6.**

Figure 5. Example of LBP calculation

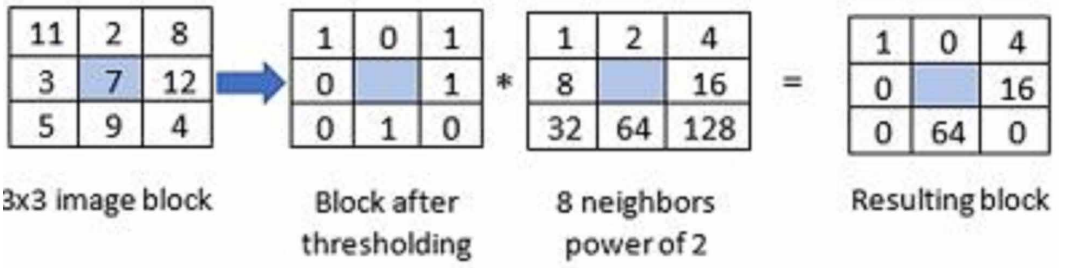
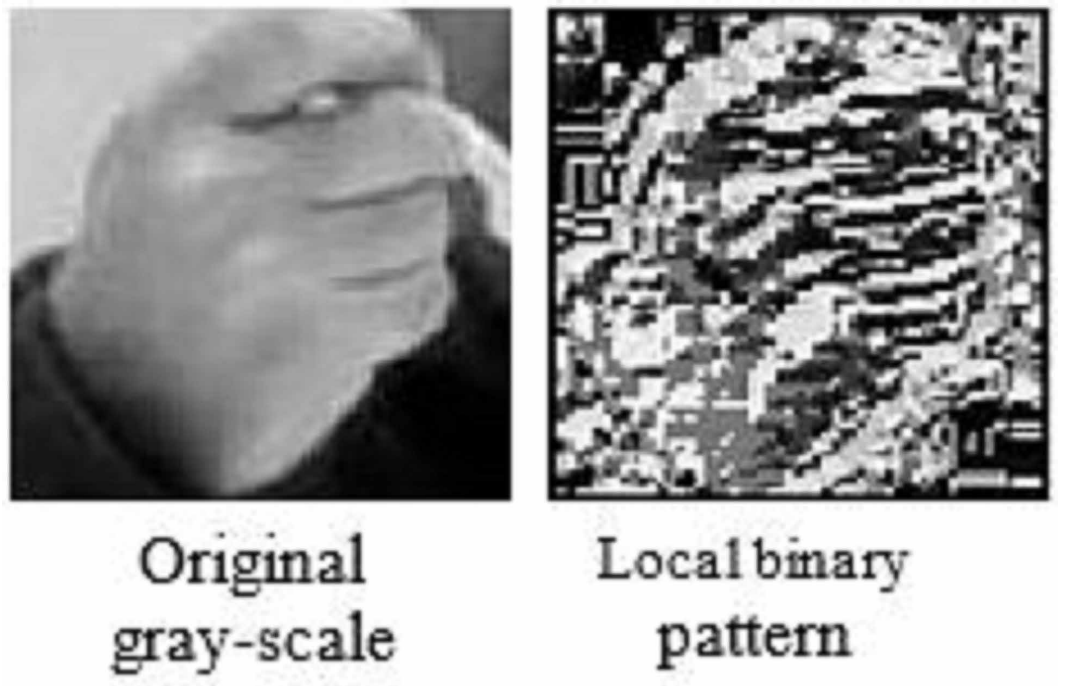


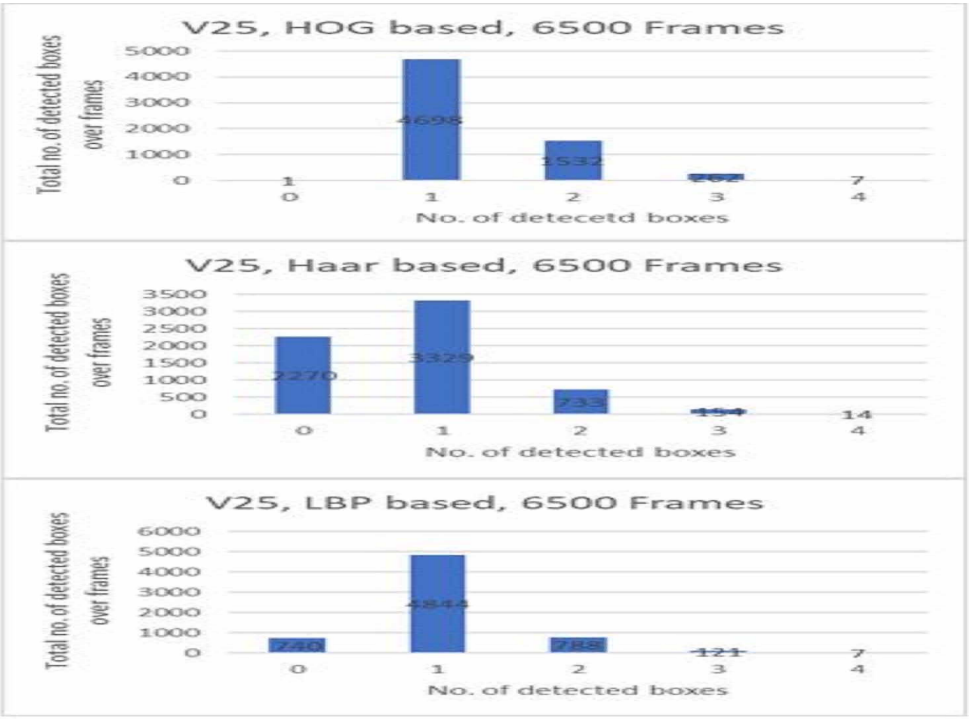
Figure 6. Example of LBP features extraction



3.3 Combined Images Templates or CIT

The features-based hand detection methods are tested via several videos for the study subjects while eating, whereby results affirm that hand detection based on Haar-like features outperform the other two methods. **Figure *7*** shows a comparison of hand detection via features-based methods. There are false positive detections result with hand detection process and should be eliminated.

Figure 7. Hand detection based on a. HOG features, b. Haar features, and c. LBP features implemented on video number 25



In this framework two filtering methods are used to reduce the false positive detections. First, any false detection located above the head region inside UBR is discarded. Second, each of the detected region is matched to a template of 30 selected hand eating postures combined into one template image as shown in Figure 8.

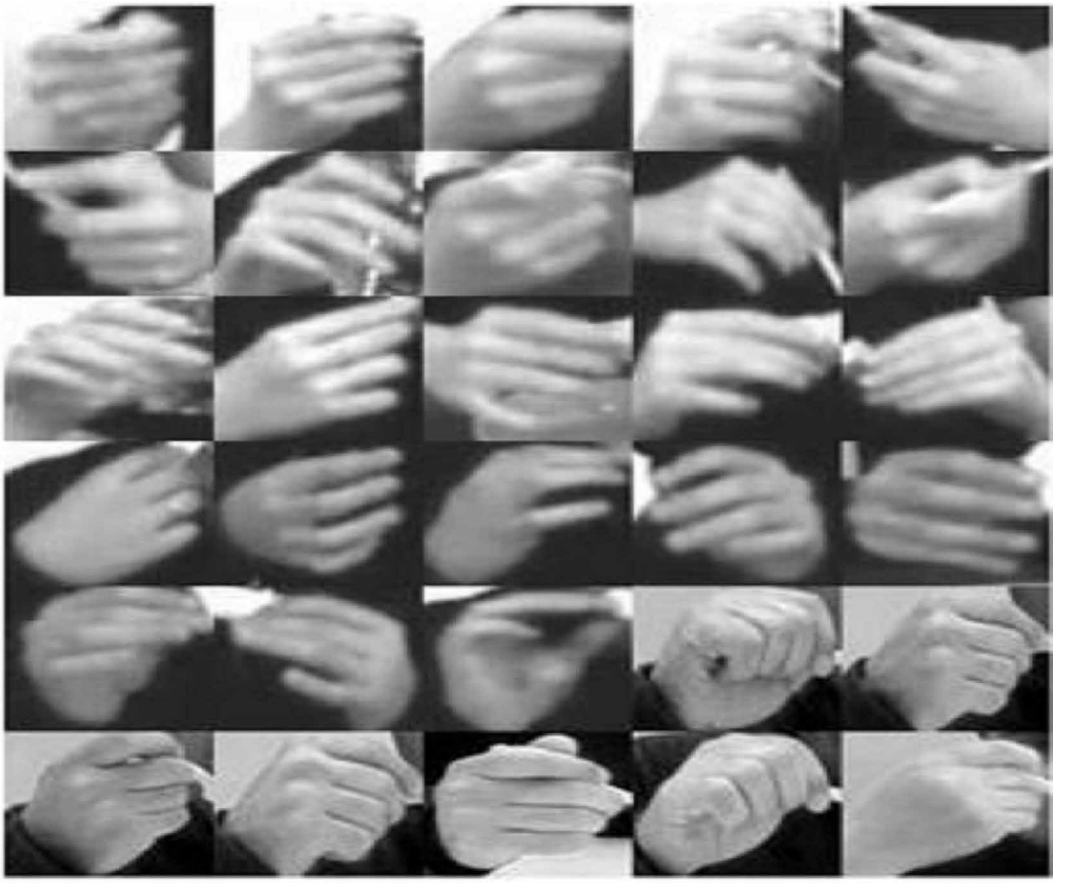
As shown, these images have been captured from videos for people while eating. Some of these videos have been recorded in the lab, at home, and/or in other locations (Dataset1) vis-à-vis the MOBISERV-AIIA dataset (Iosifidis, Marami, Tefas, Pitas & Lyroudia, 2015). Matching is computed via the normalized cross-correlation between each detected region and the combined template image (CTI). Compared to matching between each detected region and doing the 30 selected hand eating posture matching separately, the CTI method is much faster by a factor of 2.86 in the computation process. The CTI-based method execution time was 71.688 ms while the execution time for 30 template images matching was 205.571 ms; clearly, in order to achieve realtime implementation, the CTI is chosen.

After finding the maximum correlation value, the location of this maximum value is also considered. If this location is between two neighbor hand posture images, then it is discarded. This maximum correlation value is also compared to a predefined threshold. **Figure 9** shows example of using Haar like features, and then the two stages of filtering.

The hand detection is performed on the UBR to detect hand regions candidates (HRC). Assuming that there are n number of detected HRCs, the first stage of filtering is applied to discard any HRC located above the subject's head, with m boxes being selected. The second stage of filtering is applied where the CTI matching method is applied. The normalized cross correlation (NCC) is performed to find the matching between the two images as follows.

1st filtering stage:

Figure 8. Combined images template used in the proposed method



$$HRC_m = HRC_n \quad if \quad loc(HRC_n > refPoint) \tag{3}$$

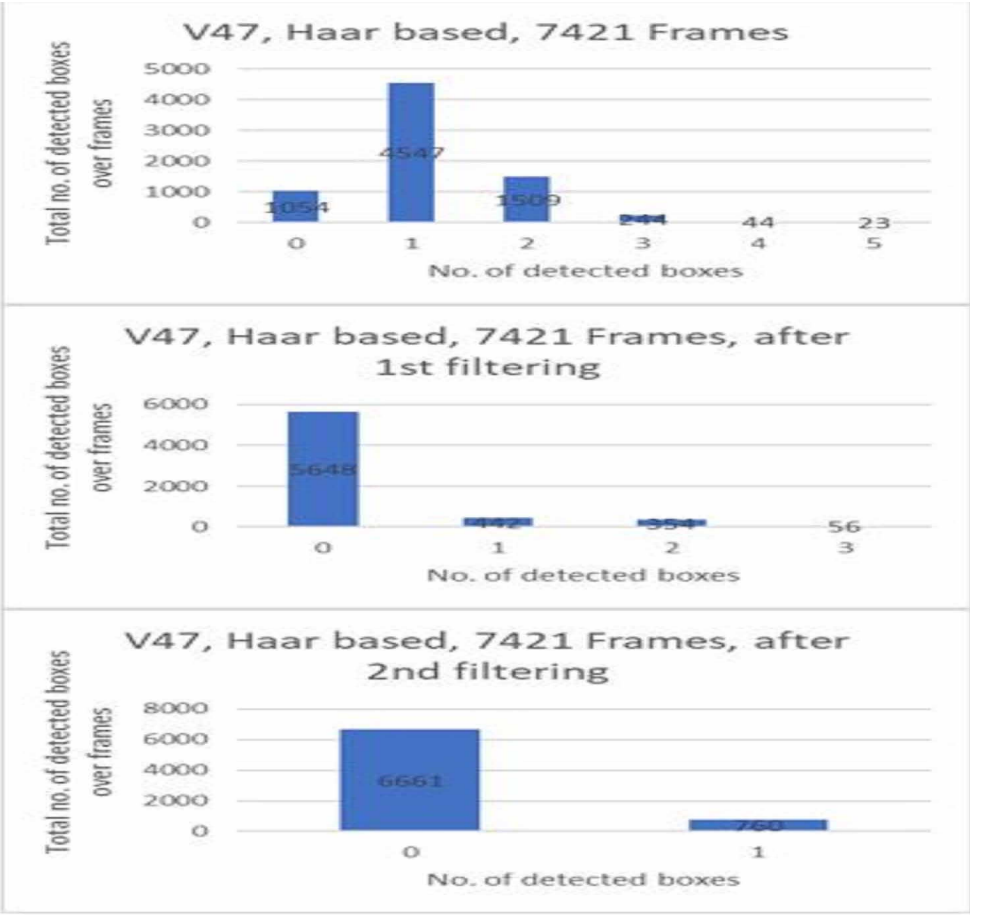
Where the *loc* is the location, and
refPoint = (x coordinate of the top line of the UBR + UBR box height)/2 …(4)
2nd filtering stage:
The normalized cross correlation between an HRC and the CTI is found by

$$c(u,v) = \frac{\left(\sum_{x,y}\left|HRC(x,y) - \overline{HRC}_{u,v}\right|\left|CTI(x-u,y-v) - \overline{CTI}\right|\right)}{\left\{\sum_{x,y}\left|HRC(x,y) - \overline{HRC}(u,v)\right|^2 \sum_{x,y}\left|CTI(x-u,y-v) - \overline{CTI}\right|^2\right\}^{\frac{1}{2}}} \ldots \tag{5}$$

Where $\overline{CTI}$ is the mean of the template, and $\overline{\mathrm{HRC}}_{\mathrm{u,v}}$ is the mean of the $HRC(x,y)$

As shown, the maximum cross correlation value and its location on CTI are found. The CTI size is 200x300 pixels, 5 horizontal images and 6 vertical; also, each of the HRCs is resized to 40x50 pixels. The matching method implemented in this framework will return two values, the maximum

Figure 9. Hand detection using Haar like features, and then after two filtering stages implemented on video 47



cross-correlation, and the location where the maximum occurs. The maximum correlation value MaxCor will be tested with predefined threshold corTh

$$MaxCor = \begin{cases} c\left(u,v\right), & c\left(u,v\right) \geq corTh \\ 0 & , \; x < corTh \end{cases} \tag{6}$$

To eliminate the potential error to occur when the location of the computed MaxCor is in the region between two hand images in the CTI, a location check is proposed where the location of MaxCor is compared to a set of predefined points, representing the locations of the images in the CTI via Euclidean distance *d* (Bishop, 2006)

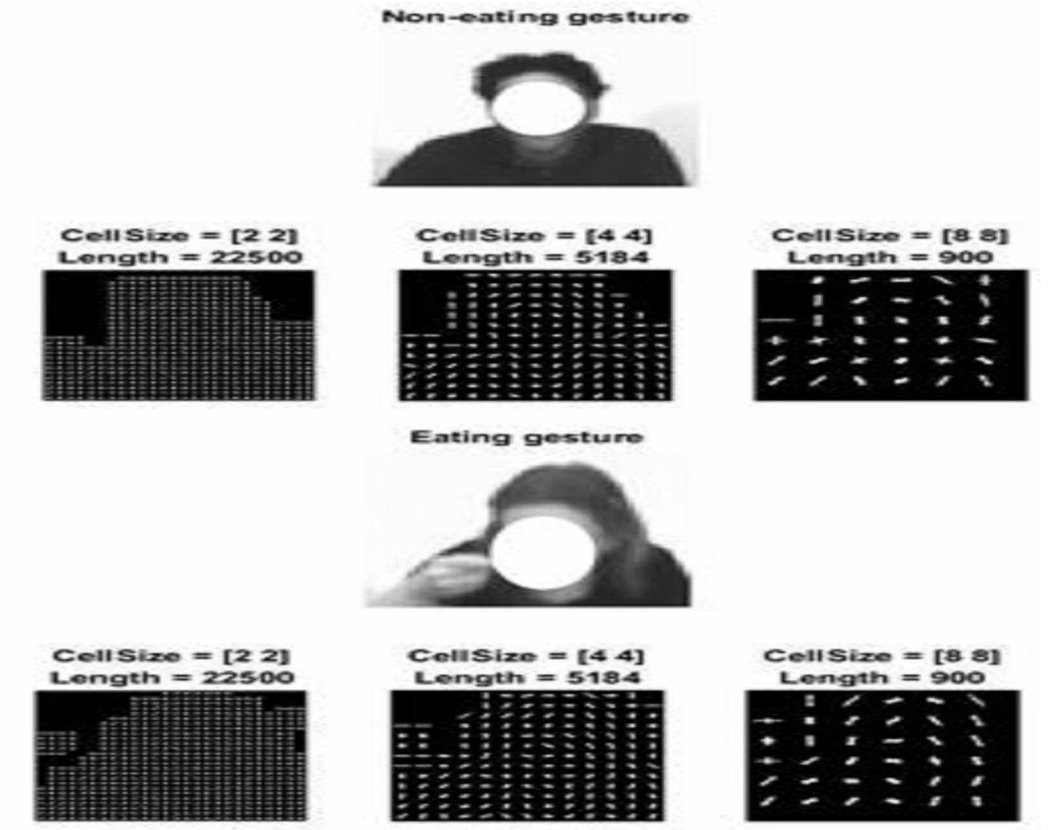
$$d\left(m\right) = \sqrt{\left(u-x\right)^2 + \left(v-y\right)^2} \tag{7}$$

The distances are then tested with a threshold dTh; if a distance is found less than dTh, the $MaxCor$ is considered. Otherwise, it will be discard.

3.4 Eating Detection

The detection of eating is performed via HOG for eating gesture features extraction, following which a trained SVM is used to decide whether this gesture is or is not eating. The HOG is used for the features extraction for the eating and non-eating gestures. **Figure 10** shows an example of the HOG features extraction for eating and non-eating gestures via 2x2, 4x4, and 8x8 cell sizes. When the hand is detected inside the UBR, the HOG is performed for the UBR based on 4x4 cells size, and then the features vector will be fed to the SVM trained classifier to decide if it is a non-eating or eating gesture.

Figure 10. HOG features extraction for eating and non-eating gestures and with 2x2, 4x4, and 8x8 cell sizes



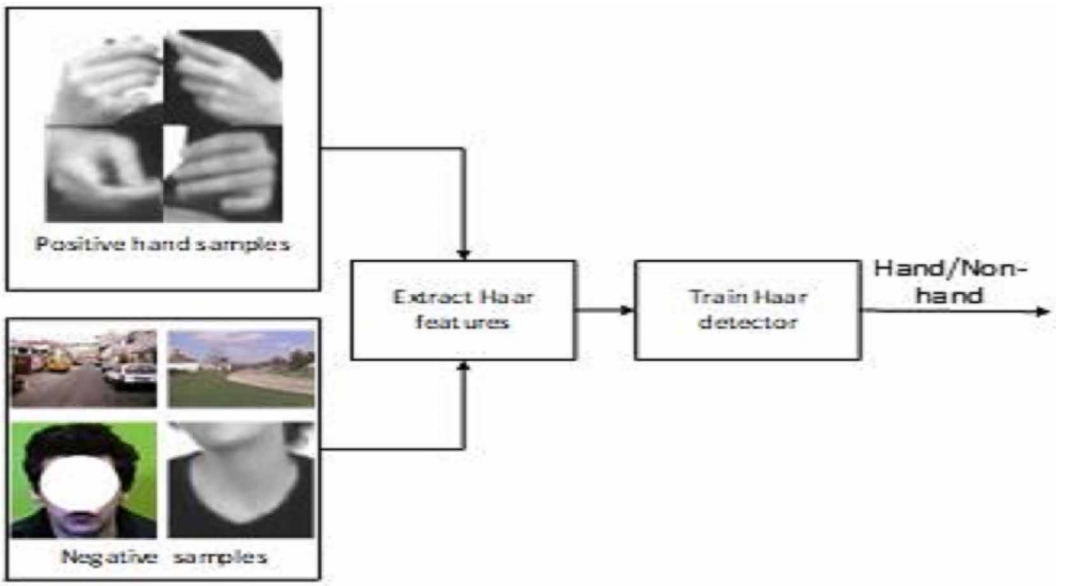
4. RESULTS AND DISCUSSION

The experimental work performed to test the proposed framework for eating monitoring is presented in this section. All experiments have been performed via Intel Core i5 2.30 GHz CPU with 6 GB RAM using the MATLAB environment. The input videos have been captured with a static camera equipped with a resolution of 320 × 240 pixels.

Three (3) hand detection features based methods have been evaluated and compared. Each classifier has been trained with 2077 positive samples for the hand while the study subjects are eating and 1393 negative samples for images that do not include any hand from INRIA (Dalal & Triggs,

2005) person dataset. **Figure 11** is an example of the Haar-like features extraction vis-à-vis the Haar detector training. The values of the corTh and the dTh are chosen to be 0.7 and 15 respectively.

Figure 11. Haar detector training



Monitoring of eating is performed using HOG for features detection and SVM for monitoring of eating. The SVM has been trained with 1129 positive samples for eating gestures and 892 negative samples for non-eating gestures with an example shown in **Figure 12.**

The positives samples for the hand detection, the positive samples for eating gestures, and the negative samples of eating gestures are taken from our recorded dataset (Dataset1) and the MOBISERV-AIIA dataset (Iosifidis, et al., 2015).

The proposed framework has been implemented on 33 videos, yielding a total of 163840 frames. These videos are from two sets. First set of videos is Dataset1, in which videos are recorded for the study subjects while they have been eating under different conditions and locations. Additionally, the MOBISERV-AIIA dataset is used with video recordings of people eating, drinking, talking, and reading a book. Long sleeves v. short sleeves as well as eating using utensils, such as fork and spoon, and with bare hands or drinking using straws or without straws are also variation scenarios being considered.

Results show that our proposed method for the hand detection performs very well. **Table 1** summarizes the results. Also, the monitoring of eating has been good and the successful detection rate is high at 90.65%. Wrong detections occur when the hand is moving inside UBR and not detected, or it can monitor wrong eating or non-eating patterns.

Figure 13 shows an example of the eating behavior for a study subject. Each colored slot represents the time (number of frames) between two successful eating events. Caregivers monitoring the patient's appetite can use this type of information.

Finally, **Figure 14** shows an example of the implementation of the proposed method where eating and non-eating gestures are successfully detected.

Compared to other vision-based monitoring systems such as the work presented in (Al-Anssari & Qader, 2016) where skin like color objects, which are not parts of the face or the hand, can resulted in wrong eating detection, the advantage of our proposed method is that the features of the face and

Figure 12. Eating and non-eating HOG features extraction and SVM training

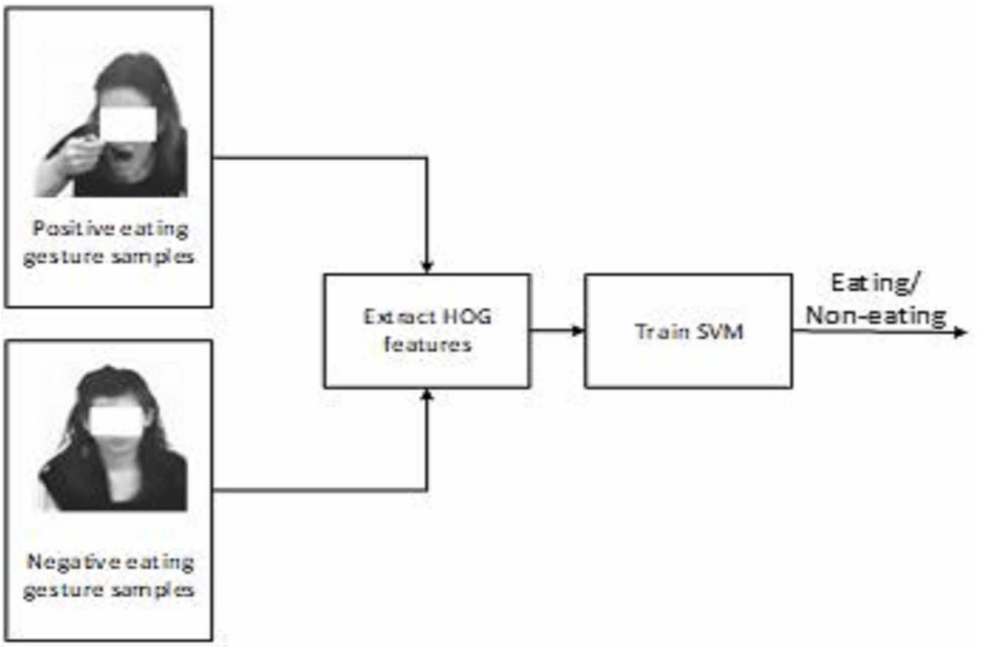


Figure 13. Show the eating behavior for a person while eating each colored slot represents the time between two successful eating events

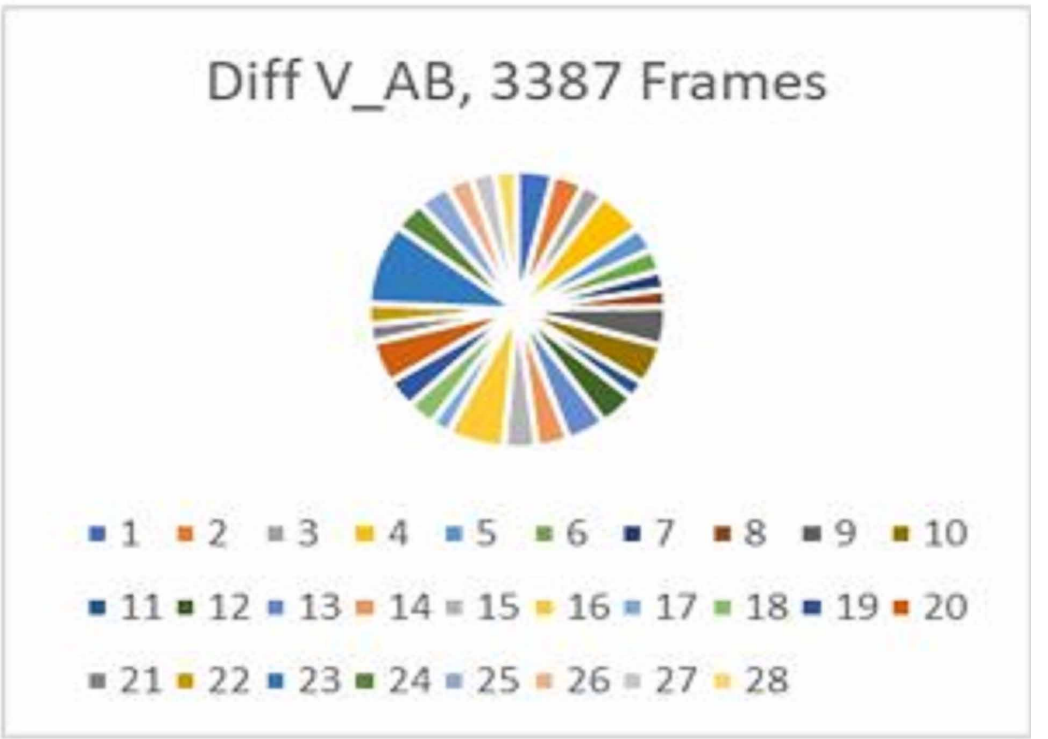
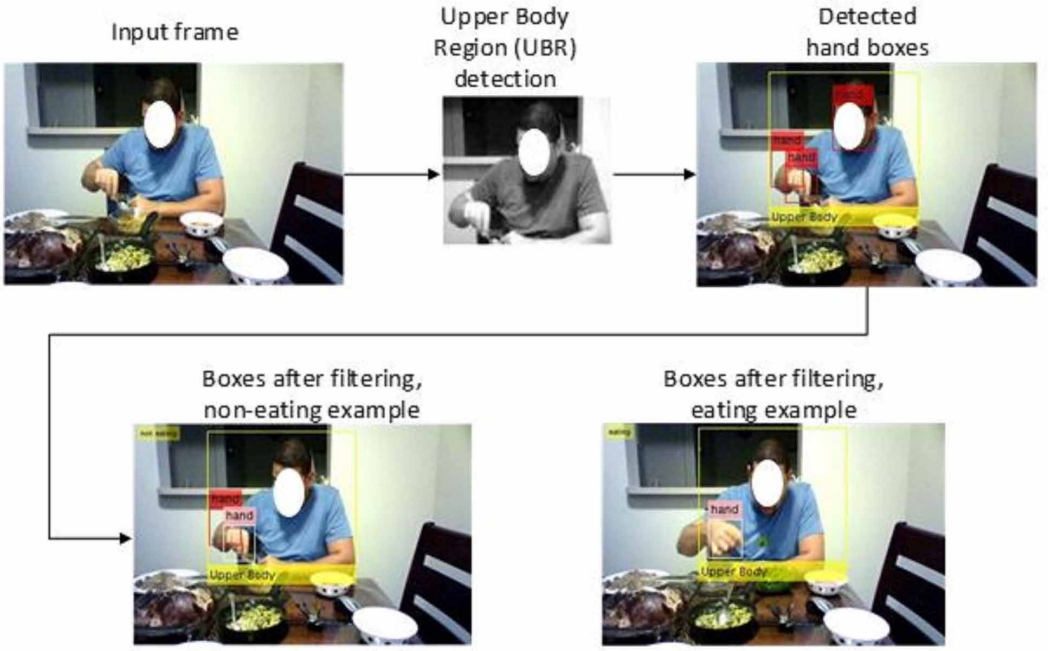


Table 1. Implementation results summary

No. of Processed Videos	No. of Processed Frames	Total no. of eating events	No. of Success Detections	Missed/Wrong Detections	Recording Time (hh:mm:ss)	Accuracy
33	163840	920	893	59	02:57:34	90.65%

Figure 14. Samples of the eating process detection and tracking showing successful eating event detection for persons who is dinning



the hands are used to locate the study subject's face and hands. In (Gao, et al., 2004) the optic flow is used for motion objects segmentation, then these objects are tracked within a specified time frame window while looking for moving objects with a consistent moving direction. When HMM is used to track certain features such as the physical distance between moving regions, their system achieved a good accuracy rate with 77%. Comparatively speaking, our method can result in detecting eating with even greater accuracy.

5. CONCLUSION

In this paper, we propose a novel vision-based eating monitoring system for Alzheimer's patients. The eating recognition is implemented via HOG. HOG extracted the features from the region of interest, which is the UBR in our case. A trained SVM has been used to distinguish between eating v. non-eating gestures. To reduce the false positive detections, the hand detection via Harr-like features has been evaluated to detect hand movement inside the UBR. When a hand is detected, the HOG is determined for the UBR with SVM for the recognition of eating v. non-eating gesture.

The proposed hand detection via Haar-like features can also result in the detection of false positives which affects the gesture recognition process. False positives can be reduced by two filtering stages: first, the detected regions above the head will be discarded; second, a CIT matching method is used for hand correlation detection. Using the proposed algorithm with a very low failure rate, our results achieve a high accuracy of 90.65%. Errors occur when hand is not detected successfully due to different hand gestures and fast motion affect.

The study has been tested on subjects with different skin color, different scenes' backgrounds, and different eating gestures. Therefore, the number of false positive detections must be eliminated. The eating recognition is enhanced with hand detecting within the UBR, while the hand detection is enhanced with the CIT matching method to improve the hand detection and cover most hand-while-eating postures.

The number of persons with Alzheimer's disease and other related dementias is increasing rapidly. The degenerative nature of the Alzheimer's translates to the need for substantial hands-on care. . The care provided for Alzheimer's society is burdening and costly. Our food intake monitoring system via physical gestures recognition method is presented as a preliminary step in designing an intervention to support caregiving and possibly malnutrition reduction goals. In addition to monitoring, our system has the potential to provide prompting cues, both audio and video. In this way, accurately detecting eating behaviors and providing corollary prompts offers a promising strategy for prolonging independence, thereby supporting both the person with the disease and their care providers. This way the burden will be reduced on the care givers. Also, the system will provide essential data of the eating habits for those patients.

As a future work, we will investigate the use of the deep learning methods for eating recognition. Recently, deep learning methods have been used by many researchers and results are encouraging. The detection of mouth of the subject and food chewing can be a useful addition to system and may further improve the eating recognition success rate.

REFERENCES

Al-Anssari, H., & Abdel Qader, I. (2016). Vision Based Monitoring System for Alzheimer's Patients Using Controlled Bounding Boxes Tracking. *IEEE International Conference on Electro Information Technology (EIT)*. doi:10.1109/EIT.2016.7888847

Al-Anssari, H. A., Abdel-Qader, I., & Mickus, M. (2018). Monitoring System for Persons With Alzheimer's Disease via Video-Object Tracking. *International Journal of Mobile Devices, Wearable Technology, and Flexible Electronics*, *9*(2), 18–36. doi:10.4018/IJMDWTFE.2018070102

Alzheimer's Disease: Facts & Figures. (2018). Retrieved from https://www.brightfocus.org/alzheimers/article/alzheimers-disease-facts-figures

Alzheimer's Disease Facts and Figures. (2017). Alzheimer's & Dementia. *The Journal of the Alzheimer's Association*, *13*(4), 325–373.

Alzheimer's Statistics. (2018). Retrieved from https://www.alzheimers.net/resources/alzheimers-statistics/

Barngrover, C. (2014). *Automated Detection of Mine-Like Objects in Side Scan Sonar Imagery.* UC San Diego. ProQuest ID: Barngrover_ucsd_0033D_14031. Merritt ID: ark:/20775/bb4506685c. Retrieved from https://escholarship.org/uc/item/4gw8g426

Bilal, S., Akmeliawati, R., Salami, M., Shafie, A., & Bouhabba, E. (2010). A hybrid method using haar-like and skin-color algorithm for hand posture detection, recognition and tracking. *2010 IEEE International Conference on Mechatronics and Automation*, 934-939. doi:10.1109/ICMA.2010.5588576

Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.

Chard, G., Liu, L., & Mulholland, S. (2009). Verbal Cueing and Environmental Modifications: Strategies to Improve Engagement in Occupations in Persons with Alzheimer Disease. *Physical & Occupational Therapy in Geriatrics*, *27*(3), 197–211. doi:10.1080/02703180802206280

Chen, Q. (2008). Hand Gesture Recognition Using Haar-Like Features and a Stochastic Context-Free Grammar. *IEEE Transactions on Instrumentation and Measurement*, *57*(8), 1562–1571. doi:10.1109/TIM.2008.922070

Chowdhury, S. A., Kowsar, M. M., & Deb, K. (2018). Human detection utilizing adaptive background mixture models and improved histogram of oriented gradients. *ICT Express*, *4*(4), 216–220. doi:10.1016/j.icte.2017.11.016

Cunha, A., Pádua, L., Costa, L., & Trigueiros, P. (2014). Evaluation of MS Kinect for Elderly Meal Intake Monitoring. *Procedia Technology*, *16*(C), 1383–1390. doi:10.1016/j.protcy.2014.10.156

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* doi:10.1109/CVPR.2005.177

Déniz, O., Bueno, G., Salido, J., & De La Torre, F. (2014). Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, *32*(12), 1598–1603. doi:10.1016/j.patrec.2011.01.004

Dong, B., & Biswas, S. (2016). Meal- time and duration monitoring using wearable sensors. *Biomedical Signal Processing and Control*, *32*, 97–109. doi:10.1016/j.bspc.2016.09.018

Gao, Hauptmann, Bharucha, & Wactlar. (2004). Dining activity analysis using a hidden Markov model. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, 2*, 915-918.

Hsieh, Liou, & Lee. (2010). A real time hand gesture recognition system using motion history image. *2010 2nd International Conference on Signal Processing Systems, 2*, V2-394-V2-398.

Huang, T. F., Chao, P. C., & Kao, Y. Y. (2012). Tracking, recognition, and distance detection of hand gestures for a 3-D interactive display. *Journal of the Society for Information Display*, *20*(4), 180–196. doi:10.1889/JSID20.4.180

Iosifidis, A., Marami, E., Tefas, A., Pitas, I., & Lyroudia, K. (2015). The MOBISERV-AIIA Eating and Drinking multi-view database for vision-based assisted living. *Journal of Information Hiding and Multimedia Signal Processing*, *6*(2), 254–273.

Johansson, L., Christensson, L., & Sidenvall, B. (2011). Managing mealtime tasks: Told by persons with dementia. *Journal of Clinical Nursing*, *20*(17-18), 2552–2562. doi:10.1111/j.1365-2702.2011.03811.x PMID:21762416

Johnson. (2012). *Computerized recognition of solar cavities*. Academic Press.

Kai, K., Hashimoto, M., Amano, K., Tanaka, H., Fukuhara, R., Ikeda, M., & Ginsberg, S. (2015). Relationship between Eating Disturbance and Dementia Severity in Patients with Alzheimer's Disease. *PLoS One*, *10*(8), E0133666. doi:10.1371/journal.pone.0133666 PMID:26266531

Kalantarian, H., Alshurafa, N., Le, T., & Sarrafzadeh, M. (2015). Monitoring eating habits using a piezoelectric sensor-based necklace. *Computers in Biology and Medicine*, *58*, 46–55. doi:10.1016/j.compbiomed.2015.01.005 PMID:25616023

Lin, J., & Ding, Y. (2013). A temporal hand gesture recognition system based on hog and motion trajectory. *Optik (Stuttgart)*, *124*(24), 6795–6798. doi:10.1016/j.ijleo.2013.05.097

Madara Marasinghe, K. (2016). Assistive technologies in reducing caregiver burden among informal caregivers of older adults: A systematic review. *Disability and Rehabilitation. Assistive Technology*, *11*(5), 353–360. doi:10.3109/17483107.2015.1087061 PMID:26371519

Mendi, E., Ozyavuz, O., Pekesen, E., & Bayrak, C. (2013). Food intake monitoring system for mobile devices. *5th IEEE International Workshop on Advances in Sensors and Interfaces IWASI.* doi:10.1109/IWASI.2013.6576082

Newell, A. (2011). Invariant encoding schemes for visual recognition. Academic Press.

Päßler, S. J., & Fischer, W. (2011). Acoustical method for objective food intake monitoring using a wearable sensor system. *2011 5th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2011,* 266-269.

Pietikäinen, M., Hadid, A., Zhao, G., & Ahonen, T. (2011). Local binary patterns for still images. *Computer Vision Using Local Binary Patterns,* 13-47.

Qiu, J., Lo, F. P., & Lo, B. (2019). Assessing Individual Dietary Intake in Food Sharing Scenarios with a 360 Camera and Deep Learning. *IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN),* 1-4.

Rasheed, S., & Woods, R. (2013). Malnutrition and quality of life in older people: A systematic review and meta-analysis. *Ageing Research Reviews*, *12*(2), 561–566. doi:10.1016/j.arr.2012.11.003 PMID:23228882

Rosa, B. G., Anastasova-Ivanova, S., Lo, B., & Yang, G. Z. (2019). Towards a Fully Automatic Food Intake Recognition System Using Acoustic, Image Capturing and Glucose Measurements. *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1-4.

Saberian, M. (2012). Multiclass boosting for fast multiclass object detection. Academic Press.

Sharma, R. (2014). Object detection using dimensionality reduction on image descriptors. Academic Press.

Shuzo, M., Lopez, G., Takashima, T., Komori, S., Delaunay, J. J., Yamada, I., Tatsuta, S., & Yanagimoto, S. (2009). Discrimination of eating habits with a wearable bone conduction sound recorder system. *2009 IEEE Sensors,* 1666-1669.

Sun, M., Liu, Q., Schmidt, K., Yang, J., Yao, N., Fernstrom, J. D., Fernstorm, M. H., DeLany, J. P., & Sclabassi, R. J. (2008). Determination of food portion size by image processing. *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology",* 871-874. doi:10.1109/IEMBS.2008.4649292

Takahashi, M., Fujii, M., Shibata, M., & Satoh, S. (2010). Robust recognition of specific human behaviors in crowded surveillance video sequences. *EURASIP Journal on Advances in Signal Processing*, *2010*(1), 14. doi:10.1155/2010/801252

Tribaldos, P. T., Serrano-Cuerda, J. J., López, M., Fernández-Caballero, A., & López-Sastre, R. (2013). People detection in color and infrared video using HOG and linear SVM. Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7931(2), 179-189

Villalobos, G., Almaghrabi, R., Hariri, B., & Shirmohammadi, S. (2011). A personal assistive system for nutrient intake monitoring. *Proceedings of the 2011 International ACM Workshop on Ubiquitous Meta User Interfaces*, 17-22. doi:10.1145/2072652.2072657

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1*. doi:10.1109/CVPR.2001.990517

Waghmare, S. (2012). Comparative study of feature-selection sliding. Academic Press.

Wei, Y., Tian, Q., & Guo, T. (2013). An Improved Pedestrian Detection Algorithm Integrating Haar-Like Features and HOG Descriptors. *Advances in Mechanical Engineering*, *5*, 8. doi:10.1155/2013/546206

Zhang, X., Gonnot, T., & Saniie, J. (2017). Real-time face detection and recognition in complex background. *Journal of Signal and Information Processing*, *8*(2), 99–112. doi:10.4236/jsip.2017.82007

Haitham Al-Anssari received his Ph.D. in Electrical and Computer Engineering from Western Michigan University in 2018. He is interested in computer vision, image processing, behavior recognition, machine learning, and real-time implementation, embedded systems, and ADAS applications.

Ikhlas Abdel-Qader (PhD), PE, is currently a professor in the Department of Electrical and Computer Engineering at Western Michigan University (WMU) in Kalamazoo, Michigan and has been a registered professional engineer (PE) since 1996. Her research and teaching interests include Digital Signal and Image Processing, Pattern Recognition, and Feature Extraction; Medical Signal Processing; Medical Imaging Systems, and Diagnostic Feature Extraction; Machine Learning and Predictive Analytics,V2X Communications and Positioning Algorithms; Fault Analysis in Power Systems; V2G and smart Grid Integration, Reliability, Efficiency, and Resilience; Cyber-Physical Security; and Non-Destructive Testing and Evaluation. She received funding in excess of $5 million from various agencies such as National Science Foundation, Michigan Department of Transportation, Calhoun County, and Texas Instruments to support her research activities. She is a Senior Member of the Institute of Electrical and Electronic Engineers (IEEE); the IEEE Acoustics, Speech, and Signal Processing Society; and the IEEE Engineering in Medicine and Biology Society. She is also a member of the Society of Women Engineer.

Maureen Mickus, Ph.D., MSG, is a gerontologist and Professor in the Department of Occupational Therapy at Western Michigan University. She received her bachelor's degree at Kalamazoo College and master's degree from the Andrus Gerontology Center at the University of Southern California. She was awarded a Ph.D. from Northwestern University and subsequently completed a post-doctoral fellowship at Michigan State University's College of Human Medicine. Dr. Mickus served on the medical school faculty at Michigan State University for ten years, teaching and conducting aging related research. She joined the faculty at Western Michigan University in 2006 where she received the Western Michigan University Distinguished Teaching Award in 2011. She has published on a variety of mental health and aging topics, including dementia and depression. Additionally, her research has involved health policy issues relating to frail elders such as the turnover of direct care staff in long term care facilities. She has also conducted a study using sensor technology to maintain persons with Alzheimer's disease in the home setting. Dr. Mickus serves on multiple boards, including the Michigan Great Lakes Alzheimer's Association.