

An Adaptability of Head Motion as Computer Input Device

Takehiko Tomikawa, Toshiaki Yamanouchi, and Hiromitsu Nishimura
Dept. of Information Media, Kanagawa Institute of Technology, Atsugi, Japan
Email: {tomikawa, yama, nisimura}@ic.kanagawa-it.ac.jp

Abstract—This paper describes an application scheme for human interface by utilizing the movements of body parts as an input device. The purpose of this paper is to assist the computer input for the person with hand disabilities, and to construct a system that can be inexpensive and easily implemented. Thus, the authors propose a combination parameters of “Euler angles” and “Translations” under body movements to perform mouse scanning behaved as alternative cursor. In other words, this is a trial to replace the pointing functionality of the mouse by utilizing “Translations” in neck or waist movements in addition to the “Pitch/Yaw/Roll” in face orientations. There are similar ways of thinking in the past, however, the usage of parameter combination as well as the possibility of practical realization can be hardly found. As a result of our experiments, it is to give an indication that our method can be applicable to function as a mouse scanning to some extent in spite of the simple system configurations by utilizing the current technique in both hardware and software. On the other hand, there are some problems remained as further considerations, such as, operability experiments by handicapped subjects, the system configurations in wireless linkage, and so on.

Index Terms—human interface, mouse substitution, Euler angles, Kinect sensor

I. INTRODUCTION

In the field of mobile phone or personal computer, so called “touch sensor” input on the display screen is widely used. This method of scanning on the display device is significant as an input unit for healthy persons of the finger. The input system based on speech requires a huge dictionary registration, it is effective only in situations where the user may aloud. The former has a role as a human interface of contact type, the latter as one of non-contact type. On the other hand, there is a limit to the means of sending information for handicapped persons of hand disabilities. Conventionally, there has been a concept of applying facial movements to the alternative mouse [1]. However, there is a problem in the robustness of the recognition due to the matching errors regarding stereo vision in three-dimensional processing [2]. Recently, there is an inexpensive product to the alternative mouse by detecting the line of sight

without any attachment to the human head [3]. The problems are still remained for the eye movements in visual fixation, narrow recognizable view range, etc. Our scheme seems to be common with the conventional method in terms of facial behavior, non-contact, or non-wearing, however, it becomes rather simple concepts by extracting Euler angles and Translations in head movements. By taking into account of the tendency as a background, the authors have focused on how to support handicapped person to perform computer input, and how to apply head movements to an alternative mouse without using hands. The following is the purpose and the proposal of this authors.

Purpose: To assist computer input for disabilities in hand and to build the system being easily implemented without cost.

Propose: To replace mouse functions by “Euler angles” and “Translations” as a hybrid scanning.

Here, Fig. 1 shows Euler angles in (a) and Translations in (b), as facial movements, used in this proposal.

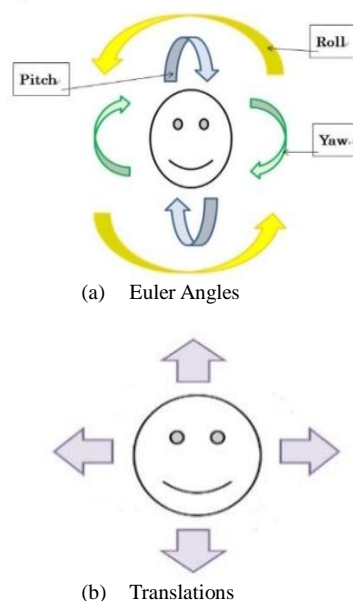


Figure 1. Euler angles and translations (a), (b)

II. PREPARATIONS

Let an arbitrary point of the head be (x, y) , the movements can be expressed as a following matrix (1).

where, the rotation angle (θ) and the scale factor (δ), and Translation (t_x, t_y). Positioning of the conversion formula in this paper means the similarity transformation based on Euler angles (Pitch, Yaw, Roll) and the depth of camera.

$$\begin{matrix} \text{rotation} & \text{scale} & \text{translation} \\ \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} & \delta \begin{bmatrix} x \\ y \end{bmatrix} & \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \end{matrix} \quad (1)$$

Then, the recognition process proceeds from head joint \rightarrow face region \rightarrow face parts (eyes, mouth), and leads to the “normal vector” on the face plane. Here, the normal vector N of the rectangle $ABCD$ can be expressed in (2) by line segments AB, AD , where, “ \times ” means cross product of the vector.

$$N = \overrightarrow{AB} \times \overrightarrow{AD} \quad (2)$$

This paper is the technique of Rotations and Translations in three dimensional head movements as a function of mouse substitution. Let the following assumptions be set as a preparation in our scheme.

- It can be done for head rotations with fixed line of sight. (Euler angles)
- It can be done for eye rotations without head movements. (contribution to the expansion of field of view)
- It can be done for visible region checking without eye and head movements. (visual field test)
- It can be done for vertical and horizontal movements without face rotations. (Translations)

For a), The Euler angles in this paper are either positive or negative value, in three dimensional rotation angles of the head or face, that does not depend on the depth of camera. Here, the cursor movements are applied as a displacement values, and proportional to the Euler angles. And, the proportional constants can be determined in consideration of {field of view, scanning range} on the screen. Where, the scanning speed means the displacement value for angular velocity of the rotating face to the cursor movements.

For b), there is a report of alternative cursor using the line of eye as a non-contact type noted in the previous paragraph. This is an advantage to utilize eye movements without head rotations, however the problems are still remained in operation stability and operability.

For c), it is performed in the “visual field examination”, meaning the range that can be recognized the presence or absence of blinking light. The field of view is the visible range of display, and does not mean scanning range of cursor. That is, in the conversion from angle to displacement, visible field on the screen becomes roughly $2\pi \times \text{depth} \times (\text{Euler Angle}) / 2\pi$ in relation to both Euler angle and depth.

For d), there are face Translations in positive or negative values that does not depend on the camera depth. For applying to an alternative cursor, the displacement due to neck and waist can also be considered in this paper.

Now, by using an image data obtained through the camera, the system proceeds {human recognizing \rightarrow body

recognizing \rightarrow head recognizing \rightarrow face recognizing}, and resulted in alternative mouse function. Here, they are recognition procedures of human body within { }, which have to run as algorithms that can be operated in stably but in real time. Further, by specifying the portion of the face { eyes, nose, and mouth }, it leads to find out the normal direction figured out in geometry.

As a whole, the authors have reached that the amount of displacement in considering with operability as a mouse substitution is followed by Euler angles and Translations. Thus, we have decided to apply our proposed scheme to alternative mouse system by utilizing appropriate existing algorithms of recognition in head or face parts. After all, it was experimentally verified that the present proposal can be realized with simple system configurations based on existing hardware and software tools. Prerequisite as a human interface in this experiments are as follows.

- In lying mode (for disabled person in hand)
- Without any attachment to the body (non-contact operation)
- Using facial, neck, or waist action (upper body as a recognition target)

The Fig. 2 illustrates the positional relationship among face, camera, and screen. They are, the case of sitting on the chair being applied for upper body and lying in bed. Assuming that the display resolution for screen projection is 480×640 [pixels] and 70 [cm] for the distance between camera and the subject. Now, they become necessary conditions as an alternative cursor in order to move within this operating range for display contents in the field of view.

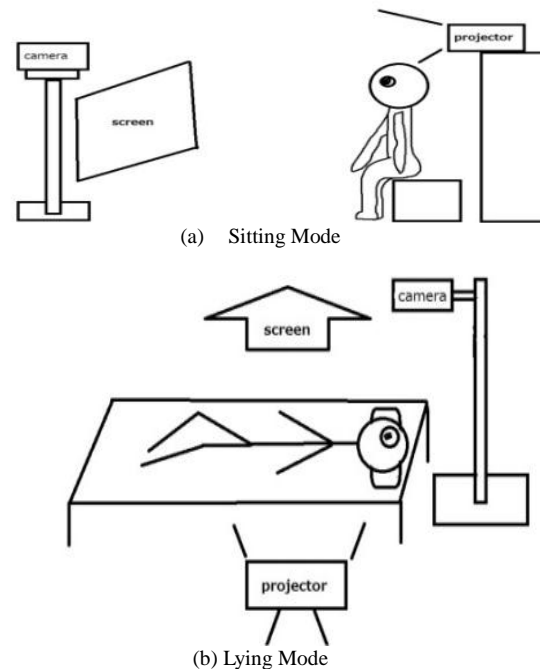


Figure 2. Experimental layout (a), (b)

III. ALTERNATIVE CURSOR

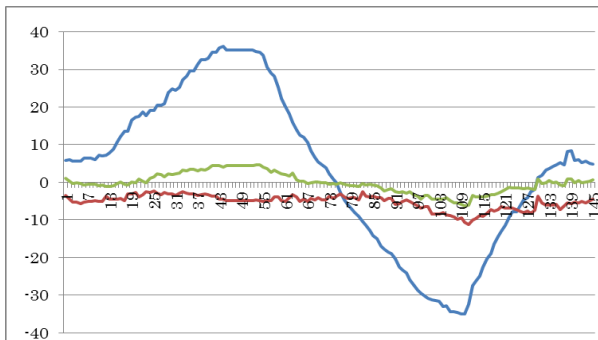
In view of assisting handicapped persons with disabilities in hand, it is conceivable use of {face, head,

or foot} as a potential. Here, we have tried to utilize only head portion as an alternative mouse in consideration with lying in bed or sitting on the wheelchair. Further, it was also performed in the case of using Translation move besides Euler angles. As a preparation, initial system parameters must be set by attempting in motional operability. Then, some evaluations by the number of subjects must be obtained through experimental trial. At this time, we have tried to verify whether the pointing operation from the face rotations is reasonable or not while scanning on the display screen with cursor moving.

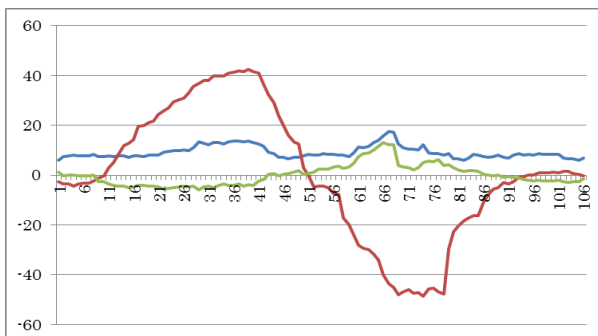
Now, typical Euler angles Pitch / Yaw / Roll are shown in Tbl. I. and Fig. 3. (a) - (c) as actual measurement examples by this authors. Where, the angles in table and figures are [degree] expression with signed values from -90 to 90 in a normal front vision as a reference, and vertical and horizontal axis indicate angles and time consumptions, respectively. Throughout the experiments, it can be seen that the rotation angles are narrowed in the order of Yaw \rightarrow Pitch \rightarrow Roll. It is considered due to the joint structure of human neck. It seems to be rather helpful tendency to summarize at the primary stage although the dynamic range of each rotation angles may not be generalized among individual differences. About Pitch and Yaw in Fig. 3 (a) (b), both of them behave in similar transition levels, while Yaw shows larger swing in operating range. In Fig. 3 (c), we can see that the Yaw and Pitch are affected by Roll operation.

TABLE I. MEAN VALUES OF PITCH/YAW/ROLL

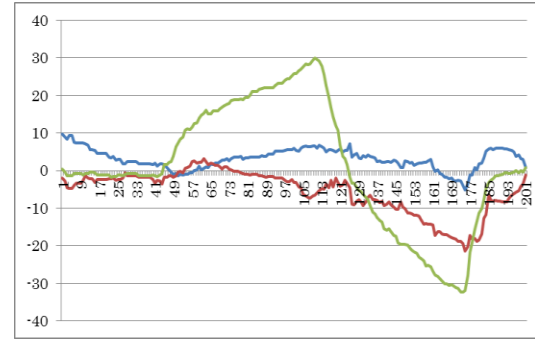
Pitch	-35	to	+37	(down to up)
Yaw	-44	to	+42	(right to left)
Roll	-32	to	+30	(right to left)



(a) Pitch



(b) Yaw



(c) Roll

Figure 3. Euler Angles (a) - (c)

In view of head rotations, it seems to be natural to utilize Pitch and Yaw as vertical and horizontal scanning, respectively. Assuming that the 6×6 matrix size = 18×18 [cm] and the distance between camera and the subject = 70[cm], (Pitch, Yaw) \cong (15°, 15°) is the field of view as a normal vision. It follows that it is sufficient ranging comparing with the measured value of Pitch and Yaw in Table I. Thus, it is necessary to adjust an angular velocity in accordance with the proportional constant for the range of movement in cursor scan, in the field of view with recognizable contents. In this experiment, we let the amount of cursor movement [pixel] on the screen be proportional to the Euler angles, and the proportional constants can be necessary in view of operation in human interface. Regarding field of view, let Pitch α and Yaw β , the distance between subject and camera be unchanged, the following relationship can be obtained in (3) and (4).

$$\text{Vertical displacement} \propto \alpha \quad (3)$$

$$\text{Horizontal displacement} \propto \beta \quad (4)$$

Now, in order to perform click action of the mouse, there are several ways existed, such as {foot action, voice action, unmoved in certain period, with blinking, open/close eye action, up/down eyebrow, and so on }. In this time, both Roll operation and time holding of cursor can be experimentally used. By taking analog value of the Roll angle γ to digital on/off function, the head inclination of left/right corresponds to left/right clicking in mouse operation. That is, left click and right click are followed by $\gamma < -T_h$ and $\gamma > +T_h$, respectively, where the value of threshold $T_h=8$ in this time. In addition, the click is set to be one-time action while occurring for Roll transaction.

IV. SCANNING EXPERIMENTS

Here, we had scanning experiments by focusing on the operation of cursor movements without regarding to input characters or symbols in keyboard function.

The various face recognition algorithms based on learning process have been proposed in the past (for example [4]). In fact, the face recognition in recent digital cameras can be automatically included. Therefore, it must be appropriate for our purpose to utilize an optimal and available system into our recognition system. Currently, the software tools for specifying the body

joints through the depth information, so called “Kinect Sensor” has been introduced [5]-[7]. In our system, it can be utilized with the product of “Face Tracking” as a software tool which specifying the joints of the body, and Euler angles [8]. Further, the face recognizing programs have been available as useful image processing [9]. The authors have decided to utilize both library tools in view of extracting Euler angles and Translations in face movements. Thus, it can lead to recognize a portion of face from geometrical locations following joint extraction from depth image. Then, it is expected to obtain the quantitative data for rotation angles related to, so called, normal direction of the face.

An experimental hardware system includes { Kinect for Windows, note type computer }, and a software system includes { human head detection / face tracking / face parts recognition or orientation, and mouse substitution }. The flow of hardware system and flow of software system are shown in Fig. 4. (a) and (b).

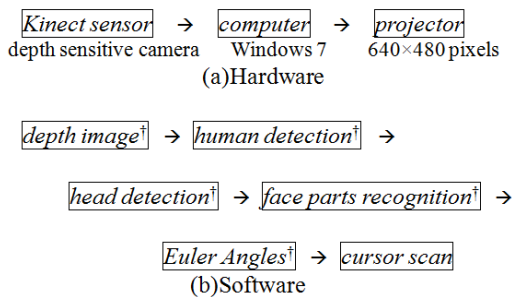


Figure 4. System flow (a) , (b)

First, we have tried an experiment to determine the operation range on the display screen in accordance with the amount of rotation angles in head action as a parameter in a polar coordinate system. Whereas, all attempts are made in the sitting mode on the chair as an upper body for camera subjects. Specifically, it was examined the scanning stability in head rotations in either 3×3 and 6×6 display matrix shown in Fig. 5. (a) and (b). Here, an each matrix element becomes 103×77 [pixels] at the 6×6 matrix in the frame of 640×480 [pixels] (excluded outer frame), but not for the projected matrix size.

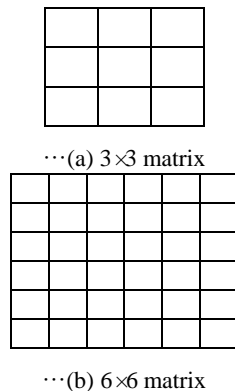


Figure 5. Scanning matrix

In order to get the differences between rotation angles and Translations, we have tried to chase indicated matrix

cell generated by randomized matrix points. In case of 3×3 matrix, the scanning over the matrix was rather easy and smooth in both using Euler angles and Translations. In case of 6×6 matrix, it was similar behavior as 3×3 matrix except some instabilities in scanning action during head rotations. In both cases, the limited actions can be seen in up/down directions by using Translation (t_y), while faster and smooth in left/right (t_x). Also, the cursor tracking with the head scanning was carried out to the corresponding position indicated by random numbers in 6×6 matrix. We have examined the position error rates while chasing indicated matrix points become 28 [%] in Euler angles, 26 to 91 [%] in Translations. Whereas, they are in the mean values of miss-chasing rate with the time interval of 2 [sec] and the number of 180 times iterations by this authors. Some higher errors can be seen in Translations depend on the matrix positions comparing with Euler angles. This is due to the difficulty for up/down shift of 91[%] errors, although the left/right shift of 26 [%] errors in Translations. It is difficult to cover the whole matrix cells by Translations, however it can be helpful to assist Euler angles to some extent. Thus the authors have arranged (3) and (4) to be hybrid type, like (3a) and (4a), where variables indicate to be proportional contributions without units.

$$\text{vertical displacement} \propto (\alpha + ty) \quad (3a)$$

$$\text{horizontal displacement} \propto (\beta + tx) \quad (4a)$$

In each case, the two terms on the right side are accelerated in the same sign, and are decelerated by the opposite sign in the horizontal displacement.

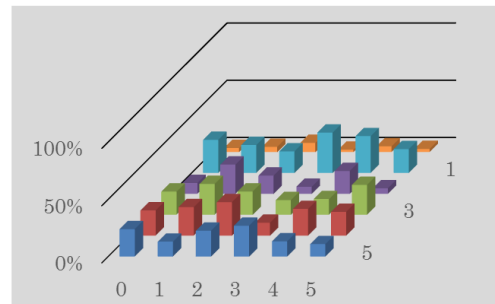


Figure 6. Scanning error rate in 6×6 matrix chasing

We have carried out the experiments in consideration of the arrangements, using random numbers in the same manner as previous trial. The tracking error rates are shown in Fig. 6 as height of each 6×6 matrix point in vertical and horizontal directions, where the average results of 1200 (300×4) trials at the depth of 70 [cm]. The error rates are distributed roughly 0 to 20 [%] in this case. The higher error rates around inside matrix comparing with the side or corner can be seen in the distribution, which is common to all subjects. It is because the chasing direction of the sidewall can be reduced by facing to the end point, not passing point. The lower error rates in the first row or higher error rate in the second row is caused by the case of the same person and seems to be the individual differences. In our experiments, it was lower error rates in case of depth =

70 [cm] comparing with 60 [cm] or 80 [cm] as shown in appendix Fig. (a) and (b). This is related to the compatibility between matrix size and operability or the tracking robustness followed by facial movements. On the other hand, for higher rates in the matrix points shown in this figure seem to be reduced by some appropriate training.

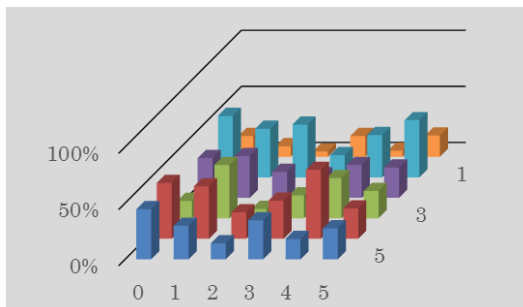
V. CONCLUSION

We have examined the applicability of facial operation by assuming the person of hand disabilities to perform mouse scanning. That is, a trial of alternative cursor based on the Euler angles Pitch / Yaw / Roll and Translations. Then, we have proposed a hybrid parameter scheme: the Pitch and Translation as vertical scanning, Yaw and Translation as horizontal, Roll and time interval as click action.

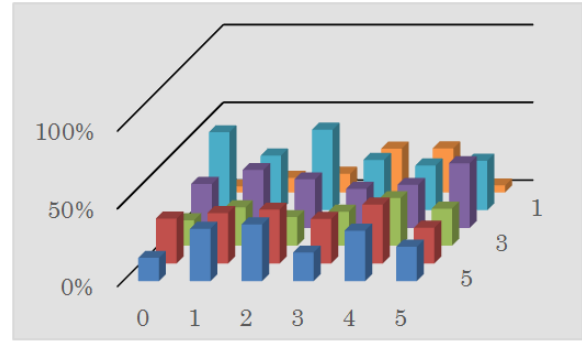
As a result, we can take an advantage of allowing inexpensive system without complexity of the recognition system by utilizing currently available tools. We have learned experimentally that the pointing operation by face action in the resolution of 6×6 matrix can be realized, although it is far behind in the keyboard scan by fingers.

Overall, to quantitative evaluation was rather difficult since it was in the mankind operation as a human interface. However, we had a measure capable of function as an alternative mouse scanning by facial movements to some extent, although operability review by the persons of disability are insufficient. On the other hand, {wired connection of Kinect sensor, web camera utilization, etc.} have been left as a future work.

APPENDIX SCANNING ERROR RATES



(a) Average of 4×300 trials, depth = 60 [cm]



(b) Average of 4×300 trials, depth = 80 [cm]

REFERENCES

- [1] Microsoft Corp. Patent: 004-362569, 2004
- [2] Y. Matsumoto. (2005). [Online]. Available: <http://kaken.nii.ac.jp/d/p/12750218.en.html>
- [3] Eye Tribe. (2014). [Online]. Available: http://www.tobiiatj.com/jpn/ET_guide.html
- [4] Sing-Tze Bow, "Pattern recognition," Marcel Dekker Inc.
- [5] S. Tsukasa and N. Kaoru, *Kinect for Windows SDK Programming Practice*, Kogaku-sha.
- [6] J. Webb and J. Ashley, *Beginning Kinect Programming with the Microsoft Kinect Sdk*, Apress.
- [7] D. Catuehe, *Programming with the Kinect for Windows: Software Development Kit*, Microsoft Press.
- [8] G. Borenstein, *Making Things See*, O'Reilly Pub.
- [9] D. L. Baggio, et. al., "Mastering open CV with practical computer vision projects," Packet Pub.

Takehiko Tomikawa was born in Japan in 1945. He is a Professor in the Department of Media Information, Kanagawa Institute of Technology. He holds the MS and PhD degrees from Shizuoka University. His current research interests are welfare assisting tools based on motion capture. He is a life time member of IEEE.

Toshiaki Yamanouchi was born in Japan in 1968. He is an assistant professor of Kanagawa Institute of Technology, Japan. He gained BE and ME degrees from Waseda University in 1990 and 1992, respectively. He worked at the University as a research associate for three years, and he moved to Kanagawa Institute of Technology in 1997. His main research field is Digital Image Processing. He is a member of IPSJ.

Hiromitsu Nishimura was born in Japan in 1972. He received his Dr. Eng. from Shin-Shu University at Japan in 2000. He is a Lecturer of the Department of Information Media, Kanagawa Institute of Technology. His current interests are adaptations of image processing, including non-visible lighted visions. He is a member of the IAPR.