





Development and Evaluation of a Mouse Emulator Using Multi-modal Real-Time Head Tracking Systems with Facial Gesture Recognition as a Switching Mechanism

Shivanand P. Guness¹ , Farzin Deravi² , Konstantinos Sirlantzis²,
Matthew G. Pepper^{2,3}, and Mohammed Sakel³

¹ net.connection Ltd., Moka, Mauritius
shivam.guness@ieee.org

² School of Digital Arts and Engineering, University of Kent, Canterbury, Kent, UK
{f.deravi,k.sirlantzis}@kent.ac.uk

³ East Kent University Hospitals Trust, Canterbury, Kent, UK
Matthew.Pepper@ekht.nhs.uk, msakel@nhs.net

Abstract. The objective of this study is to evaluate and compare the performance of a set of low-cost multi-modal head tracking systems incorporating facial gestures as a switching mechanism. The proposed systems are aimed to enable severely disabled patients to access a computer. In this paper, we are comparing RGB (2D) and RGB-D (3D) sensors for both head tracking and facial gesture recognition. System evaluations and usability assessment were carried out on 21 healthy individuals. Two types of head tracking systems were compared - a web camera-based and another using the Kinect sensor. The two facial switching mechanisms were eye blink and eyebrows movement. Fitts' Test is used to evaluate the proposed systems. Movement Time (MT) was used to rank the performance of the proposed systems. The Kinect-Eyebrows system had the lowest MT, followed by the Kinect-Blink, Webcam-Blink and Webcam-Eyebrows systems. The 3D Kinect systems performed better than the 2D Vision systems for both gestures. Both Kinect systems have the lowest MT and best performance, thus showing the advantage of using depth.

Keywords: Assistive technology · Facial gesture recognition · Fitts' test · Eye blink detection · Eyebrow movement

1 Introduction

The World Health Organization (WHO) estimated that 1 billion people around the world live with some form of disability [36]. Approximately 10 million people in UK have disabilities with a neurological diagnosis. For a multitude of reasons, the number of people with profound disability stemming from neurological disorders is increasing with a resulting impact on their quality of life and that of

their caregiver. The cost of caring for neuro-disabled persons in Europe has been estimated as 795 Billion Euro [17]. The value of assistive technologies in improving the quality of life of people with disability and also reduce carer strain is emphasized in a 2010 Royal College of Physicians Report [30].

For many individuals with disability access to a computer and/or communication aid may help mitigate the effect of communication impairments. Often this can be achieved through the identification of suitable access sites e.g. hand, foot, arm or head. Some patients, however, are profoundly disabled that they might be unable to talk but can only make small head movements and facial gestures such as eye blink or eyebrow movement. In some cases there may not even be enough head movement to enable the use of an access device such as a head tracker like SmartNav [2] and so the only remaining access site may be small facial gestures. Although there are other options available - e.g. the use of eye gaze, existing systems using eye gaze technology such as MyTobii [3] are complex, expensive and set-up/configuration places a significant burden on both the user and the caregiver.

The motivation for the work reported in this paper is the need for low-cost, reliable head tracking with an automatic facial gesture recognition system to help severely disabled users access electronic assistive technologies. The objective is to develop a multi-modal head tracking system, which uses facial gestures as a switching mechanism thus enabling severely disabled patients whose control is restricted to small head movements and facial gestures to be able to access a computer.

2 Background

Pistori [29], states that assistive devices using computer vision can have a great impact in increasing the digital inclusion of people with special needs. Computer vision can improve both the devices used for mobility i.e. controlling motorised wheel chairs, sign language detection and head trackers. Similarly, Betke et al. [7], describe the advances made in the development of assistive software and the use of emerging technology can lead to the creation of intelligent interfaces using both assistive technology and human computer interaction (HCI). The example of the CameraMouse [9] is used as an interface system with different assistive devices and software such as Midas Touch [6], Dasher [34] etc. are included to highlight the use of HCI and assistive devices.

Abascal et al. [4], highlighted some opportunities and challenges that designing human-computer interfaces suitable for the disabled can pose. For people suffering from disabilities, HCI can be used to design better interfaces which could be accessible to people with disabilities and thus improve socialisation, better access to communication facilities and have a greater control over their environment.

2.1 Device Evaluation

Fitts' test [14] was developed in 1954 to model human movement. The result of the experiments showed that the rate of performance of the human motor system is approximately constant over a wide range of movement amplitudes. Mackenzie et al. [23], adapted the Fitts' Law for assessing HCI. This work was later embedded in an International Standard for HCI, ISO 9431-9:2000 [18] providing guidelines for measuring the users' performance, comfort and effort. The performance of the device was measured by making the user perform tasks using the device. There are six types of tasks - one-direction, multi-directional, dragging, free-hand tracing (drawing), and, hand input, grasp and park (homing/device switching). ISO 9431-9:2000 [18] requires that the input device be tested for at least 2 different Index of Difficulty (ID). Index of Difficulty (ID) is a measure of the difficulty of the task [5]. In Douglas et al. [12], the validity and practicality of the ISO framework using both multi-directional and the one-direction Fitts' Tests for two devices namely a touch-pad and a joystick was investigated.

2.2 Gesture Detection

In this paper, the interest is in processing video information to recognise blink and eyebrow movement gestures. The detected gestures can be to emulate a mouse click or a switch action to access and control a computer/communication aid.

Grauman et al. [15] proposed two systems called BlinkLink and EyebrowsClicker. The BlinkLink software tracked both the motion within the eye region and the eye region itself. The EyebrowsClicker tracked the eyebrows region and detected the rising and falling of the eyebrows. To initialise the location of the eye and eyebrows regions, the user has to perform the gestures and by analysing the area of motion on the face, the respective regions are detected. A template of each region is generated. The correlation score of the eye region and a template of both the closed eye and open eye were compared to detect an eye blink. For eyebrows gesture, the distance from the eyes and the eyebrows are monitored to detect the rise and fall motion of the eyebrows. Blink detection had an overall success rate of around 95.6% and was tested on 15 healthy individuals and one person suffering from Traumatic Brain Injury (TBI). EyebrowsClicker had an overall success rate of 89% and it was tested with six individuals, but the software had to be reinstated twice during the data capture session because the tracking of the eyebrows was lost. There has been no further published work on this system.

Malik et al. [25] proposed a blink detection method using histogram of Local Binary Patterns (LBP) [27]. A template of open eye was generated using the average histogram of LBP from a sample of 50 images of an open eye. The histogram of LBP of images of the eye region were compared against the template using the Kullback-Leibler Divergence (KLD) method. In KLD, the distance between two distribution is zero only if the distributions are identical. KLD was found to be robust against both the precision of the eye detection and

the variation in the window size of the detected eye region. The eye region are obtained using the Viola-Jones [33] algorithm implemented in OpenCV. The proposed algorithm was tested against the ZJU Eye blink Database [28] and resulted in a 99.2% blink detection rate. Missimer et al. [26] proposed a blink detection algorithm based on the analysis of the differences in three consecutive images. Blobs are generated from the merging of two difference images produced. Three points are used for tracking, the centre of the upper lip and the upper part of both eyebrows. In addition, optical flow is used to track these three points. The eye templates are generated based on the tracked points and used to train the system. The system is reported to having a success rate of 96.6% and was tested on 20 healthy individuals.

Yunqi et al. [37] proposed an eye blink detection algorithm which was used in a drowsiness driver warning system. The proposed system used Haar-like [32] features and AdaBoost to detect the face of the user. Some pre-processing was performed on the image and an edge detection algorithm was used to find the eye corners, the iris and the upper eyelid for each eye. The curvature of the upper eyelid was compared with the line connecting the two eye corners and if most of the upper eyelid curvature was under this line, the eye was considered closed. The algorithm was tested on images captured during a real driving session and 94% accuracy was obtained for the eye state detection.

In Zhang et al. [38], proposed a Gaze based assistive application on a smart-phone to enable the user to communicate. The application can recognise six gestures from both eyes namely look up, look down, look right, look left, look center and closed eyes. The algorithm used OpenCV [8] and Dlib-ML [21]. Before using the device, calibration must be performed to create templates for each gesture. The template is created by making the users perform the gesture and capturing the image of the eye region when the action is performed. The algorithm detected the gestures with an accuracy of 86% on average. The accuracy rate decreased to 80.4% for people wearing glasses, increased to 89.0% for people wearing contact lenses and increased to 89.7% for people without glasses.

In Val et al. [11], eye blinks are used to control a robot. An infra-red emitter and an optical sensor were used to detect the eye blink. The blinks are used to navigate the robotic assistive aid, for example a right eye blink would cause the robot turn right and a left eye blink would make the robot turn left. A combination of the left blink followed by a right blink would cause the robot stop. In Krolak et al. [22], the proposed method uses two active contour [20] models - one for each eye - for detecting eye blinks. Haar-like features [32] are used to detect the face and the location of the eyes are determined using known geometrical proportion of the human face.

In Tuisku et al. [31], the evaluation of a system called Face Interface was conducted. The system used voluntary gaze direction for moving the cursor around the screen and facial muscle activation for the selecting objects on a computer screen. Face Interface used two different muscle activation - frowning and raising the eyebrows. A series of points were presented to the user. The time to complete the tasks and the accuracy of the activation were used as performance

measure. The pointing tasks were conducted using three different target diameters (i.e. 25, 30, 40 mm), seven distances (i.e., 60, 120, 180, 240, 260, 450, and 520 mm), and eight pointing angles namely (0° , 45° , 90° , 135° , 180° , 225° , 270° , and 315°). It was found that for distances between 60 mm and 260 mm, tasks performed using the raising eyebrow selection technique were faster than those using the frowning technique. Also, the overall time taken to complete the tasks were 2.4 s for the frowning technique and 1.6 s for the raising technique. The *IP* of the frowning techniques was 1.9 bits/s and 5.4 bits/s for the eyebrow raising technique.

The systems reported here were limited in that they would only work with frontal facial images and were not robust in coping with posture changes. The work reported here aims to address these shortcomings by making use of the depth data available from RGB-D sensors.

3 Materials and Methods

The systems evaluated in this work incorporate a camera and an algorithm for tracking the head movement and detection of the eye blink or eyebrow movement facial gestures. The camera is either the Microsoft Kinect for Windows [1] sensor which can provide 3D (RGB-D) data or a Logitech web camera which can only provide 2D (RGB) data. Raw data is extracted in the form of images and depth maps. The efficacy of head tracking and gesture recognition of 3D vision-based system is compared to 2D vision-based systems using a modified Fitts' test.

3.1 Device Evaluation

Fitts' Test. Fitts originally proposed a method to model the human hand movement in order to improve human-machine interactions [13]. Each task has an *ID* which is based on the size of the target and the distance of the target from the starting point. The *ID* represents the cognitive-motor challenge imposed on the human to accomplish the task and is measured in bits as shown in Eq. (1).

$$ID = \log_2\left(\frac{D}{W} + 1\right) \quad (1)$$

where D represents the distance from the starting point to the target and W is the width of the target.

$$MT = a + b \times ID \quad (2)$$

The relationship between MT and ID is shown as a linear relationship where a is the y-intercept and b is the gradient of the line represented in Eq. (2). The Index of Performance (*IP*) in bits/second of a device is given in Eq. (3).

$$IP = \frac{1}{b} \quad (3)$$

where b is the gradient of the line described in Eq. (3). A positive value of *IP* indicates that the device gets more difficult to use as the interaction becomes

more challenging. Equation (4) is used to calculate the Effective Throughput (TP_e) in bits/second.

$$TP_e = \frac{ID_e}{MT} \tag{4}$$

where MT is the mean movement time, in seconds, for all trials within the same condition. It represents the overall efficiency of the device in facilitating interactions.

$$ID_e = \log_2\left(\frac{D}{W_e} + 1\right) \tag{5}$$

ID_e , is the effective index of difficulty, in bits, and is calculated from the distance (D) from the start location to the target and W_e , the effective width of the target. W_e , is the effective width of the target and it is calculated from the observed distribution of the target selection coordinates.

$$W_e = 4.133 \times SD \tag{6}$$

where SD is the standard deviation of the selection coordinates [12].

The experiments showed that the rate of performance of the human motor system is approximately constant over a wide range of movement amplitudes. Fitts' Law [14, 23] states that MT should increase with an increase in the ID i.e. as the difficulty of the task increases, the time taken to complete the task also increases. Fitts' Law was adapted in Mackenzie et al. [23], to assess HCI devices. Therefore, it was thought Fitts' test is an appropriate tool for assessing the performance of the head tracking and gesture recognition system.

3.2 Gesture Detection

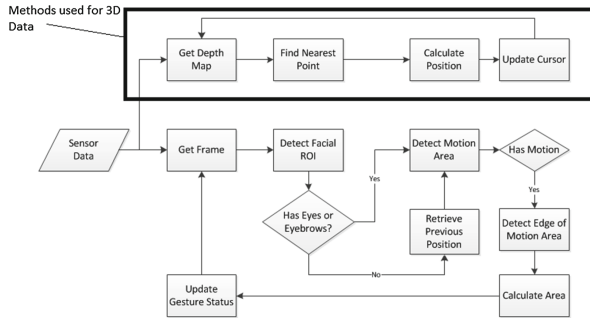


Fig. 1. Algorithm to detect blink and eyebrows movements.

Figure 1 shows an overview of the 3D head tracking and facial gesture recognition system. The facial gesture recognition system is the same for both the 2D vision system and the 3D Kinect system. Depth data is used only to filter the region

of interest when processing the facial image - only objects within a meter of the 3D sensor were included in the region of interest and all other background is removed before further processing.

The facial gesture recognition system used the RGB data from the sensors. Facial areas of interest such as the head, eyes region, left eye and right eye are detected using a Haar-Cascade [32]. To detect a blink, closure of both eyes has to be detected for a period of 1 s or more and then return to the open state. If closure of only one eye is detected, the system assumes there is no blink. Only the transition from open eye to close eye and to open eye again is recognised as a blink.

In the case of the eyebrows detection, the two states of the eyebrows (raised, down) are monitored. In the eyebrows raised state the facial eyebrows muscles are contracted in order to raise the eyebrows and the down state, the muscles are relaxed and the eyebrows revert to their original location. The eyebrows region is detected using the location above the eye region. The state of the eyebrows is initially set to down. To recognise eyebrows movement both eyebrows have to be raised for a period of 1 s or more and subsequently return to the down state. Only the transition from down to raised and then to the down state again will be recognised as a valid eyebrows movement.

4 Experimentation

4.1 Setup

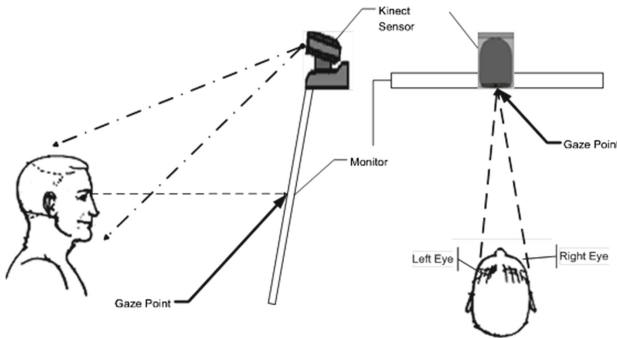


Fig. 2. Experimental set-up.

The participant was asked to perform a series of Fitts' Tests [14,24]. The participants were allowed to repeat the gestures until the click action was detected and thus this caused the movement time to increase. The Fitts' Test was used to evaluate two devices: a 2D vision based head tracker using the Logitech web camera and a 3D head tracking system using the Kinect device. The experiment

was performed using the two facial gestures (blink and eyebrows movement) as a switching mechanism. It has been reported that spontaneous eye blink can change from 20 to 30 blinks/min depending on the mental task the person is performing [19], and can decrease to about 11 blinks per minute during visually demanding tasks [35]. Therefore, the intentional blink time threshold was set to 1000 ms to distinguish between intentional and unintentional facial gestures and to prevent spontaneous blinks from being detected. The activation time of the eyebrows movement switch was also set to 1000 ms (Fig. 2).

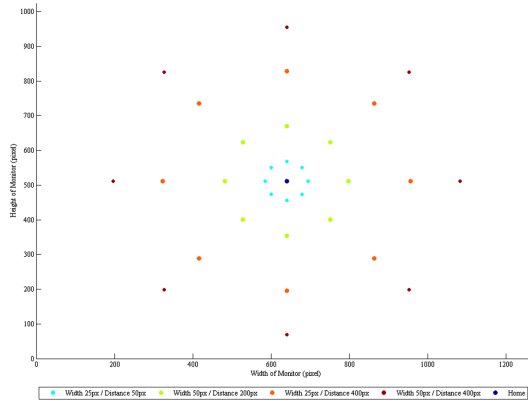


Fig. 3. Target locations (incorporating 8 distinct movement orientations).

The screen used is a 17 in. LCD monitor with a resolution of 1280 by 1024 pixels. The target is selected at random from a set of pre-designated locations as shown in Fig. 3 and presented to the participant. The participant then has to move the cursor using head movement and select the target with the equivalent of a mouse click using the different facial gestures being evaluated. Once a target has been chosen, the participant has to move the cursor back to and select the target at the central location on the screen. This ensures that the same start point is used for each target selection. The choice of the stimulus target locations are based on earlier work by Guess et al. where the points were configured to perform a range of selection tasks with 8 target directions/orientations [16].

4.2 Sensors

Two sensors were used. The first was a standard Logitech web camera. The web camera captured 640×480 pixel RGB images at a rate of 30 frames per second. The second sensor was the Kinect for Windows sensor [1]. The Kinect sensor consists of a structured light based depth sensor and an RGB sensor. The Kinect sensor operates at a 30 Hz rate and generates 640×480 depth and RGB images. The depth range of the Kinect sensor in default mode is 800 mm to 4000 mm and in near mode is 500 mm to 3000 mm. In this experiment the

Kinect sensor operated in near mode. Both the web camera and the Kinect sensor were selected because they are relatively inexpensive devices that can be readily obtained.

4.3 Depth Data

The depth data obtained from the Kinect sensor is used to reduce the search area for the different Haar-Cascade features. This will reduce the computational load and will avoid background distractions, such as people, movements and changes in lighting and therefore increase the performance. A mask is created from the depth data and the object within 1000 mm of the sensor is selected. The mask is used on the colour image to remove all the objects which are more than 1000 mm from the sensor.

5 Result

The experiment was carried out with 21 healthy individuals who completed the tests with all 4 devices. The *MT* in Fitts' Test is the time taken to move to the target location from the starting point and performing the task. To be able to compare the devices and the effect of the facial gesture, we have broken the task in two. Task 1 involves moving the cursor to the target location using the movement of the head. Task 2 encapsulates Task 1 and also involves selecting the target by using one of the facial gestures as a switching mechanism.

In Figure 4, the Kinect-eyebrows has a lower *MT* that the Kinect-blink system for an ID greater than 1.9 bits. Overall for Task 1, it can be seen that the Kinect-eyebrows system has the lowest *MT*, followed by the Kinect-blink, the webcam-blink and finally the webcam-eyebrows, which took the most time to complete (Fig. 5).

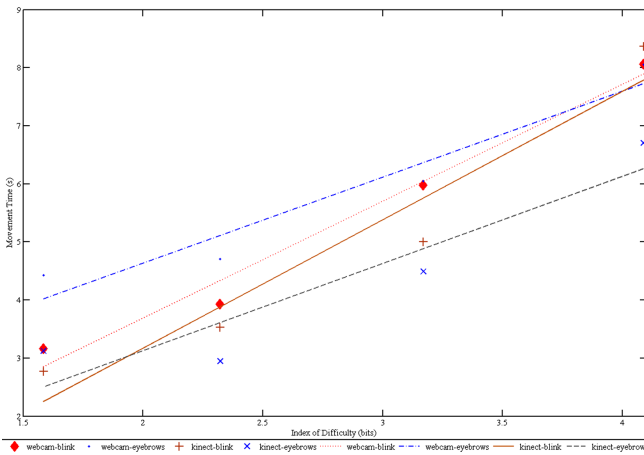


Fig. 4. Fitts' test result for Task 1 (movement to target).

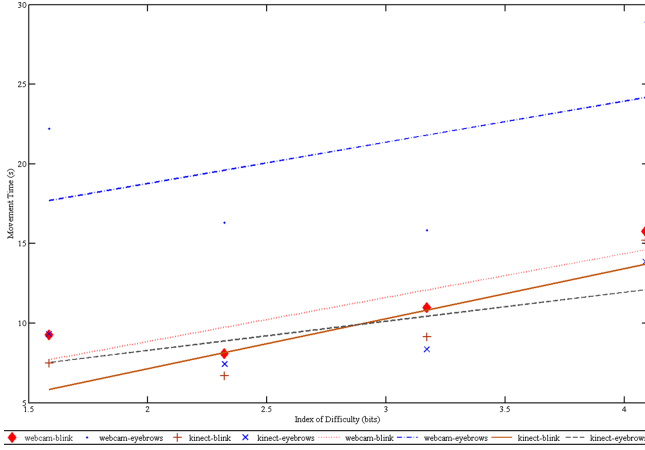


Fig. 5. Fitts’ test result for Task 2 (performing the facial gesture).

Table 1. Overall index of performance (IP) and effective throughput (TP_e) of tested devices

Device	Task 1		Task 2	
	IP	TP_e	IP	TP_e
Webcam-Blink	0.36	0.74	0.32	0.41
	$(R^2 = 0.98)$		$(R^2 = 0.78)$	
Webcam-Eyebrows	0.39	0.67	0.55	0.37
	$(R^2 = 0.90)$		$(R^2 = 0.21)$	
Kinect-Blink	0.5	0.89	0.45	0.6
	$(R^2 = 0.89)$		$(R^2 = 0.78)$	
Kinect-Eyebrows	0.68	0.95	0.67	0.64
	$(R^2 = 0.81)$		$(R^2 = 0.48)$	

From Table 1, it can also be seen that both the IP and the TP_e for moving the cursor to the designated target (Task 1) were better than that of the combination of moving and the click action (Task 2) using the different facial gestures for all devices. This is to be expected as the clicking/selection method has an effect on the performance and efficiency of the system used. Also, both the IP and TP_e of the 3D Kinect system were better than those of the 2D Vision system. R^2 is the coefficient of determination and measured as a percentage of how well the data fits the linear model [10]. If we look at the R^2 values Task 1 are higher than those of Task 2, this would indicate that Task 1 follows the linear model more closely than Task 2. Also, another interesting observation is the fact that using the Blink gesture with both the web camera and the Kinect yield similar R^2 values whereas the R^2 values of the Eyebrows movement gesture are lower.

In Table 2, the IP for different devices are presented when performing Task 1 and a combination of Task 1 followed by Task 2 for different target orientations.

Table 2. *IP* of Task 1 and Task 2 in bits/second

Orientation	<i>IP</i> of Task 1 (bits/second)								<i>IP</i> of Task 2 (bits/second)							
	0	45	90	135	180	225	270	315	0	45	90	135	180	225	270	315
Webcam-Blink	0.51	0.56	0.42	0.52	0.48	0.4	0.49	0.56	0.63	0.91	0.44	0.12	0.28	0.38	0.7	0.47
Webcam-Eyebrows	0.94	0.29	1.43	0.67	0.84	1.57	0.83	0.57	0.34	-0.35	-0.73	3.74	0.1	0.13	0.22	1.07
Kinect-Blink	0.27	0.57	0.49	0.5	0.3	0.35	0.68	0.77	0.29	0.88	0.21	0.37	0.12	0.3	0.37	0.52
Kinect-Eyebrows	0.56	0.51	1.26	0.35	0.53	1.23	1.94	0.65	0.31	1.54	0.43	0.22	0.21	-5.41	2.82	1.1

Table 3. *TP_e* of Task 1 and Task 2 in bits/second

Orientation	<i>TP_e</i> of Task 1 (bits/second)								<i>TP_e</i> of Task 2 (bits/second)							
	0	45	90	135	180	225	270	315	0	45	90	135	180	225	270	315
Webcam-Blink	0.67	0.88	0.67	0.8	0.81	0.59	0.88	0.69	0.41	0.45	0.39	0.34	0.47	0.33	0.44	0.41
Webcam-EyebrowsEyebrows	0.96	0.62	0.61	0.71	0.62	0.64	0.62	0.62	0.68	0.34	0.32	0.35	0.35	0.3	0.35	0.36
Kinect-Blink	0.82	0.8	0.91	0.88	0.73	0.83	1.04	0.98	0.45	0.54	0.63	0.64	0.51	0.62	0.71	0.64
Kinect-Eyebrows	0.85	0.99	1.12	0.78	0.79	0.9	1.39	0.92	0.54	0.55	0.68	0.44	0.54	0.63	1.06	0.67

A one-way ANOVA test was performed on the *TP_e* for the different orientations and gestures of both Task 1 and Task 2. For the comparison by orientations, $p < 0.01$ ($p = 0.001$ and $p = 0.007$) for Task 1 and Task 2, it can be said that there is a significant difference between the mean of the different orientations i.e. the *TP_e* are different based on the orientation of the movement. For the comparison by gesture, only Task 2 had $p < 0.01$ ($p = 0.0093$). This indicates that there is a significant difference between the mean of the *TP_e* based on the gesture being performed. This would point out that there is a difference in the performance of the two facial gestures being investigated. The mean *TP_e* of Task 2 is greater due to the increased challenge of both moving and selecting/clicking. Also, there are no sufficient evidence of any difference between the means of *TP_e* based on the orientation and gesture for either Task 1 or Task 2. This would indicate the gesture recognition for the sample used might be invariant to the orientation of the task being performed.

6 Discussion

Using facial gestures as a switch is possible in real time but the use of such gestures may cause a drop in the overall *IP* of the systems. *IP* and *TP_e* values in Table 1 using the four different systems were obtained with participants successfully reaching and selecting all targets. As it can be seen in results for the overall *IP* (Table 1), the R^2 value which represents the goodness of fit of the

fitted line for the Kinect 3D system is greater than 0.7 i.e. the line accounts for more than 70% of the variance. In contrast, the Webcam-Eyebrows device R^2 is 0.21, and thus accounts for only 21% of the variance. This could also indicate that the presence of outliers has a large influence on the fitted line and thus the gradient. As the IP calculation from Eq. (3) is based on the inverse of the slope, it is also being influenced by outliers at very low and very high indices of difficulty. It should be born in mind that each of the points in Figs. 4 and 5 are obtained from the mean of data obtained from 21 users and 8 directions giving 64 data points. In the presence of such outliers relying on TP_e as a measure of performance might be better.

It can be seen that there is a decrease in the TP_e of all the four different devices after the switching action is included. The reduction in the TP_e of the 2D Vision system is 45% and 44% for the blink and eyebrows devices respectively. Similarly, the decline in the TP_e of the 3D Kinect system is 32% and 35% for the blink and eyebrows devices respectively. The higher total TP_e value indicates that the Kinect system, utilizing 3D information, has resulted in better performance when the two tasks of moving and selecting are combined and thus improved the ease of use of the system as a whole. It has also been shown that the TP_e for Task 2 based on gesture are from different populations - with eyebrows having a higher mean TP_e . There is no evidence to support a difference in performance based on sensor or device. This also supports the impact to the improved performance of the gesture detection algorithm.

In addition, the facial gesture detection rate affected the MT for the different devices. In this implementation of the Fitts' test, the tasks were considered completed only when the switch was activated and click action performed.

7 Conclusion

Both Kinect systems have lower MT and higher IP and TP_e than the Webcam based systems thus showing that the introduction of the depth data had a positive impact on the head tracking algorithm. This could be explained by the ability to throw away unnecessary data at an early stage in processing using depth information and thereby speeding up subsequent stages to create a more smooth experience for the users. In this work, we have looked at only blink and eyebrows movement gestures, further work will have to be carried out on additional gestures such as mouth opening/closing and tongue movement. We now intend to conduct translational research with neurological patients.

References

1. Kinect for Windows. <http://www.microsoft.com/en-us/kinectforwindows/>
2. SmartNav. <http://www.naturalpoint.com/smartnav/>
3. Tobii Technology. <http://www.tobii.com/>

4. Abascal, J., Nicolle, C.: Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interact. Comput.* **17**(5), 484–505 (2005). <https://doi.org/10.1016/j.intcom.2005.03.002>. <http://iwc.oxfordjournals.org/cgi/doi/10.1016/j.intcom.2005.03.002>
5. Accot, J., Zhai, S.: Beyond Fitts' law: models for trajectory-based HCI tasks. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 295–301 (1997). <https://doi.org/10.1145/258549.258760>. <http://dl.acm.org/citation.cfm?id=258760>
6. Betke, M., Gips, J., Fleming, P.: The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Trans. Neural Syst. Rehabil. Eng.* **10**(1), 1–10 (2002). A Publication of the IEEE Engineering in Medicine and Biology Society
7. Betke, M.: Intelligent interfaces to empower people with disabilities. In: Nakashima, H., Aghajan, H., Augusto, J.C. (eds.) *Handbook of Ambient Intelligence and Smart Environments*, pp. 409–432. Springer, Boston (2010). https://doi.org/10.1007/978-0-387-93808-0_15
8. Bradski, G.: The OpenCV library. *Dr. Dobb's J. Softw. Tools* **25**, 120–125 (2000)
9. Cloud, R., Betke, M., Gips, J.: Experiments with a camera-based human-computer interface system. In: *Proceedings of the 7th ERCIM Workshop "User Interfaces for All," UI4ALL 2002*, pp. 103–110 (2002). <http://cstest.bc.edu/~gips/UI4ALL-2002.pdfwww.cs.bu.edu/faculty/betke/papers/Cloud-Betke-Gips-UI4ALL-2002.pdf>
10. Cox, D.R., Snell, E.J.: *Analysis of Binary Data*, vol. 32. CRC Press, Boca Raton (1989)
11. del Val, L., Jiménez, M.I., Alonso, A., de la Rosa, R., Izquierdo, A., Carrera, A.: Assistance system for disabled people: a robot controlled by blinking and wireless link. In: Lytras, M.D., Ordonez De Pablos, P., Ziderman, A., Roulstone, A., Maurer, H., Imber, J.B. (eds.) *WSKS 2010. CCIS*, vol. 111, pp. 383–388. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16318-0_45
12. Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing pointing device performance and user assessment with the ISO 9241, part 9 standard. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1999*, pp. 215–222. ACM, New York (1999). <https://doi.org/10.1145/302979.303042>
13. Fitts, P.M., Radford, B.: Information capacity of discrete motor responses under different cognitive sets. *J. Exp. Psychol.* **71**, 475–482 (1966)
14. Fitts, P.: The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* **47**(6) (1954). <http://psycnet.apa.org/journals/xge/47/6/381/>
15. Grauman, K., Betke, M., Lombardi, J., Gips, J., Bradski, G.: Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. *Univ. Access Inf. Soc.* **2**(4), 359–373 (2003). <https://doi.org/10.1007/s10209-003-0062-x>
16. Guiness, S.P., Deravi, F., Sirlantzis, K., Pepper, M.G., Sakel, M.: Evaluation of vision-based head-trackers for assistive devices. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 4804–4807 (2012). <https://doi.org/10.1109/EMBC.2012.6347068>. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6347068
17. Gustavsson, A., et al.: Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol. J. Eur. Coll. Neuropsychopharmacol.* **21**(10), 718–79 (2011). <https://doi.org/10.1016/j.euroneuro.2011.08.008>. <http://www.ncbi.nlm.nih.gov/pubmed/21924589>

18. ISO TC 159/SC 4: ISO 9241-9:2000, ergonomic requirements for office work with visual display terminals (VDTs) - part 9: requirements for non-keyboard input devices. International Organization for Standardization (2002)
19. Karson, C.N.: Spontaneous eye-blink rates and dopaminergic systems. *Brain: J. Neurol.* **106**(Pt 3), 643-53 (1983). <http://www.ncbi.nlm.nih.gov/pubmed/6640274>
20. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. *Int. J. Comput. Vis.* **1**(4), 321-331 (1988). <https://doi.org/10.1007/BF00133570>
21. King, D.E.: Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res. (JMLR)* **10**, 1755-1758 (2009). <http://dl.acm.org/citation.cfm?id=1577069.1755843>
22. Krolak, A., Strumillo, P.: Vision-based eye blink monitoring system for human-computer interfacing. In: 2008 Conference on Human System Interactions. Institute of Electrical & Electronics Engineers (IEEE), May 2008. <https://doi.org/10.1109/hsi.2008.4581580>
23. MacKenzie, I.S.: Fitts' law as a research and design tool in human-computer interaction. *Hum. -Comput. Interact.* **7**(1), 91-139 (1992). https://doi.org/10.1207/s15327051hci0701_3
24. MacKenzie, I.S., Buxton, W.: A tool for the rapid evaluation of input devices using Fitts' law models. *ACM SIGCHI Bull.* **25**(3), 58-63 (1993). <https://doi.org/10.1145/155786.155801>. <http://portal.acm.org/citation.cfm?doid=155786.155801>
25. Malik, K., Smolka, B.: Eye blink detection using local binary patterns. In: 2014 International Conference on Multimedia Computing and Systems (ICMCS), pp. 385-390, April 2014. <https://doi.org/10.1109/ICMCS.2014.6911268>
26. Missimer, E., Betke, M.: Blink and wink detection for mouse pointer control. *ACM Press, New York* (2010). <https://doi.org/10.1145/1839294.1839322>. <http://portal.acm.org/citation.cfm?doid=1839294.1839322>
27. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recogn.* **29**(1), 51-59 (1996). [https://doi.org/10.1016/0031-3203\(95\)00067-4](https://doi.org/10.1016/0031-3203(95)00067-4)
28. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcam. In: 2007 IEEE 11th International Conference on Computer Vision. Institute of Electrical & Electronics Engineers (IEEE) (2007). <https://doi.org/10.1109/iccv.2007.4409068>
29. Pistori, H.: Computer vision and digital inclusion of persons with special needs: overview and state of art. In: *Computational Modelling of Objects Represented in Images* (2018). <https://www.taylorfrancis.com/books/e/9781351377133/chapters/10.1201%2F97813515106465-6>
30. Royal College of Physicians: Medical rehabilitation in 2011 and beyond. Report of a working party, November 2010, Royal College of Physicians & British Society of Rehabilitation Medicine. RCP, London (2010)
31. Tuisku, O., Surakka, V., Vanhala, T., Rantanen, V., Lekkala, J.: Wireless face interface: using voluntary gaze direction and facial muscle activations for human-computer interaction. *Interact. Comput.* **24**(1), 1-9 (2012). <https://doi.org/10.1016/j.intcom.2011.10.002>. <http://iwc.oxfordjournals.org/cgi/doi/10.1016/j.intcom.2011.10.002>
32. Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vis.* **57**(2), 137-154 (2004). <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>
33. Viola, P., Jones, M.: Robust real-time object detection. *Int. J. Comput. Vis.* **57**, 137-154 (2001). <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>

34. Ward, D.J., Blackwell, A.F., MacKay, D.J.C.: Dasher—a data entry interface using continuous gestures and language models, pp. 129–137. ACM Press (2000). <https://doi.org/10.1145/354401.354427>. <http://portal.acm.org/citation.cfm?doid=354401.354427>
35. Wilson, G.F.: An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *Int. J. Aviat. Psychol.* **12**(1), 3–18 (2002)
36. World Health Organization: World report on disability. Technical report, WHO, Geneva (2011). http://whqlibdoc.who.int/publications/2011/9789240685215_eng.pdf
37. Yunqi, L., Meiling, Y., Xiaobing, S., Xiuxia, L., Jiangfan, O.: Recognition of eye states in real time video, pp. 554–559, January 2009. <https://doi.org/10.1109/ICCET.2009.105>. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4769528>
38. Zhang, X., Kulkarni, H., Morris, M.R.: Smartphone-based gaze gesture communication for people with motor disabilities. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017, pp. 2878–2889. ACM, New York (2017). <https://doi.org/10.1145/3025453.3025790>