Estimation of User Interest from Face Approaches Captured by Webcam

Kumiko Fujisawa[1] and Kenro Aihara[1,2]

[1] The Graduate University for Advanced Studies, Sokendai
[2] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo, Japan
{k_fuji,kenro.aihara}@nii.ac.jp

Abstract. We propose a methodology for estimating a user's interest in documents displayed on a computer screen from his or her physical actions. Some studies show that physical actions captured by a device can be indicators of a user's interest. We introduce the ongoing pilot study's results, which show the possible relationship between a user's face approaching the screen, as captured by a webcam, and their interest in the document on the screen. Our system uses a common user-friendly device. We evaluate our prototype system from the viewpoint of presuming an interest from such a face approach and the practicality of the system, and discuss the future possibilities of our research.

Keywords: Interface design, knowledge acquisition, user interest, motion capture.

1 Introduction

Although keyboards and mice are standard input devices for personal computers, many new devices are coming into regular use. Nintendo Wii's motion-sensitive controller is a popular example of such futuristic input devices. Video capturing devices can also be used as a means of input whereby people can control PC software, games, or other machines by moving their hands or bodies.

Techniques to detect and analyze body (including the face) movements are becoming more accessible. In particular, face tracking technologies are now used in household electronic goods [e.g., 1, 2, 3].

There has been a lot of research on new input devices and on using devices to detect user reactions in the field of human-computer interaction (HCI). These devices and systems tend to be heavy or distracting in some way, and user experiments involving them have had to be conducted under extraordinary conditions. To capture natural and emotional behaviors, more common situations are needed.

Our research focuses on how to capture the users' natural behaviors in response to information displayed on the screen via user-friendly (low impact) devices such as webcam. We are planning to identify user actions reflecting their interests using these devices and put them to practical use in the real learning situations. Therefore, we need light-weight and effective data for estimating the user's interest.

R. Shumaker (Ed.): Virtual and Mixed Reality, LNCS 5622, pp. 51–59, 2009.

We are proposing a methodology to estimate the users' interests by using face tracking data captured by a webcam. We describe our preliminary test results showing the potential effectiveness of such a methodology. This is part of our ongoing research on new interactive systems for supporting users in acquiring knowledge.

2 Previous Work

The topic of using sensors to recognize user actions has attracted the attention of researchers in the field of computer science for a long time, and quite a lot of devices have been developed. We will introduce some of these in the following paragraphs.

2.1 How to Capture User Actions

For a system to be able to capture whole-body movements, users sometimes have to wear sensor devices. For example, Sementile et al. [4] proposed a motion capture system based on marker detection using ARToolkit. The system consists of markers with patterns that act as reference points to collect a user's articulation coordinates. This system was used to generate humanoid avatars with similar movements to those of the user. Another example is the emotion recognition sensor system (EREC) that detects a user's emotional state [5]. It is composed of a sensor globe, a chest belt, and a data collection unit. Eye tracking cameras are often used, and the EMR eye tracker is one example [6]. This system uses a dedicated computer to make a head-mounted camera track the eye movements of the user.

Jacob categorized input devices in terms of the aspect of HCI [7, 8]. He studied the use of the hands, foot position, head position, eye gaze, and voice to manipulate a computer and described the devices that could be manipulated. Picard and Daily [9] summarized the body-based measures of an affect and described the typical sensor devices that can detect the action modalities, such as the facial or posture activities. They introduced video, force sensitive resistors, electromyogram electrodes, microphones, and other electrodes.

2.2 Body Action, Interest, and Capture Devices

Mota and Picard [10] described a system for recognizing the naturally occurring postures and associated affective states related to a child's interest level while he or she performed a learning task on a computer. Their main purpose was to identify the naturally occurring postures in natural learning situations. The device used in this research was composed of pressure sensors set in a chair. They used a hidden Markov model to remove the noise from the data and estimated the relationship between a posture and interest. They found evidence to support a dynamic relationship between postural behavior patterns and the affective states associated with interest.

Wakai et al. [11] proposed a technique for quantitatively measuring the change in interest from a portrait measured with a camera. They focused on the pose of eye gaze and the distance to the object. Their system detected eye gazes and the posture from the video image, and they succeeded in detecting changes in the direction of interest

when the experiment's participants were asked to choose one favorite advertisement from a total of three. They used three cameras set behind each of the advertisements to detect the approach posture from the silhouette.

3 Proposed System

We propose a system that uses a familiar webcam to detect a user's face approaching the screen. We intend to use this system to select the natural movements which relate to the interest from a large amount of real-life samples. Figure 1 shows the basic structure of our prototype system.

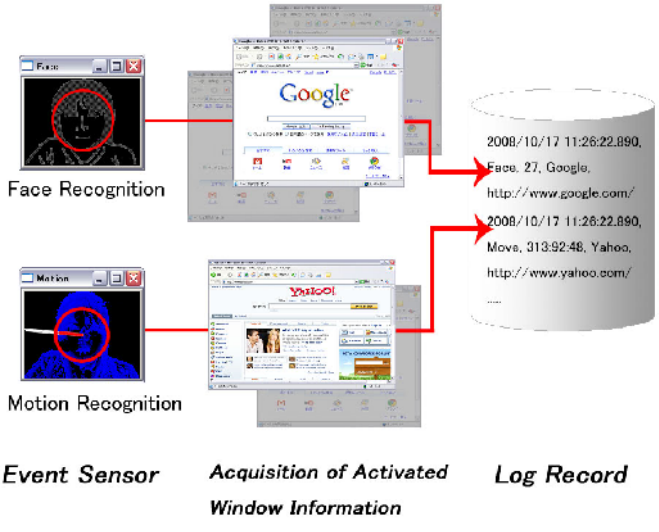

Fig. 1. Basic structure of prototype system

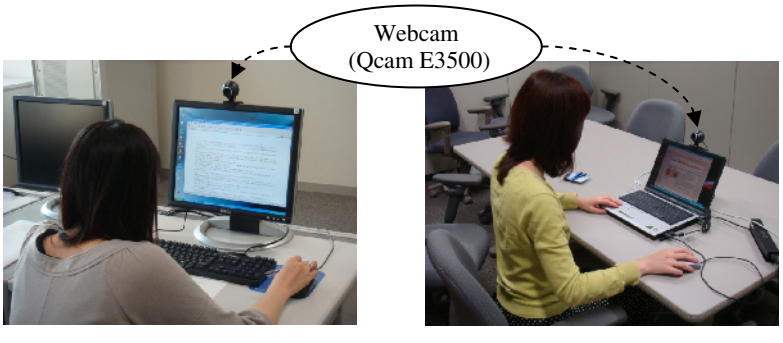

Fig. 2. Participants of the web experiment (left: with desktop PC; right: with notebook PC)

The webcam (Qcam E3500) mounted on the top of the screen (Fig. 2) captures the images of the user and the event sensor of the system detects the size of the face (face recognition part) and the body movements (motion recognition part). When the event sensor is activated, the log part records the timestamp in milliseconds, the active window's name, the URL (only when a browser is active), and the recognized face or motion size. The sensor detects the target event from each frame, and the maximum frame rate is 30 fps.

The image acquired from the camera is reduced to 1/2 size for the purpose of faster processing. The original image is 240 pixels in height by 320 pixels in width, and the converted image is 120 x 160 pixels.

3.1 Detection of Face Approaches

We used OpenCV [12] to detect the events in which a face approached the screen, as judged from the expansion of the recognized face. The face-detection function in the OpenCV library has a cascade of boosted classifiers based on Haar-like features. It excels at detecting blurred or dimmed images, although objects other than a face might be distinguished as one depending on the brightness difference in the rectangular area. We used a prepared front-face classifier. The size of the face is calculated from the sum of the width and height (in pixels) of the front face image detected by the system.

The event of a face approaching the monitor is determined on the basis of each user's baseline size. We obtained 100 baseline size data before starting the detection. The average and standard deviations of the size were calculated for each user. Face approaches were counted by comparing them with the average and standard deviations.

3.2 Detection of Other Actions

To investigate how other physical actions are related to a user's interest, the size of the user's body motions were also recorded. Body motions were calculated by summing up the widths and the heights of the motion segment detected by OpenCV.

4 Pilot Study Results

The implemented motion detector was tested in a pilot study. We evaluated its function and the relationship of the body movements to the user's interest.

4.1 Participants

Nineteen university undergraduate and graduate students from 19 to 31 years old (average age: 21.74 yrs.) participated in our preliminary test. All of them were right-handed. Seven of them did not use any visual correction devices, but the rest wore either contacts or glasses.

4.2 Procedure

Participants were asked to adjust their monitor and keyboard positions to suit their preferences. Then, they were asked to gaze at the monitor. This was the baseline session, and it lasted more than three minutes.

During the experimental session, the participants were asked to watch and evaluate the web pages according to the links on the screen. To distribute the degree of interest on a web page, we prepared 10 links to topics on the arts, academics, current news, fiction, and commercial goods. The participants were asked to visit all the links shown on the screen and other links if they wanted, and after visiting a link or viewing a new page, they were asked to rate on a scale of one to ten each page by using the following points: degree of reading (from 'did not read' to 'read carefully'), interest (from 'not interesting at all' to 'very interesting'), amusement (from 'not amusing at all' to 'very amusing'), novelty (from 'not novel at all' to 'very novel'), benefit (from 'not beneficial at all' to 'very beneficial') and easiness (from 'not easy at all' to 'very easy'). The duration of the experiment was around one hour. The face recognition and motion recognition parts of our system recorded data in both the experimental session and in the baseline session.

After this experimental session, all participants were asked whether they cared about the camera on the screen. We used these answers to evaluate the system's impact on the user.

4.3 Collected Data

The user's evaluation of each page and their actions (face approach and motion) while each page was being shown were totaled for reflecting whole tendency. The face size data in the experimental session was compared with a threshold value. We used the following function (1) and a size that was bigger than the threshold (T) value was counted as a face approach. T value was determined by adding on averaged x values in baseline period ($avg(X_{baseline})$) to standard deviation of x values ($stdev(X_{baseline})$) multiplied by coefficient α (from 1 to 3) value.

$$\begin{aligned} x &= face.height + face.width \\ T &= avg(X_{baseline}) + \alpha \times stdev(X_{baseline}) \\ f(x) &= \begin{cases} 1 & if\ x_{experiment} > T \\ 0 & otherwise \end{cases} \end{aligned} \quad (1)$$

We used only one threshold for the motion detection, because all of the users' avg. + 2 stdev. values and avg.+3stdev. values exceeded the maximum size of the recognition area (280 pixels), and most of the values recognized by the prototype system were near the maximum. We also measured the duration of each page's access.

4.4 Preliminary Results

We eliminated two of the 19 participants' log data because of a system failure in recording and an answering mistake. The face approaches were counted on each web page, and the Pearson's product-moment correlation coefficients between the counted

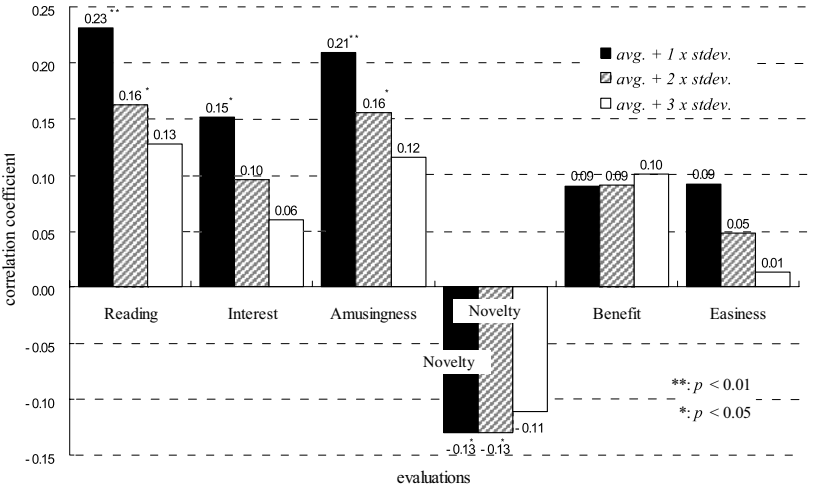

Fig. 3. Comparison of correlation coefficients between counted number of face approaches and each evaluation value from three different thresholds

Table 1. Matrix of correlation coefficients between variables

	Reading	Interest	Amusing	Novelty	Benefits	Easiness	Face	Motion	Duration
Reading	1.00	0.80(**)	0.75(**)	0.17(*)	0.50(**)	0.57(**)	0.23(**)	0.06	0.04
Interest	0.80(**)	1.00	0.87(**)	0.12	0.53(**)	0.58(**)	0.15(*)	0.04	0.07
Amusing	0.75(**)	0.87(**)	1.00	0.12	0.51(**)	0.55(**)	0.21(**)	0.10	0.08
Novelty	0.17(*)	0.12	0.12	1.00	0.32(**)	-0.07	-0.13(*)	-0.17(*)	-0.03
Benefits	0.50(**)	0.53(**)	0.51(**)	0.32(**)	1.00	0.09	0.09	0.07	0.07
Easiness	0.57(**)	0.58(**)	0.55(**)	-0.07	0.09	1.00	0.09	0.03	-0.02
Face	0.23(**)	0.15(*)	0.21(**)	-0.13(*)	0.09	0.09	1.00	0.47(**)	0.12
Motion	0.06	0.04	0.10	-0.17(*)	0.07	0.03	0.47(**)	1.00	0.07
Duration	0.04	0.07	0.08	-0.03	0.07	-0.02	0.12	0.07	1.00

** p < 0.01

* p < 0.05

number of face approaches and each evaluation value were calculated for three thresholds (Fig. 3).

For all the thresholds, all the coefficients showed the same tendency in positive and negative values. Face approach counts using avg.+1stdev. threshold showed these tendency more clearly. Significant ($p<0.01$) correlations were found between the face approach frequency and the degree of reading (0.23**) and amusement (0.21**) at the avg.+1stdev. threshold. The results at the avg.+2stdev. threshold showed similar

tendencies. Although the level of significance was low ($p<0.05$), interest was positively correlated (0.15*) and the novelty was negatively correlated (-0.13*).

The face count, motion count, page access duration time, and all the page evaluation points are summarized in the correlation matrix (Table 1). The motion count and duration did not show any significant correlation to the reading, interest, or amusement factors, but the face approach and motion counts were significantly correlated. The N size of a motion was much smaller than the others', because four of the participants did not leave motion logs in the baseline period and their data was treated as a missing value. The reading, interest, and amusement factors highly correlated with each other.

Results from questions and answers about webcam were shown below (Table 2). Most participants said that the existence of the camera didn't matter to them. Even if they had no experience in using personal computers with a webcam, they could browse without being concerned. This result showed that the camera on the computer was not considered to be out of the ordinary.

Table 2. Questions and answers about webcam

	Did you care about the camera on the screen?	**Have you ever seen a camera like this?**	**Have you ever used a PC with a camera like this?**
Yes	3	13	6
No	16	6	13

N=19

5 System Evaluation

We evaluated our prototype system from the viewpoint of presuming an interest from the face approach frequency and the practicality of the system.

5.1 Effect of Interest Estimation by Using Face Approach

There was a positive correlation between the face approach frequencies acquired by the system and the degree to which the user had actually read the page and the degree of amusement they felt while viewing it. In addition, the reading factor was highly correlated to the interest and amusement factors. These results showed that our system could be used to estimate whether a user had actually read the contents with positive emotions such as interest or amusement. However, these relationships became vague as the threshold value increased.

Motion count and page access duration were not useful data for estimating the user's interest in this experiment. The size of the acquired motion or the range of the value might be the reason for this problem. It is necessary to reexamine the sensitivity of the system or the acquisition method for characterizing an individual's behavioral pattern for the precision enhancement in addition to the face approach.

5.2 System Practicality

In our experiment, most of the participants reported that the existence of the camera did not bother their activities when exploring the web. Although the data size of a captured image was very small, the results showed the possibility to estimate a user's emotional status. This prototype system seems to be applicable to the current PC environment.

Moreover, our system did not directly record facial images, but recorded only numeric size data. In addition to the familiarity of a device, such data might be comparatively less intrusive to the user. That is, it might be useful data for the information provider, and it can be collected while maintaining the user's privacy.

6 Conclusion

We described our prototype system to identify a user's interest in the target on a PC screen using a common webcam. Although our methodology is very easy to use, the preliminary results showed the possibility of using it to identify actions reflecting the user's interest in a target. For the detailed analysis, we will evaluate this tendency of each individual.

Analyzing the whole-body movement with dedicated equipment continues to have a very important role, because our experiment is based on the relations revealed by past experiments. On the other hand, experimental analyses using real-world equipment in real-life situations are becoming more and more important now that there is an ever-increasing variety of input devices. Developing an application for an actual environment will soon be an important issue.

This technique might could be applied to various fields, although it would be necessary to set the appropriate thresholds for judging face approaches under varying circumstances. The degree of reading showed a positive correlation, and this seemed to indicate the possibility of utilizing our technique for user profiling in systems that make recommendations. To estimate the user's interest in a website, researchers have used the viewing duration, mouse movement, or eye gaze [13]. If our system proves to be valid in future experiments, it may be useful for estimating users' actions on the web.

The results of our experiments are presently being validated in real-world learning situations. We plan to analyze additional movements and revise the prototype system while applying it in an actual environment.

References

1. NIKON COOLPIX 5900, http://www.nikon-image.com/jpn/products/camera/compact/coolpix/5900/features01.htm
2. Polaroid t831, http://www.polaroid.com/global/detail.jsp?PRODUCT%3C%3Eprd_id=845524441767954&FOLDER%3C%3Efolder_id=2825744883387 93&bmUID=1224770795670&bmLocale=en_US

3. OLYMPUS FE-360, http://olympus-imaging.jp/product/compact/fe360/
4. Sementille, A.C., Lourenço, L.E., Brega, J.R., Rodello, I.: A motion captures system using passive markers. In: Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry, Singapore, June 16-18 (2004)
5. Kaiser, R., Oertel, K.: Emotions in HCI: an affective e-learning system. In: Goecke, R., Robles-Kelly, A., Caelli, T. (eds.) Proceedings of the Hcsnet Workshop on Use of Vision in Human-Computer interaction, Canberra, Australia, November 1, 2006. ACM International Conference Proceeding Series, vol. 56, 237, pp. 105–106. Australian Computer Society, Darlinghurst, Australia (2006)
6. Prendinger, H., Ma, C., Yingzi, J., Nakasone, A., Ishizuka, M.: Understanding the effect of life-like interface agents through users' eye movements. In: Proceedings of the 7th international Conference on Multimodal interfaces, ICMI 2005, Torento, Italy, October 4-6, 2005, pp. 108–115. ACM, New York (2005)
7. Jacob, R.J.: Human-computer interaction: input devices. ACM Comput. Surv. 28(1), 177–179 (1996)
8. Jacob, R.J.: The future of input devices. ACM Comput. Surv. 28(4es), 138 (1996)
9. Picard, R., Daily, S.B.: Evaluating affective interactions: Alternatives to asking what users feel. In: The 2005 CHI Workshop Evaluating Affective Interfaces (2005)
10. Mota, S., Picard, R.W.: Automated posture analysis for detecting learner's interest level. In: 1st IEEE Workshop on Computer Vision and Pattern Recognition, CVPR HCI 2003 (2003)
11. Wakai, Y., Sumi, K., Matsuyama, T.: Estimation of Human Interest Level in Choosing from Video Sequence. In: The actual use of vision technology Workshop (2005)
12. OpenCV, http://opencv.jp/
13. Hijikata, Y.: User Profiling Technique for Information Recommendation and Information Filtering. Jinkouchinougakkaishi 19(3), 365–372 (2004)