https://doi.org/10.20965/ijat.2017.p0378

Paper:

Three-Dimensional Input System Employing Pinching Gestures for Robot Design

Kiyoshi Hoshino[†] and Keita Hamamatsu

University of Tsukuba
1-1-1 Tennodai, Tsukuba 305-8573, Japan
[†]Corresponding author, E-mail: hoshino@esys.tsukuba.ac.jp
[Received October 5, 2016; accepted February 21, 2017]

Several studies of input interfaces capable of recognizing the gestures have been conducted but most of them use the user's fingers to enter the position data. These finger-based input interfaces are difficult to provide a so-called click & drag function (as in a mouse) and some of them request for the user to take uncomfortable gestures. When people pinch any objects, however, basically their thumb and index finger come into contact with each other or separate them from each other. These pinching gestures provide superior benefits as the gestures, which may contribute to the input interfaces. This study proposes the method for detecting 3D finger positions and estimating 3D hand postures in pinching gestures based on information on depth images captured by a depth sensor, especially from the viewpoint of robot design. That produces benefits including button-clicking-like input operation by means of contact between the fingers; user's comfortable gestures as in daily life; clicking action independent of input of positions and postures; and clear identification between ON and OFF. As the evaluation of the 3D input interface proposed here, the authors design real products with the system and a 3D printer, suggesting that the users can design precise and fine 3D objects with his/her comfortable daily gestures with highest usability.

Keywords: three-dimensional input system, pinching gestures, depth sensor, robot design

1. Introduction

"Strati" [1], for instance, is the world's first 3D-printed electric car, and was manufactured using a large-scale 3D printer. Not only car body and car components, however, but also robot body and the body components should be together designed with design, function, and features so as to provide safety, comfortability, and highest usability. But it always needs skilled hands with a lot of time, labors, and costs. Three-dimensional input system with highest usability is needed for that.

To become skilled in the use of tools, typified by 3DCAD and 3DCG [2–7], for drawing the three-dimensional shapes of objects on a computer, high-level of expertise and skill are often required. One of the reasons is less intuitive property of input interfaces as input devices, such as a mouse device, in designing three-dimensional objects. Against the recent background of widespread of 3D technology in output devices, as seen in familiarization of 3D display techniques and commercial-availability of inexpensive 3D printers, the demand for intuitive 3D input interfaces has risen. It is desirable that these interfaces enable any gestures that people take in the real world to be entered with no need for a special input device to ensure their intuitiveness. To satisfy this requirement, input interfaces need to recognize the daily and common gestures by people. Several studies of input interfaces capable of recognizing the gestures have been conducted but most of them use the user's fingers to enter the position data [8–11]. These finger-based input interfaces are difficult to provide a so-called click & drag function (as in a mouse) and some of them request for the user to take uncomfortable gestures [12–15]. To address this problem, our study focused on pinching gestures.

When people pinch any objects, basically their thumb and index finger come into contact with each other or separate them from each other. These pinching gestures provide superior benefits as the gestures, which may contribute to the input interfaces. These benefits include button-clicking-like input operation by means of contact between the fingers; user's comfortable gestures as in daily life; clicking action independent of input of positions and postures; and clear identification between ON and OFF.

Wilson et al. have proposed the method for recognizing pinching gestures using a RGB camera as one of GUI operation methods in the desktop environment [16, 17]. This method makes elliptical approximation for internal region defined by the thumb and index finger and detect the position and posture based on the obtained information on the long and short axes of the ellipse. However, only data on relative variation for the posture parameter can be acquired; accordingly, it is unsuitable for delicate input.

Moreover, Fukuchi et al. have incorporated the pinching gesture recognition function in gesture input for the table-top entertainment system [18, 19]. They uses such technique that the gravity center of the internal region defined by the two fingers is assumed to be the coordinate

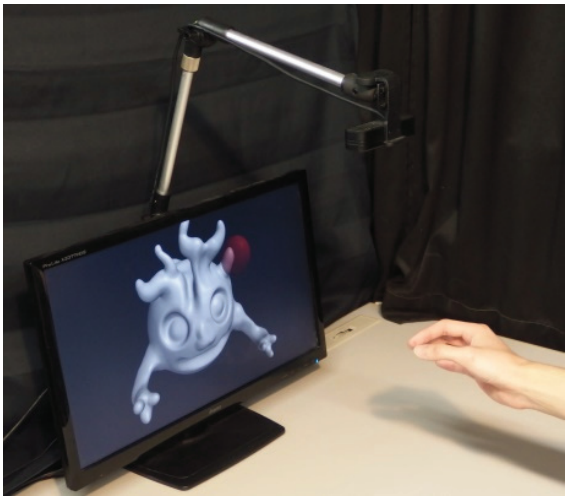

Fig. 1. Configuration of the system.

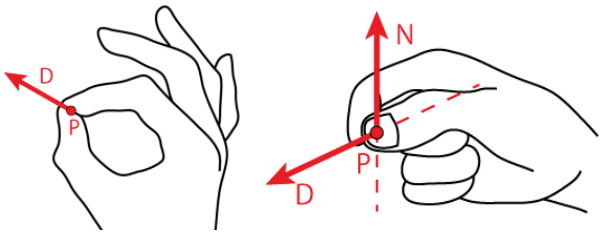

Fig. 2. System output.

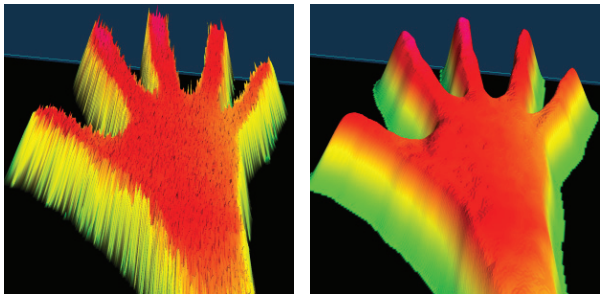

Fig. 3. Noise cancellation using the filters.

Fig. 4. Depth-gradient-based extraction of the contour region.

for position input, a line connecting the gravity centers of the internal region and the arm region is used as posture input. Assuming that multiple users use the system simultaneously, this technique provides high-speed processing and allows for the detection of postures in the absolute direction; however, it detects the postures in only the direction close to the arm and does not capture the delicate movements of the fingers. In addition, both of the studies, which define the positions based on the gravity center of the internal region defined by the two fingers, are incapable of detecting these positions themselves.

This study therefore proposes the method for detecting 3D finger positions and estimating 3D hand postures in pinching gestures based on information on depth images captured by a depth sensor. Moreover, the authors propose a 3D input interface using the method proposed here.

2. Configuration of System

The system configuration is as shown in **Fig. 1**. As known from this figure, the image of a user's hand is captured by a depth sensor attached to the top of a monitor. The sensor DepthSense 325 (SoftKinetic) is used. This inexpensive Time-of-Flight (TOF) type depth sensor is capable of capturing the images with a resolution of 320×240 pixels at max. 60 fps.

System input is depth map data, in which 3D coordinates are defined for each of pixels acquired by the depth sensor. In the output, the finger positions, hand posture, and whether the fingers are in contact with each other or not are detected in pinching gestures as shown in **Fig. 2**.

3. Pinching Gesture Recognition

3.1. Filtering

The depth information acquired by the depth sensor has mixed noises and across the image is preprocessing for noise cancellation and smoothing using median and Gaussian filters, as shown in **Fig. 3**.

3.2. Extraction of the Internal Contour Regions

To detect pinching gestures on the images, a focus is placed on the closed space defined by the fingers. To extract this close space as a region on the image, the depth gradient between the individual pixels on the depth image is used. The gradients in the x and y direction are calculated for the individual pixels on the noise-cancelled images using a 3×3 Scharr filter and the norm of the obtained 2D gradient vector is assumed to be the gradient value for individual pixels. Then, binarization is performed based on the gradient values across the image, making it possible to extract the region corresponding to the hand contour as shown in **Fig. 4**.

Next, whether any pinching gestures are taken or not is detected. The contour regions are compared between the present and previous frames and the number of pixels of the region separated from the largest contour region is calculated. If the calculated value is larger than or equal to

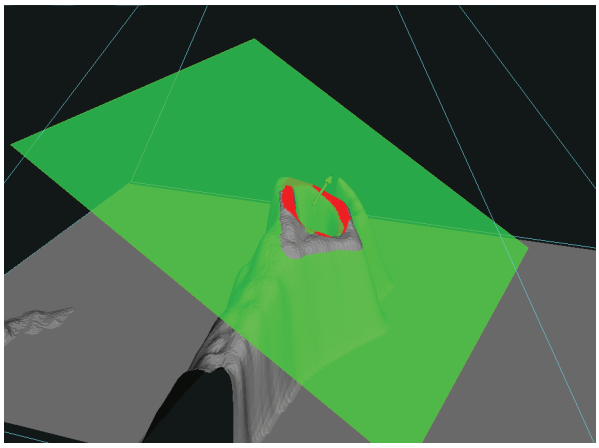

Fig. 5. Plane approximation.

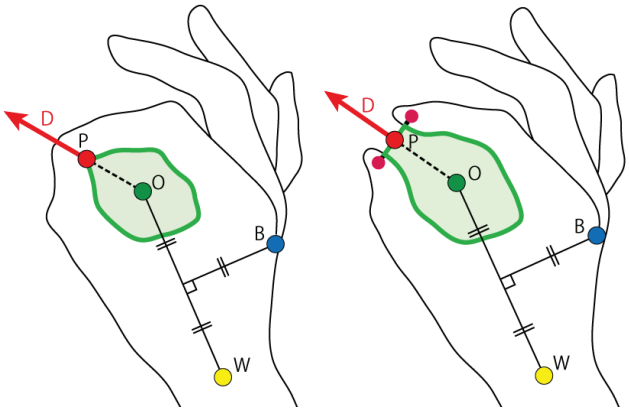

Fig. 6. Relationship between the internal contour region and the feature points.

the threshold, it is determined that two fingers are in contact with each other, namely a pinching gesture is taken. The detailed parameters for the pinching gesture are calculated based on the internal contour region as shown in **Fig. 4(c)**, which is detected in this process. If the area of the separated contour region is smaller than the threshold, it is determined that no pinching gesture occurs.

3.3. Plane Approximation by Multi-Regression Analysis

To estimate the hand posture, the shapes of the hand in the proximity of the fingers are approximated in a 3D plane based on 3D information on the hand and fingers around the internal contour as described in Section 3.2. The sampling points for the hand are extracted to use in approximation. First, the convex hull area of the internal region is obtained. Then, the original convex hull area and the whole contour region found in Section 3.2 are subtracted from the region obtained from addition of image expansion gain to the area to extract the region as shown in **Fig. 4(d)**. The pixels within this region form a group of points, which have 3D coordinates, together, making it possible to approximate the group of points in the special plane by multi-regression analysis using the least-square method, as shown in **Fig. 5**.

3.4. Extraction of Feature Points

To determine the finger positions and hand postures, some feature points are defined in the plane obtained Section 3.3. First, the gravity center is calculated on the depth image based on the concave hull area obtained Section 3.3 and the point in the 3D space obtained by projecting the found gravity point into the 3D approximation plane is assumed to be O. Second, a sphere, which may cover the finger region around this O point, is defined and the point close to the boundary of the sphere is extracted to extract the region around the wrist. Third, the point obtained by projecting the gravity center into the plane is defined as the point W, which indicates the wrist position, in the same manner as that in obtaining the point O for this region. Fourth, out of the vertexes of a quadrate having a line segment OW connecting the points O and W, as a diagonal, a point B on the back of the hand is defined to be B. Finally, among the pixels within the internal contour region obtained in Section 3.2, the point obtained by projecting the furthermost point from the point B into the approximation plane is assumed to be the point P. This point P is defined to be the position coordinate for the finger and used as system output. Moreover, assuming that the vector $\overrightarrow{OP}$ be a directional vector and the normal vector in the approximation plane be an upward vector, the 3D hand postures may be defined in the approximation plane, as shown in **Fig. 6**.

4. Pinching Gesture Detection by Finger Detection

When the fingers are not in contact with each other, the internal contour region as shown in the Section 3 cannot be extracted, leading to failure to define the finger position and hand postures. To address this problem, the finger positions are detected from the depth image and the contour region is cut off based on the finger position, making it possible to detect almost the same region as that obtained when the fingers are in contact with each other.

5. Pinching Gesture Recognition

5.1. Extraction of Feature Points

When the fingers are not in contact with each other, the finger positions and hand postures cannot be defined because the internal contour region as shown in the Section 3 may not be extracted. The finger positions are detected on the depth image and the contour region is cut off based on the obtained finger positions, making it possible to detect almost the same region as that obtained when the fingers are in contact with each other.

5.2. Finger Position Detection Using Image Isolation Levels

The finger positions on the depth image correspond to the end point or extremal point when seen from the hand region. A large difference in depth may be observed between the pixels containing these points and the pixels in the proximity of them. To avoid this difference, the procedure described below is followed to evaluate the difference in depth between the individual pixels and those in the proximity of them and the finger region is extracted based on the values obtained from the evaluation.

First, 32 line segments radially extending from the target pixels on the depth image are defined. The lengths of these line segments are fixed in the 3D coordinate system and their size are preset to that about three times the actual finger size. Second, the individual pixels corresponding to the line segments and the target pixels are compared. If one or more pixels, of which difference in depth from the target pixels is larger than or equal to the thresholds, are detected among the pixels on the line segment extending toward one direction, it is assumed that the target pixels be isolated in the direction. In contrast, if the differences in depth between the pixels on the line segment and the target pixels are all within the threshold, they are assumed to have smooth continuity. This evaluation is performed on the line segments extending in 32 direction and the isolation levels of the target pixels from the pixels in the proximity of them are quantified using the grades of each 32 direction. These values are conveniently referred to as isolation levels and the isolation levels for the individual pixels are used for finger region extraction.

A schematic diagram explaining isolation evaluation is shown in **Fig. 7**. The center points shown in **Figs. 7(a)** and **(b)** represent the target pixels. The line segments in 16 directions are conveniently defined and the isolation levels for the target pixels are quantified; for example, out of 16 line segments, 15 segments are isolated in **Fig. 7(a)** and 10 segments in **Fig. 7(b)**.

5.3. Extraction of Finger Positions and the Internal Region

To obtain the distribution of the end region as shown in **Fig. 8(a)**, the isolation evaluation is performed on all the pixels of the depth image. The region, which is rich in higher isolation level of pixels, corresponds to the finger region. The gravity center of the pixel area (**Fig. 8(c)**), which belongs to both of a high isolation region and the contour region, is selected among the points most close to the tips of the fingers in the region, making it possible to detect he finger positions as shown in **Fig. 8(d)**. This method independent of the hand silhouette shape is capable of detecting robustly the finger positions even from the area where any change in hand posture and finger position is difficult to detect. Finally, the pseudo internal region is extracted by separating the gradient region based on the line segment connecting the thumb position point and the point of the finger most close to the thumb point (**Fig. 9**). This region can be treated in the same manner as that of the internal contour region, making it possible to define the feature points. This enables the finger positions and hand posture to be estimated even when the fingers are not in contact with each other (**Fig. 6(b)**).

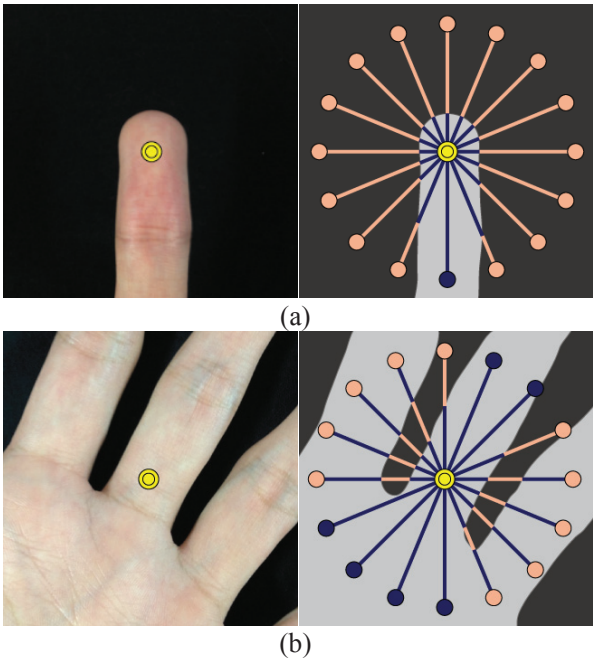

Fig. 7. Isolation evaluation.

Fig. 8. Finger position detection based on the isolation distribution.

Fig. 9. Pseudo internal contour region.

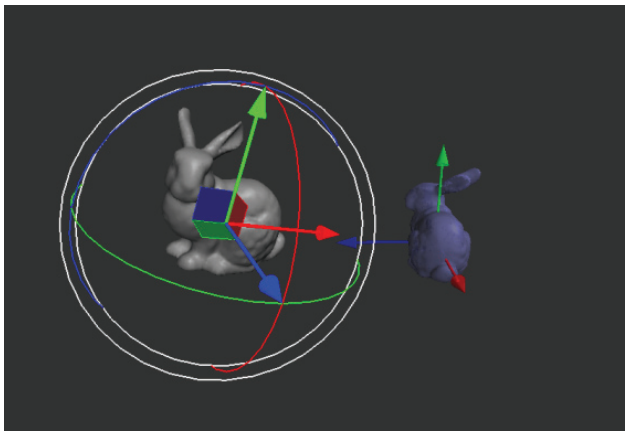

Fig. 10. An image viewed on the screen during the evaluation test.

6. System Evaluation

6.1. Time Required for Data Entry

As illustrated in **Fig. 10**, the subjects were asked to manipulate an object in a virtual space viewed on a computer screen and place it in a pre-determined position with given posture to measure the time required for them to finish this task. In this experiment, the subjects were asked to perform the same task as mentioned above using two different techniques; one being proposed one and the other being GUI manipulation using a computer mouse. The measured times were compared to evaluate the usefulness of the proposed technique as an input interface.

On the screen visible to the subjects, a virtual 3D space as shown in **Fig. 10** appears with an animal figure and a translucent object therein. The subjects changed the position and posture of the animal to be manipulated using drag-and-drop operation by means of pinching gestures so as to fit on the target object. On the other hand, to perform GUI manipulation, the GUI viewed around the animal to be manipulated is operated by computer mouse drag-and-drop operation. GUI manipulation allows us to enter information on the position and posture of the animal independently with six degrees of freedom. These two techniques are used to move the animal closer to the target object. Once the position and posture of the animal have gotten closer to those of the target object at a certain level or more, the task finishes and the next target object appears. This matching task was repeated on the third target object and the amount of time required was calculated. Assuming that this task was one experimental session, the subjects were asked to use two different input techniques to perform the task ten times. Three university students (two males and one female) served this experiment.

The result of the experiment is shown in **Table 1**. It was verified in all the three subjects that the proposed input technique required less time to perform the tasks than the mouse-based GUI input technique. The average of times, which the proposed technique required to perform tasks ten times, was less than one-third of that of the mouse-based technique in all the three subjects. Thus, it may be concluded that the proposed input technique is very useful in performing the tasks such as manipulation of 3D objects as a 3D input interface. Moreover, from the viewpoint of a small difference in working hours between the subjects and less requirement of proficiency, the proposed system could be said to be an intuitive-type interface.

Table 1. Average time required to perform the task for each subject.

	Computer mouse	Proposed method
Subject A	105.2 s	27.7 s
Subject B	129.2 s	42.9 s
Subject C	123.4 s	28.7 s

In the earlier system [20] proposed by the authors, to detect the directions of pinching gestures, a closed region defined between the thumb and the middle finger was always used. Specifically, the directions of pinching gestures could be detected only when "the system is assumed to be ON." For this reason, the directions of pinching gestures could not be estimated as long as the closed region was detected by a depth sensor. Moreover, only five points within the region defined between the thumb and the middle finger were used. Accordingly, the directions could not be detected favorably. In contrast, the proposed input technique does not necessarily require the closed region defined between the thumb and the middle finger, while taking advantage of multiple points within the region, allowing us to detect the postures at high precisions even if the hand is considerably departed from the depth camera.

6.2. An Example of Figures Created by Pinching Gestures

Figure 11 shows an example of 3D figures created by a subject, who has skilled in the proposed system, using pinching gestures. In the figure, some figures were created by drawing according to the shape data on the display screen and the other by outputting the shape data on a 3D printer. All the figures, which have complicated shapes, were created by pinching gestures over 4 to 5 hours only. It can be well understood that the proposed system allows us to create intuitively complicated 3D figures.

7. Conclusion

High-level of expertise and skill are often required, so as to become skilled in the use of tools, typified by 3DCAD and 3DCG, for drawing the three-dimensional shapes of objects on a computer. One of the reasons is less intuitive property of input interfaces as input devices, such as a mouse device, in designing three-dimensional objects. Against the recent background of widespread of 3D technology in output devices, as seen in familiarization of 3D display techniques and commercial-availability

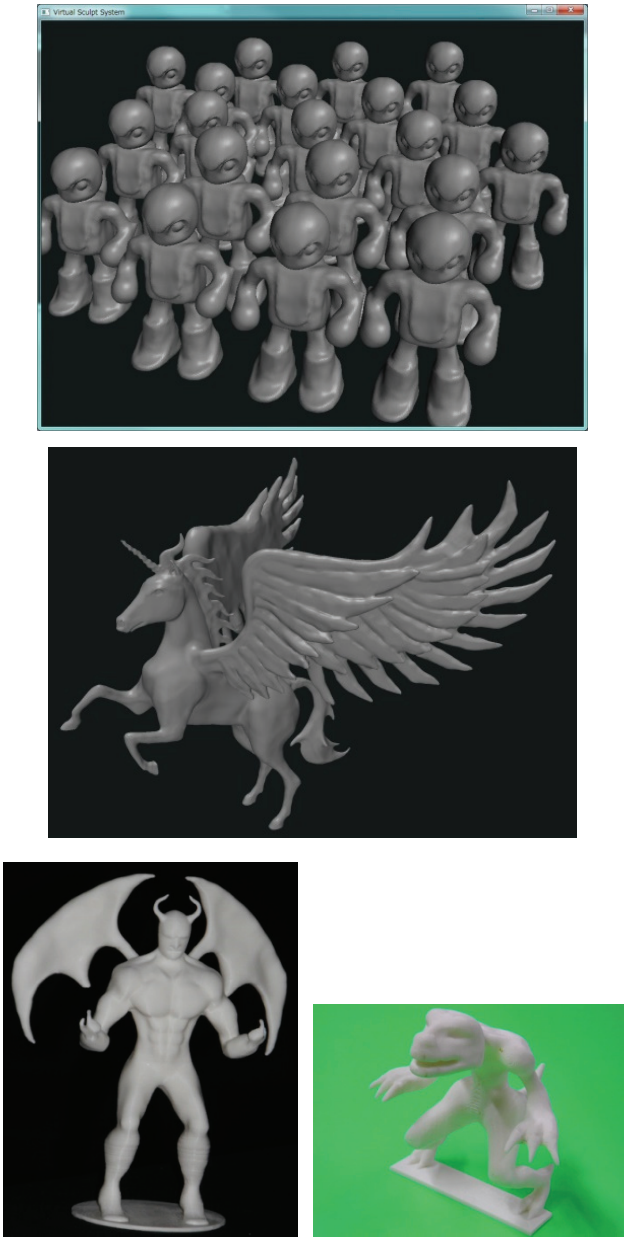

Fig. 11. Examples of designed 3D shapes with gestures and 3D-printed figures.

of inexpensive 3D printers, the demand for intuitive 3D input interfaces has risen. It is desirable that these interfaces enable any gestures that people take in the real world to be entered with no need for a special input device to ensure their intuitiveness. To satisfy this requirement, input interfaces need to recognize the daily and common gestures by people. Several studies of input interfaces capable of recognizing the gestures have been conducted but most of them use the user's fingers to enter the position data actually. These finger-based input interfaces are difficult to provide a so-called click & drag function (as in a mouse) and some of them request for the user to take uncomfortable gestures.

To address this problem, this study focused on pinching gestures. When people pinch any objects, basically their thumb and index finger come into contact with each other or separate them from each other. These pinching gestures provide superior benefits as the gestures, which may contribute to the input interfaces. These benefits include button-clicking-like input operation by means of contact between the fingers; user's comfortable gestures as in daily life; clicking action independent of input of positions and postures; and clear identification between ON and OFF.

Focusing on finger pinching gestures as an intuitive 3D input interface for working in a 3D virtual space, this study has proposed an interface, which enables the finger positions and hand postures to be input simultaneously using the pinching gestures. The proposed method, which allows the finger positions and hand posture to be recognized with no pinching gesture, may use the information on whether the fingers are in contact with each other or not as input independent of the finger positions and hand postures.

As the evaluation of the 3D input interface proposed here, the authors design the real products with the system and a 3D printer, suggesting that the users can design precise and fine 3D objects with his/her comfortable daily gestures with highest usability. This kind of 3D input device will help engineers design the robot body and the components, without their skilled hands, a lot of time, labors, and costs.

Acknowledgements

A part of this research was conducted with the assistance of the Strategic Information and Communication R and D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communication, KDDI Foundation, and the Adaptable and Seamless Technology Transfer Program through Target-driven R&D (A-STEP) of Japan Science and Technology Agency (JST). The authors would like to extend our sincere gratitude to these organizations.

References:

[1] "3D-printed Car by Local Motors – The Strati." https://www.youtube.com/watch?v=daioWlkH7ZI [Accessed September 7, 2014]

[2] Y. Boz, O. Demir, and I. Lazoglu, "Model based feedrate scheduling for free-form surface machining," Int. J. on Automation Technology (IJAT), Vol.4, No.3, pp. 273-283, 2010.

[3] F. Tanaka, "Current situation and problems for representation of tolerance and surface texture in 3D CAD model," Int. J. on Automation Technology (IJAT), Vol.5, No.2, pp. 201-205, 2011.

[4] E. Kunii, T. Matsuura, S. Fukushige, and Y. Umeda, "Proposal of consistency management method between product and its life cycle for supporting life cycle design," Int. J. on Automation Technology (IJAT), Vol.6, No.3, pp. 272-278, 2012.

[5] K. Takasugi, H. Tanaka, M. Jono, and N. Asakawa, "Development of a forging type rapid prototyping system (relationship between hammering direction and product shape)," Int. J. on Automation Technology (IJAT), Vol.6, No.1, pp. 38-45, 2012.

[6] S. Kanai, T. Shibata, and T. Kawashima, "Feature-based 3D process planning for MEMS fabrication," Int. J. on Automation Technology (IJAT), Vol.8, No.3, pp. 406-419, 2014.

[7] M. M. Isnaini, Y. Shinoki, R. Sato, and K. Shirase, "Development of a CAD-CAM interaction system to generate a flexible machining process plan," Int. J. on Automation Technology (IJAT), Vol.9, No.2, pp. 104-114, 2015.

[8] Y. Sato, Y. Kobayashi, and H. Koike, "Fast tracking of hands and fingertips in infrared images for augmented desk interface," Proc. Fourth IEEE Int. Conf, on Automatic Face and Gesture Recognition, pp. 462-467, 2000.
[9] C. Von Hardenberg and F. Bérard, "Bare-hand human-computer interaction," ACM Proc. the 2001 workshop on Perceptive user interfaces, pp. 1-8, 2001.
[10] K. Oka, Y. Sato, and H. Koike, "Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems," Proc. 5th IEEE on Automatic Face and Gesture Recognition, pp. 429-434, 2002.
[11] F. Bérard, J. Ip, M. Benovoy, D. El-Shimy, J. R. Blum, J. R. Cooperstock, "Did 'Minority Report' get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task," IFIP Conf. on Human-Computer Interaction, pp. 400-414, 2009.
[12] Z. Zhang, "Vision-based interaction with fingers and papers," Proc. Int. Symposium on the CREST Digital Archiving Project, pp. 83-106, 2003.
[13] S. Malik and J. Laszlo, "Visual touchpad: a two-handed gestural input device," ACM Proc. the 6th Int. Conf, on Multimodal Interfaces, pp. 289-296, 2004.
[14] S. Malik, A. Ranjan, and R. Balakrishnan, "Interacting with large displays from a distance with vision-tracked multi-finger gestural input," Proc. 18th annual ACM symposium on User interface software and technology, pp. 43-52, 2005.
[15] Z. Zhang, "Microsoft kinect sensor and its effect," IEEE multimedia, Vol.19, No.2, pp. 4-10, 2012.
[16] A. D. Wilson, "Robust Computer Vision-Based Detection of Pinching for One and Two-Handed Gesture Input," Proc. UIST'06, ACM Press, pp. 255-258, 2006.
[17] A. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk, "Bringing physics to the surface," Proc. UIST '08 ACM Proc. 21st annual ACM symposium on User interface software and technology, pp. 67-76, 2008.
[18] K. Oka, Y. Sato, and H. Koike, "Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems," Proc. 5th IEEE on Automatic Face and Gesture Recognition, pp. 429-434, 2002.
[19] K. Fukuchi, T. Sato, H. Mamiya, and H. Koike, "Pac-pac: pinching gesture recognition for tabletop entertainment system," ACM Proc. the Int. Conf. on Advanced Visual Interfaces, pp. 267-273, 2010.
[20] K. Hamamatsu and K. Hoshino, "Detection of pinching gestures using a depth sensor and its application to 3D modeling," 2013 IEEE/SICE Int. Sympo. on System Integration (SII2013), TA2-K.1, pp. 814-819, 2013.

Name:
Kiyoshi Hoshino

Affiliation:
Professor, Graduate School of Systems and Information Engineering, University of Tsukuba

Address:
1-1-1 Tennodai, Tsukuba 305-8573, Japan

Brief Biographical History:
1993- Assistant Professor, Tokyo Medical and Dental University
1995- Associate Professor, University of the Ryukyus
2002- Associate Professor, University of Tsukuba
2008- Professor, University of Tsukuba
1998-2001 Senior Researcher of PRESTO project, Japan Science and Technology Agency (JST)
2002-2005 Project Leader of SORST project, JST

Main Works:
• "Hand Gesture Interface for Entertainment Games," R. Nakatsu, M. Rauterberg, and P. Ciancarini (Eds.), Handbook of Digital Games and Entertainment Technologies (ISBN: 978-981-4560-52-8), Springer, pp. 1-20, 2015.

Membership in Academic Societies:
• Robotics Society of Japan (RSJ)
• Institute of Electronics, Information and Communication Engineers (IEICE)
• Japanese Society for Medical and Biological Engineering (JSMBE)

Name:
Keita Hamamatsu

Affiliation:
Graduate School of Systems and Information Engineering, University of Tsukuba

Address:
1-1-1 Tennodai, Tsukuba 305-8573, Japan

Brief Biographical History:
2009-2013 Undergraduate Student, University of Tsukuba
2013-2015 Master Candidate, University of Tsukuba
2015- Nintendo Co., Ltd.

Main Works:
• "Detection of Pinching Gestures Using a Depth Sensor and Its Application to 3D Modeling," 2013 IEEE/SICE Int. Symposium on System Integration (SII2013), TA2-K.1, pp. 814-819, 2013.