# REAL-TIME HEAD NOD AND SHAKE DETECTION FOR CONTINUOUS HUMAN AFFECT RECOGNITION

*Haolin Wei*[1], *Patricia Scanlon*[2], *Yingbo Li*[1], *David S. Monaghan*[1], *Noel E. O'Connor*[1]

[1]CLARITY: Centre for Sensor Web Technologies, Dublin City University, Ireland
[2]Bell Labs Ireland, Alcatel Lucent Dublin, Ireland

## ABSTRACT

Human affect recognition is the field of study associated with using automatic techniques to identify human emotion or human affective state. A person's affective states is often communicated non-verbally through body language. A large part of human body language communication is the use of head gestures. Almost all cultures use subtle head movements to convey meaning. Two of the most common and distinct head gestures are the head nod and the head shake gestures. In this paper we present a robust system to automatically detect head nod and shakes. We employ the Microsoft Kinect and utilise discrete Hidden Markov Models (HMMs) as the backbone to a to a machine learning based classifier within the system. The system achieves 86% accuracy on test datasets and results are provided.

## 1. INTRODUCTION

Nonverbal behaviors such as head gestures, body language, facial expression and eye contact play an import role in daily communications. As the most common head gestures, head nod and shake are usually used as semantic functions (e.g. nodding means yes, and shaking means no), affect indication (e.g. nodding means approval or acceptance) and conversational feedback (e.g. keep conversation flow), at least in Western Europe. Therefore, the detection of head nods and shakes can be seen as a valuable module for achieving affect recognition and natural human-computer interaction. In this paper we describe a new system that detects head nod and shake in real time. We use Microsoft Kinect and Kinect for Windows SDK to estimate head pose robustly. The direction of head movement is then determined based on the head pose and used by a discrete Hidden Markov Model (HMM) classifier as the observation sequence to detect whether head nod or shake occurs.

## 2. RELEATED WORK

Much work has been done on head nod and shake detection stretching back over a decade. The related work presented in [1] proposed a head gesture recognition system for interfaces. The IBM PupilCam is first used to obtain the location of the user's face. Based on the face location, a Timed Finite Sate Machine is used to detect head nod and shake and the results are used to drive a perceptual dialog-box agent (e.g., nod=YES). Similarly, the authors of [2] present a system that uses a customized IR camera for pupil tracking and a discrete Hidden Markov Model to detect head nod and shake. Kawato and Ohya [3] have described another system to detect head nods and shakes in real time by directly detecting and tracking the between-eyes region using a webcam. Combining the circle frequency filter together with skin color information and template, the between-eyes region could be detected and tracked. A rule based detection algorithm is then applied to the movement of the between-eyes region to detect head nods and shakes. Because of its simple rule based detection, some non-regular head nods and shakes may not be detected. It should be noted that all the systems mentioned above need to track the eye pupil position in order to detect head nod and shake, and will not be able to detect any head gesture if the user's eyes are closed. In [4] the authors present another method to detect head nodding and shaking in real time from video streams. The AdaBoost algorithm is first used to detect the user's face and based on the physiological information of the eye location in the face, eye location can be obtained in each frame. The direction of head movement is calculated based on eye location and used as an observation sequence for a discrete HMM to detect head nods and shakes. The authors in [5] present a new method for head nods and shakes detection by using 3D cylindrical head model (CHM) and dynamic template to estimate head pose and use the accumulative Hidden Markov Models to detect head nod and shake.

In this paper, we present a new method to robustly detect head nods and shakes in real time using the Microsoft Kinect. Despite a lot of work on head nod/shake in the past, to the best of our knowledge this has not been widely explored with the Kinect due to its relatively recent introduction. We first use the Microsoft Kinect for Windows SDK to estimate the head pose of the user. The change of head pose in each frame indicates the direction of the head movements that are then used as an observation sequence by a discrete Hidden Markov Models (HMMs) to detect if a head nod or shake occurs. The

proposed system runs fast and can detect the head nods and shakes in real time on a standard desktop PC. The approach can also robustly detect non-obvious and non-regular head nods and shakes.

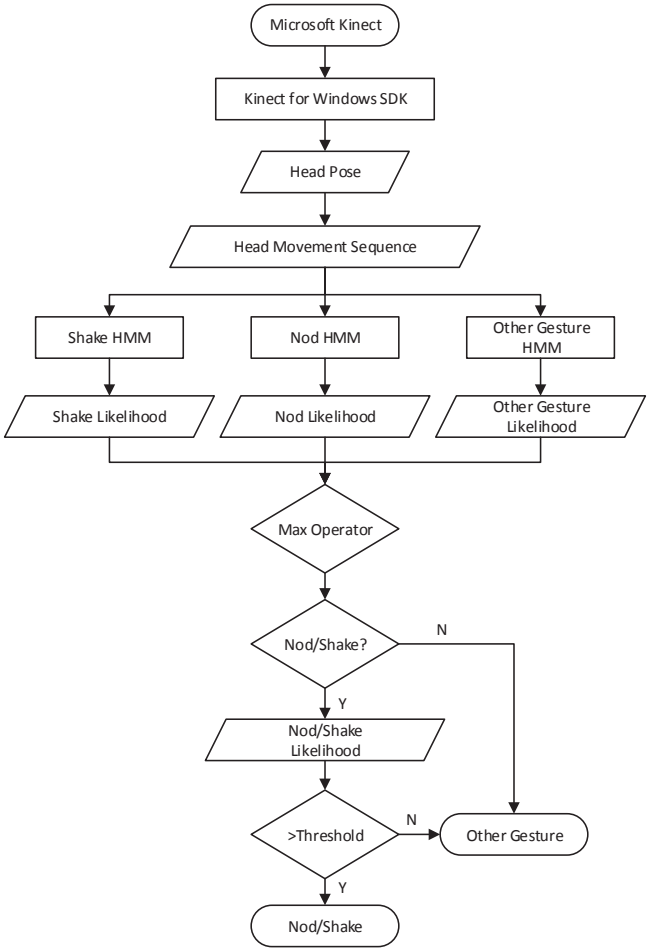


**Fig. 1**. System Overview

## 3. SYSTEM OVERVIEW

The overall architecture of the system is shown in Fig.1. The head pose is first obtained from Kinect through the Kinect for Windows SDK. Then, the head pose in a temporal window is analyzed as a sequence of head movements. Finally, we use three HMMs to detect the presence of head shake, head nod and other head gestures in this sequence of head movements. The largest likelihood value is selected as the detection result. In order to further distinguish head nod and shake from other head gestures, a predefined threshold is used. More details are described in this section.

### 3.1. Head Pose Estimation

Head pose estimation has received a lot of attention recently as a key element of human behavior analysis. With depth cameras such as Microsoft Kinect becoming available at commodity prices, the research focus of head pose estimation have shift from 2D video data based to depth data based and have shown very good results compared to 2D approach [6, 7]. The Microsoft Kinect supports the capture of 2D RGB streams and 3D depth streams at 30 frames per second, based on infrared projection and light coding techniques. However, the Kinect depth information is not very accurate and much nosier compared to the data obtained from other devices, such as a laser-scanner, for example. In order to estimate the head pose, the method described in [8] was used. The method utilizes a regularized maximum likelihood deformable model fitting (DMF) algorithm to reduce the effect of the noisy depth map acquired from the Kinect and to improve the accuracy of the estimation results. As this method has been implemented in the recent release of Kinect for Windows SDK, we use it directly to obtain the head pose of the user. The SDK gives head pose with respect to the Kinect by three angles: pitch, roll and yaw, as illustrated in Fig.2. The angles are expressed in degrees, with values ranging from -90 to +90 degrees.

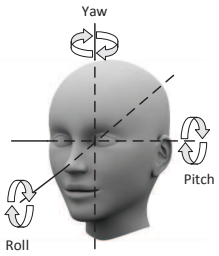


**Fig. 2**. Yaw, Pitch and Roll

### 3.2. Head Nod and Shake Detection

Although head nods and shakes could be performed differently by different people in terms of intervals and amplitudes, some common characteristics still exist for the head movement to be recognized as nods or shakes. In this paper we consider a nod as the head tilted in an alternating down and up manner, whereas a shake is rotation of the head horizontally from side-to-side. This is shown in Fig.3. The vertical and horizontal movement could be represented by pitch and yaw in terms of head pose as shown in Fig.2. By comparing the difference of pitch and yaw in two adjacent frames, the direction of head movement can then be determined. Following [2], the direction is represented by five directional symbols (Up, Down, Left, Right and Still). Based on the five states of head movements, three Hidden Markov Models (HMM) termed nodHMM, shakeHMM and otherHMM are trained. The nodHMM consists of three states Up, Down and Still,

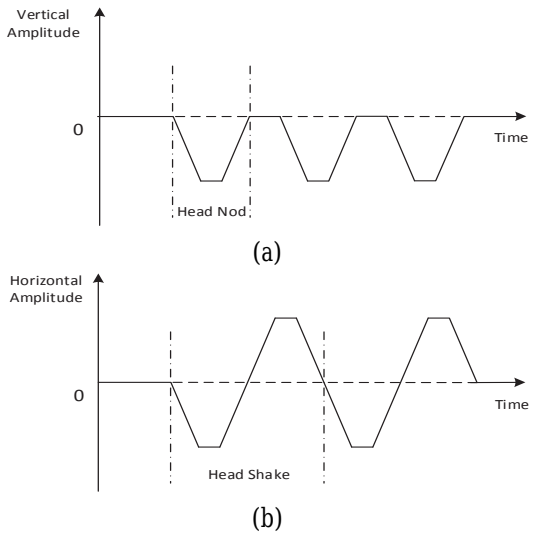


**Fig. 3**. (a) Typical Nod Sequence. (b) Typical Shake Sequence.

whereas the shakeHMM consists Left, Right and Still. Both HMMs have five observation states Up, Down, Left, Right and Still. To further distinguish other head movements (E.g. moving up, moving down, moving left and moving right) from the actual head nods and shakes, we first build an additional HMM, termed otherHMM, which consists of five states Up, Down, Left, Right and Still to recognize head gestures except head nods and shakes, and then we compare the nod or shake likelihood values to a predefined threshold. The state transitions of head nod, head shake and other gestures is shown in Fig.4.

In order to analyze head movement continuously, we choose a window size of 0.6 seconds similier as [4], corresponding to 18 frames/sec, which we found sufficient to detect both slow as well as subtle head nods and shakes. During the training phase, we extract the head pose using the method mentioned in section 3.1 for each frame and formed an observation sequence of 18 frames. Since it is impossible for 18 frames to compromise all the actions of head nod and shake we consider the sequence containing down as head nod and any obvious Left or Right as head shake too. The Baum Welch algorithm [9] is used to train the nodHMM, shakeHMM and otherHMM based on the observation sequence. In the testing phase, the forward-backward procedure [9] is used to compute the log likelihood for the input observation sequence on three HMMs. The largest likelihood value is selected, and if a head nod or shake is detected, it is further compared to a predefined threshold. If the likelihood value is larger than the predefined threshold the observation sequence is considered to be a nod or a shake, otherwise it is considered to be other head gesture such as still or looking upward.

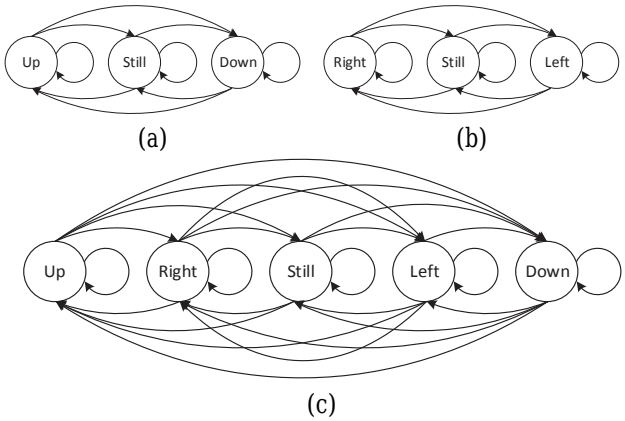


**Fig. 4**. (a) Transition of nodHMM's hidden states. (b) Transition of shakeHMM's hidden states. (c) Transition of otherHMM's hidden states

| | Recognized As | | |
|---|---|---|---|
| | Head Nods | Head Shakes | Other |
| Head Nods | 22 | 0 | 3 |
| Head Shakes | 0 | 23 | 2 |
| Other | 4 | 1 | 20 |

**Table 1**: Recognition Results for Training set

| | Recognized As | | |
|---|---|---|---|
| | Head Nods | Head Shakes | Other |
| Head Nods | 21 | 0 | 4 |
| Head Shakes | 0 | 22 | 3 |
| Other | 5 | 1 | 19 |

**Table 2**: Recognition Results for Testing set

## 4. EXPERIMENTS RESULTS

We collected a database of manually performed head nods, shakes and other gestures to train the HMMs. Microsoft Kinect and Kinect studio were used to capture the head motion. In total 150 samples with 50 head nods, 50 head shakes and 50 other head gestures (including still, look upward, look downward, look leftward and look rightward) were collected and manually annotated. These head nods and shakes are of obvious motions of nod and shake in different motion magnitudes. Thus, we ensured that the trained HMM classifiers were suitable for the head nods or shakes with small or big head motions. A random 50% of the each gesture class is selected for training to estimate the parameters of nod, shake and other HMMs.

After training, the estimated parameters and the detection algorithm were implemented on an Intel Core i7 3.4GHz machine with Windows 7 with the Kinect placed under the monitor. The details of recognition results are shown in Table 1 and

2. This performance appears to be comparable, if not better, to the results obtained by other methods, such as in [2] and [4]. Future work will investigate this more fully by applying those techniques to our dataset.

From the results we can see there is no misclassification among head nods and head shakes. Most missed head nods are due to the head gestures such as look downward and look upward and missed head shakes are due to look leftward and look rightward motions.

A demonstration system has been developed to visualize the estimated head pose value and show the detection results in a bar chart form, shown in Fig.5. When the head is detected by the Kinect, a 3D mesh will appear on the face. At the same time, the detection of head nod and shake begins to work. We visualize the Pitch, Yaw and Roll data from Kinect for Windows SDK. The real-time data of Pitch, Yaw, Raw, number of sequence, and the head position relative to the position of the Kinect is displayed. We finally show the detection results of Nod, Shake and None by HMM classifiers with above data. The classifier with the maximum probability is the final detection results.

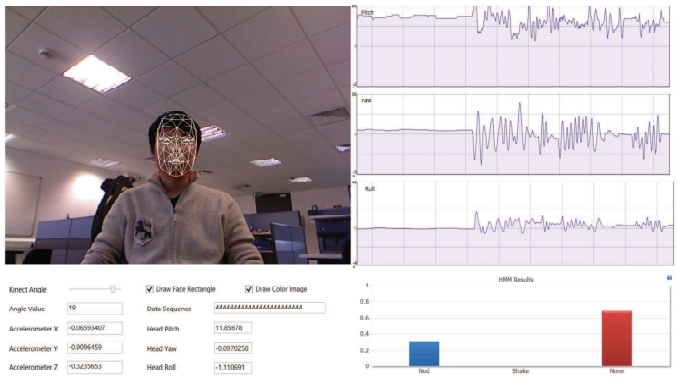


**Fig. 5**. Screenshot of the system in operation.

## 5. CONCLUSION

We have presented a system for real time detection of head nods and shakes. The Microsoft Kinect and Kinect for Windows SDK is used to estimate the head pose. Based on the difference of head pose angle in two consecutive frames, the direction of head movement could be determined. The head movement direction sequence is then feed into nod and shake HMMs for head gesture detection. A program is designed to visualize the angle value and detection results over time. It runs at 30 frames and a recognition accuracy of 86% is achieved. Our work extends the application of Microsoft Kinect towards future research of semantic analysis of head gesture.

## 6. ACKNOWLEDGEMENTS

This work is co-funded by Bell Labs Ireland and the Irish Research Council under the Enterprise Partnership scheme. This work is partly supported by the EU FP7 project REVERIE, ICT-287723

## 7. REFERENCES

[1] James W. Davis, "A perceptual user interface for recognizing head gesture acknowledgements," in *In ACM Workshop on Perceptual User Interfaces*, 2001, pp. 15–16.

[2] Ashish Kapoor and Rosalind W. Picard, "A real-time head nod and shake detector," in *in Proceedings from the Workshop on Perspective User Interfaces*, 2001.

[3] Shinjiro Kawato and Jun Ohya, "Real-time detection of nodding and head-shaking by directly detecting and tracking the "between-eyes"," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, Washington, DC, USA, 2000, FG '00, pp. 40–, IEEE Computer Society.

[4] Wenzhao Tan and Gang Rong, "A real-time head nod and shake detector using hmms," *Expert Systems with Applications*, vol. 25, no. 3, pp. 461 – 466, 2003.

[5] Ohryun Kwon, Junchul Chun, and Poem Park, "Cylindrical model-based head tracking and 3d pose recovery from sequential face images," in *Hybrid Information Technology, 2006. ICHIT '06. International Conference on*, 2006, vol. 1, pp. 135–139.

[6] M.D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, and H. Pfister, "Real-time face pose estimation from single range images," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.

[7] Gabriele Fanelli, Thibaut Weise, Juergen Gall, and Luc Van Gool, "Real time head pose estimation from consumer depth cameras," in *Proceedings of the 33rd international conference on Pattern recognition*. 2011, DAGM'11, pp. 101–110, Springer-Verlag.

[8] Qin Cai, David Gallup, Cha Zhang, and Zhengyou Zhang, "3d deformable face tracking with a commodity depth camera," in *Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III*. 2010, ECCV'10, pp. 229–242, Springer-Verlag.

[9] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.