Touchless User Interface Utilizing Several Types of Sensing Technology

● Akihiro Minagawa ● Junichi Odagiri ● Yoshinobu Hotta ● Satoshi Nakashima
● Liu Wei ● Fan Wei

Recently users have come to demand a natural user interface (NUI) so that they can operate devices naturally, as an alternative to a graphical user interface (GUI). At Fujitsu Laboratories, with the aim of achieving an NUI, we are developing touchless user interfaces that make it possible to monitor users' behavior through intelligent sensing technologies, such as gesture recognition, eye tracking and speech recognition, and then understand users' intent to enable them to operate a device in a natural way. It is difficult to achieve this by using individual sensing technologies. Therefore, we have developed a way to ensure a device operation by integrating multiple sensing technologies for natural motion detection. Both gesture recognition and eye tracking technologies are combined in the developed interface, which gives the users the feeling that they are interacting more naturally and effectively with devices than they would if only individual sensing technologies were utilized. This paper describes an overview of the developed sensing technologies, the merits and problems regarding their combination, the developed interface, and future work.

1. Introduction

Many research institutions have recently been carrying out research on natural user interfaces (NUIs) to reduce users' burden by letting them use interfaces that respond to natural human motions.[1),2)]

At Fujitsu Laboratories, in an effort to improve the ease of use of interfaces at the front end of information and communications technology (ICT), which is the point of contacting with users, we are studying NUIs that make use of technologies for sensing human behavior including gesture recognition, eye tracking and speech recognition (**Figure 1**).

However, making separate use of these individual sensing technologies to build an interface may make it difficult to achieve all operations or end up forcing unnatural motions on the user.

Accordingly, as one approach to an NUI, we have developed a "touchless user interface technology" that realizes improved ease of use by integrating the gesture recognition technology used for the docomo Tablet ARROWS Tab LTE F-01D and the eye tracking assistance feature used for FMV ESPRIMO FH98/JD.

2. Outline of gesture recognition technology

This section outlines the gesture recognition technology owned by Fujitsu Laboratories. This gesture recognition technology, which is called hand gesture recognition technology, allows a user to operate a device by waving a hand in front of it as shown in **Figure 2 (a)**.[3)]

In essence, this technology can be used to calculate the motion of the skin (hand) by detecting the difference of the skin from two successively captured images, as shown in **Figure 2 (b)**. For the calculated motion, an operation associated in advance, e.g., rightward movement of the cursor and downward scroll of the screen, is performed.

Of operations by hand gesture, those achieved by waving a hand, such as operations made by instantaneous motion including deciding on a menu item and scrolling in a certain direction by a certain amount, generally impose less burden on users. Meanwhile, operations that use hand information over an extended period of time, e.g., keeping the hand in a certain position to maintain the same cursor position and moving the cursor according to the hand position, require the

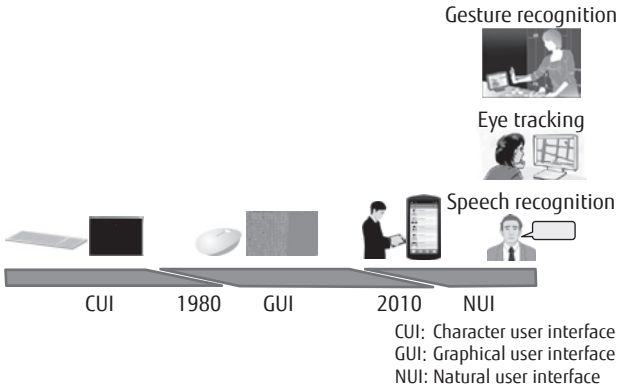


Figure 1
History of user interfaces.

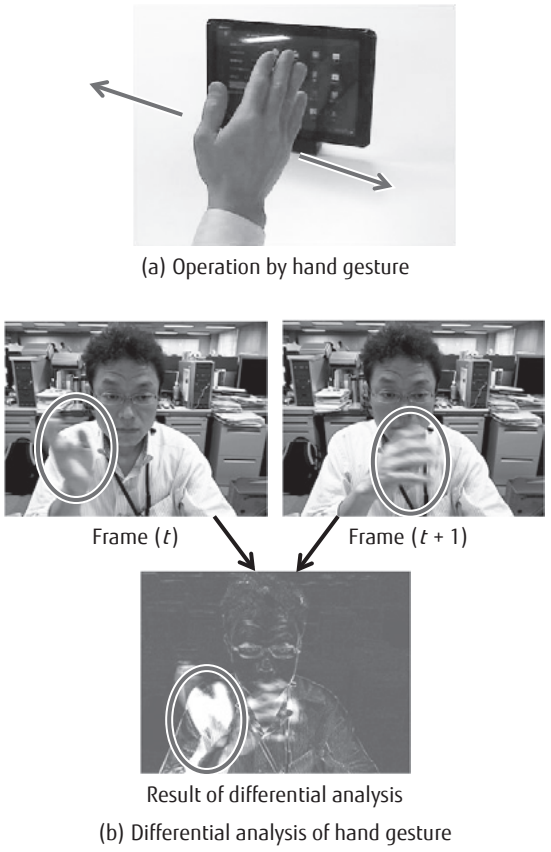


Figure 2
Outline of gesture recognition by hand gesture.

user to keep the hand up. It poses issues such as a greater burden on the user and an inclination to force the user to move in an unnatural way.

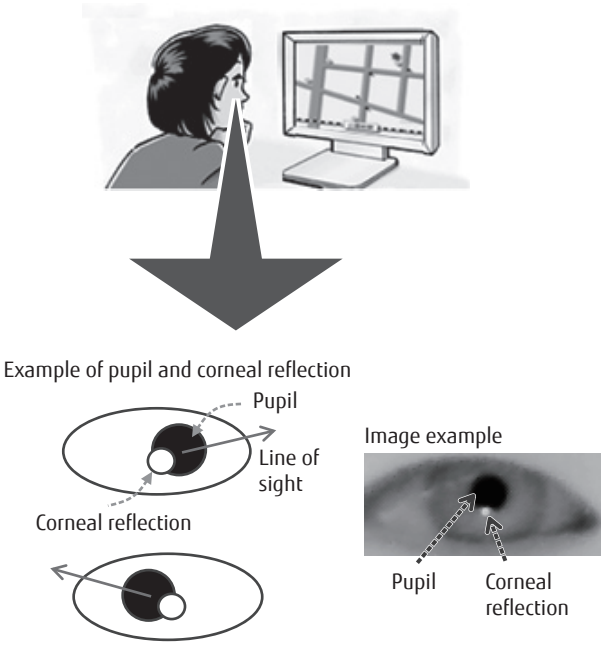


Figure 3
Outline of eye tracking technology.

3. Outline of eye tracking technology

The eye tracking technology of Fujitsu Laboratories uses detection by reflection of light on the cornea (corneal reflection method). In order to detect corneal reflection, an LED is used to emit near-infrared (IR) light and an IR camera is used to capture the image from which the pupil of the eye and corneal reflection are detected. Based on their positional relationship, the direction of sight is calculated. The principle is illustrated in **Figure 3**. See reference 4) for the details of this technology.

An interface based on the line of sight works well in operations such as moving the cursor to a certain position. On the other hand, to achieve a command input such as a decision operation, it may involve unnatural line-of-sight motions such as having to provide a command input with a certain key and move the line of sight to the position of the key. This is one issue with such an interface in terms of operation.

4. Merits and issues regarding integration

A multimodal interface integrating these sensing technologies can solve the respective issues and realize an easy-to-use interface.[5] A multimodal interface

is a generic name for an interface that uses multiple sensing technologies and combines various input modes according to the respective sensing methods. Conventional multimodal interfaces, however, are constructed by loose coupling of the respective modes. For this reason, motions used multimodally and motions that use the same sensing technologies individually are basically unvaried. It means that, with operations that cannot be replaced by other sensing technologies, the issues with individual sensing technologies as shown below will remain as they are, and this is a problem that has to be avoided.

1) Achievement of operations made by natural motions (avoidance of motions imposing high burden)

Of interfaces, easy operations and operations that do not place a high burden on the user even if performed for a long time are generally desired. For this reason, it is important to avoid the operations that require the user to maintain the same position for a long time or the operations that require the user to move in an unnatural posture.

2) Extraction of new information by integration of sensing information (detection of motions according to the operation intent)

When motions with and without operation intent are similar or identical, whether there is operation intent or not must be judged for each motion. In an example of gesture recognition, it corresponds to deciding whether a hand motion needs to be recognized as a certain hand movement or whether it is a hand motion made by the user unconsciously. In an example of speech recognition, it corresponds to deciding whether an utterance is not related to device operation or whether it is intended for operation. With a touchscreen, the user's operation intent can be judged by the physical action of touching. With a touchless interface, however, this judgment is difficult to make. This problem is evident especially with NUIs, which are operated by natural motions, and correct recognition of the user's operation intent is important for NUIs.

In addition, there is a new issue with multimodal interfaces, which is described below.

3) Increased cost due to the increased number of sensors

Larger numbers of sensors used lead to increased cost of devices. Accordingly, reducing the number of sensors is one important issue with multimodal interfaces.

In order to realize an interface based on the NUI framework, these issues need to be resolved.

5. Developed technologies

Based on the perspectives described in the previous section, we have prototyped an integrated interface capable of touchless screen scroll and scaling. The system of operation includes scroll operation and provision of the center location of scaling by the line of sight and scaling operation by gesture. The following describes the developed interface by referring to the three issues mentioned in the previous section.

First, avoidance of motions which impose a high burden on the user has been achieved by two methods. One is assignment of the optimum operation by integration. As described in the "Outline of gesture recognition technology" and "Outline of eye tracking technology" sections, gesture information and line-of-sight information can be integrated for operation by gesture recognition and eye tracking. Gesture information is more suitable for command operations and line-of-sight information is better suited for cursor operations, which means that the respective operations can be combined in a complementary manner to eliminate the need for awkward operations. This leads to more natural motions (**Figure 4**).

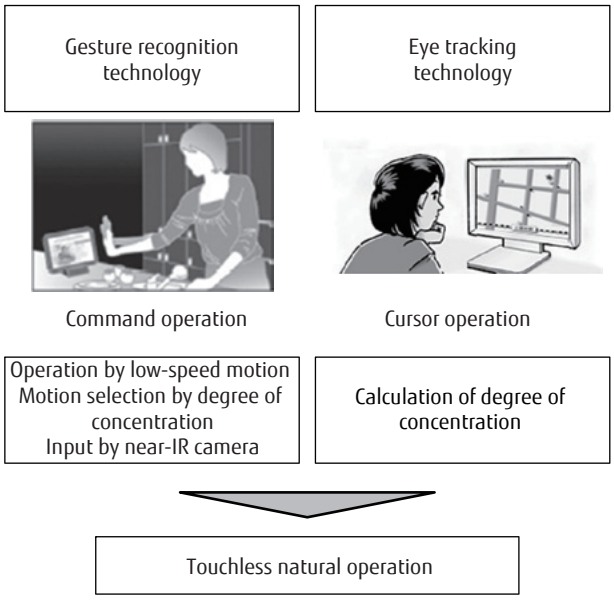


Figure 4
Gesture recognition technology and eye tracking.

The other is to avoid the unnatural motions by changing motions that are associated with operations. For gesture recognition, large motions or motions that are not usually made may be intentionally used to make the device recognize the operation intent. Use of hand gestures unavoidably forces the user to make large motions because motions need to be detected. To address this problem, we have developed gesture recognition that allows a device to be operated by holding up the hand. The user does not need to make high-speed or large motions, and can achieve operations by making low-speed or small motions alone.

For the second issue of detecting motions according to operation intent, we have decided to take advantage of the benefits of multimodal interfaces to determine the presence of operation intent based on certain sensing information and apply it to the detection result of other sensing information, thereby resolving the issue. Line-of-sight information not only indicates the coordinates at which the user is currently looking but also indicates whether or not the user is paying close attention to one thing over a certain period of time. If the degree of attention paid by the use is high, any motion detected by another sensor can be regarded as being made for operation. Conversely, if the degree of attention is low, the motion is likely to be not intended for operation. Based on such observation, the degree of concentration can be calculated from line-of-sight information and an original filter for gesture recognition can be created from the degree of concentration to detect only the intended gesture. This allows suppression of unnecessary gesture detection even in the present case, in which small motions are assigned.

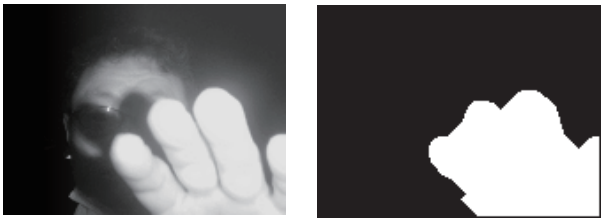
To address the third issue of increased cost due to the increased number of sensors, we have reduced the number of sensors by using one sensor for detection of more than one piece of information. Specifically, we have used the near-IR camera and LED intended for detecting the line of sight also for gesture recognition to share the sensor and thus reduce the cost.

With this shared use of the sensor, an image from the near-IR camera is used as the input image for gesture recognition. In this image, objects closer to the camera are captured with higher brightness and more distant objects are darker. Accordingly, to accommodate this characteristic, we have changed the gesture recognition method.

The developed hand gesture recognition detects the face first and identifies the location of the face. Then, it decides whether there is any object in the picture that could be a hand in front of the face. When an operation is an intended one, the hand is held before the face as a natural gesture and the hand is closer to the camera. Based on this characteristic and the fact that, of the hand skin and face skin, which have almost the same reflectance, the one closer to the camera has higher brightness, the technology determines whether there is any area with higher brightness than that of the face. Any area that is brighter, or closer to the camera, is detected and whether the area is the hand or not is separately judged (**Figure 5**). In this way, comparing the brightness of the face and hand allows the hand to be detected at a distant location. With this detection method, the hand can be detected even if it is about 50 cm away from the camera, and this makes detection possible at a longer distance than the existing gesture recognition that makes use of near-IR light.

Table 1 shows the result of evaluating the prototype above in a 9462-frame image sequence.

This sample includes 261 attempts (5348 frames) of motions with operation intent and the remaining approximately 40% (4114 frames) represents motions without operation intent in the image sequence. This sample has been evaluated in terms of reproducibility



(a) Input image (b) Area judged as hand

Figure 5
Example of input image of held up hand from near-IR camera and area judged as hand.

Table 1
Evaluation result of prototype interface.

	Detection rate and number of detections
Reproducibility	85.4% (223/261)
Relevance	97.0% (223/230)

and relevance, which are defined as follows:

Reproducibility: Of the motions with operation intent, the rate of operations achieved as intended

Relevance: Of the operations, the rate of motions with operation intent

As a result, reproducibility of 85.4% and relevance of 97.0% have been achieved. The non-detection in reproducibility was mostly because of non-detection of the hand when the distance between the face and hand was not long. Meanwhile, the major factor of the misdetection in relevance was, in addition to misjudgment on the hand, misdetecting the face as the hand after the hand disappeared from the screen.

6. Future issues

On top of further improvement of the detection accuracy, future issues include the following two.

1) Relation between ease of use and number of operation commands

The present prototype allows intuitive operations in two directions including screen scaling, and determines which of the two types of operations is intended according to the shape of the hand held up. However, an increased number of operation commands may require the user to learn the shapes and may generate misjudgment as well. To ensure that the number of operation commands is easy to use for the user, measures are necessary such as assigning small command sets according to situations of use so as to prevent placing an increased burden on the user.

2) Provision of "feedback to user" feature

With hand gesture detection using a camera, the position and state of the hand were generally made changeable so that the user could make motions in the detection area by presenting the image to the user showing whether the hand was in the detection area. However, images from a near-IR camera as in the present study are different from visible images that can be usually obtained, which makes it difficult to associate an image with the real space even if it is presented. It confuses the user and makes accurate feedback difficult. For presentation technology to provide a feedback feature in such environment, further study is necessary in the future.

7. Conclusion

This paper presented, as one approach to NUI, the development of touchless user interfaces by multimodal interfaces integrating gesture recognition and eye tracking technologies. We have integrated hand-holding gesture recognition and eye tracking technologies using a near-IR camera to prototype an interface that can be operated by natural motions and imposes less burden on the user. As a result of an experiment, it has been found out that non-detection may occur depending on the distance between the hand and face.

In the future, we intend to verify the feasibility of the prototype mentioned above to work on easier-to-use integration and build an integrated interface with less non-detection by improving the accuracy.

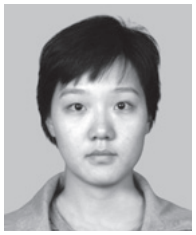
References

1) D. Wigdor et al.: Brave NUI World: Designing Natural User Interfaces for Touch and Gesture. Elsevier, 2011.
2) Y. Katsumura et al.: "Natural UI" on the horizon: Making PC Operation Human, Nikkei Online Edition October 20, 2011 (in Japanese).
http://www.nikkei.com/article/DGXNASFK1702S_X11C11A0000000/
3) H. Yoshizawa et al.: Development of and Future Prospects for Tablets Devices. *Fujitsu Sci. Tech. J.*, Vol. 49, No. 2, pp. 208–212 (2013).
4) T. Kogure et al.: Eye Tracking Technology to Assist PC Use. *FUJITSU*, Vol. 64, No. 3, pp. 293–297 (2013) (in Japanese).
5) F. Takahashi et al.: Magic UI. *Nikkei Electronics*, No. 1081, April 30, 2012, pp. 32–61 (2012) (in Japanese).

Akihiro Minagawa
Fujitsu Laboratories Ltd.
Mr. Minagawa is currently engaged in research related to image processing and recognition.

Satoshi Nakashima
Fujitsu Laboratories Ltd.
Mr. Nakashima is currently engaged in research and development related to eye tracking technology.

Junichi Odagiri
Fujitsu Laboratories Ltd.
Mr. Odagiri is currently engaged in research and development related to eye tracking technology.



Liu Wei
Fujitsu Research and Development Center Co., Ltd.
Ms. Wei is currently engaged in research related to image processing and recognition.

Yoshinobu Hotta
Fujitsu Laboratories Ltd.
Mr. Hotta is currently engaged in research related to image processing and recognition.

Fan Wei
Fujitsu Research and Development Center Co., Ltd.
Mr. Wei is currently engaged in research related to image processing and recognition.