Measuring Grasp Posture Using an Embedded Camera

Naoaki Kashiwagi[1], Yuta Sugiura[1], Natsuki Miyata[2], Mitsunori Tada[2],
Maki Sugimoto[1], Hideo Saito[1]

[1]Keio University
kasiwagi@hvrl.ics.keio.ac.jp
{sugiura, maki.sugimoto, hs}@keio.jp

[2]National Institute of Advanced Industrial Science and Technology
{n.miyata, m.tada}@aist.go.jp

Abstract

This paper proposes a measurement system for grasping postures using a fisheye camera. We attached a fisheye camera to the cap of a cylindrical object, such as a jar or a bottle. The fisheye camera was used to determine the position of the fingertips in 3D via image processing. The grasping posture utilized when opening or closing a jar or a bottle was then reconstructed. If an object, a model of a hand and the grasp type are given, it is possible to estimate the grasping posture of the whole hand, even using partially captured data. Preliminary experimental results show that the system is able to reconstruct the grasping posture of various users.

Figure1: Proposed method

1. Introduction

Measurement of the human grasping posture is required in many fields, such as human–computer interactions and ergonomics. Such measurements are utilized to enable researchers to understand how human beings use daily objects. The results are then employed in refining the design of products and services. In this paper, we focus on how people open cylindrical objects such as jars or bottles. Measurement of the motion to open cylindrical objects in daily life can be utilized to check the validity of the design of such products.

Today, there are many methods available for measuring the human grasping posture. One popular way involves fixing sensor equipment in the environment. The LeapMotion product [1], for example, captures hand postures via infrared LEDs and monochromatic IR cameras. However, it does not work well when measuring object grasping because part of the hand is obstructed by the object. Vicon [2] is an optical-motion capture system that uses reflective markers attached to the hand. Marsico et al. [3] and Wang et al. [4] use a data glove that is designed to measure hand postures. While these methods are effective to avoid occlusion problem, the unnatural sensations caused by the markers and the glove often lead to the subject displaying different behavior than usual.

In this paper, we have developed a system for the observation of a bare hand in a grasping posture. The method involves the use of computer vision – more specifically, a fisheye camera is attached to the cap of the object being grasped. Figure 1 shows our proposed method. The left-hand side of the figure depicts the system and the right-hand side displays a digital hand model constructed with the system. Computer vision was used to capture elements such as skin color and polar coordinates, which enabled the fingers to be detected via the fisheye camera. If a cylindrical object shape is given and the camera disposition is defined with respect to the model, it is possible to calculate 3D fingertip coordinates from the 2D fingertip coordinates detected in the image. If the type of grasp is known, then the whole 3D hand posture can be estimated using the fingertip position.

2. Related work

As for measurement devices for the bare hand, two types of body-worn devices have been proposed. The first estimates finger posture by sensing the shape (or other properties) of the wrist, the dorsal side of the hand, and the forearm [5, 6, 7]. The second type of device, which is called Digits [8], directly measures hand posture by using an

978-1-5090-4941-7/17 $31.00 © 2017 IEEE
DOI 10.1109/WACVW.2017.14

infrared laser line projector and an infrared camera.

The vision-based method is also administered for measurement of the human grasping posture. HandyAR [9] uses computer vision for skin-color detection or hand segmentation, allowing human hand posture to be estimated. The CyclopsRing [10] takes a direct image of a hand using a small fisheye camera worn on the finger. The hand posture is then determined via machine learning. It is difficult, however, to detect arbitrary grasping postures through vision-based methods due to the difficulty of separating the hand region from mixed images of both the object and the hand.

Attaching a sensor device to the object is another option when investigating bare-hand interaction with an object. PrintSense [11] is a capacitive sensing system that analyzes interactions such as touch, proximity, pressure, and folding using flexible, printed conductive electrode arrays. The SensorTape [12] is a modular, dense sensor network in the form of a tape. Each module senses its orientation via an Inertial Measurement Unit (IMU), as well as determining proximity through time-of-flight infrared. However, neither device allows the user's hand to touch the object surface directly. Touch & Activate [13] employs a vibration speaker to determine changes in acoustic signals. While this device allows for direct touch of the object by the hand, it basically "identifies" previously learned patterns, which limits the number of capturable grasping postures.

Wrap & Sense [14] is a bare-hand grasp observation system that uses band-type sensing equipment consisting of infrared distance sensors placed in an array. Assuming that the type of grasp is a "power grasp," the whole hand posture can be determined, if the 3D shape of the object is known. However, it is difficult to measure the motion when the hand's side edge cannot be detected, where motions like opening and closing jars are concerned.

3. Our method

Our system aims to observe bare-hand grasping posture by attaching a fisheye camera to the cap of a cylindrical object. To capture a complete image, each fingertip is detected via vision-based methods. Once the type of grasp has been determined, 3D fingertip coordinates are calculated from the 2D fingertip coordinates. By fitting a digital hand model to the fingertip coordinates, the whole hand posture is estimated. Figure 3 shows the workflow for proposed system.

3.1. Hand segmentation

First, a grasping image is captured using a fisheye camera. Then, skin-color segmentation is used, and every pixel in the image is categorized as either a skin-color pixel or a non-skin-color pixel. It can be assumed that the majority of

Figure 2: Fisheye camera

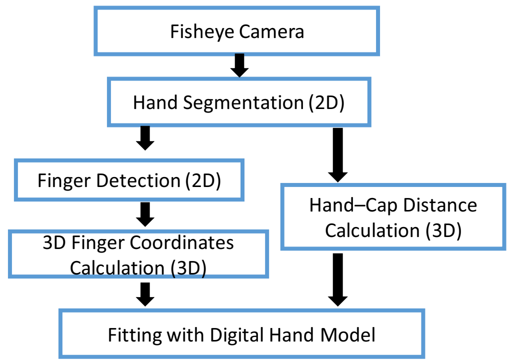


Figure 3: Workflow for proposed system

the skin-color portion of the image will be the hand region, which is the largest section. We retrieve the point exhibiting the maximum distance value from the Distance Transformation of the segmented image. The segmentations results are given in Figure 4.

3.2. Fingertip detection

The fingertips are detected from the contours of the hand via image processing. First, we use log-polar transformation on the segmented image of the hand. To determine the center of the log-polar transformation, we use the point representing the maximum distance value from the Distance Transformation. The results are shown on the left-hand side of Figure 5. The vertical axis gives the distance from the center to the contour and the horizontal axis gives the angle. Next, we check the pixels in the log-polar image from top left to top right. If consecutive white pixels are found, we check whether the neighbor pixels on the previous line are white or not. If the pixels on the previous line are white, we give the pixels on the current line the same ID as those on the previous line. Where there are no white pixels on the previous line, a new ID is given to the current line. The right-hand side of Figure 5 shows the segmented image. In each blob, the middle pixels at the top of the line can be defined as the fingertips.

Once five points have been detected as the fingertips in polar coordinates, these points are converted to Cartesian

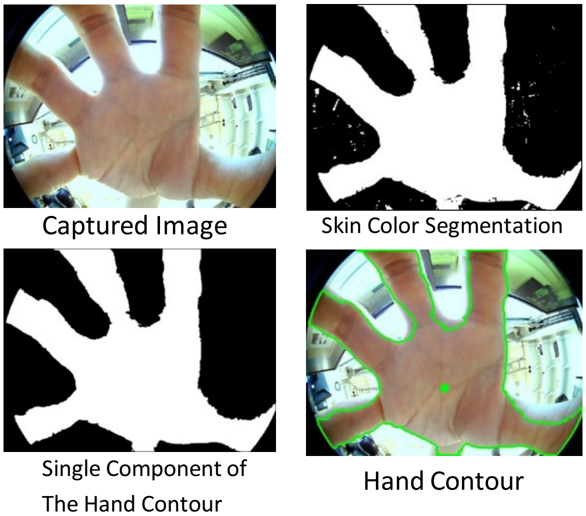


Figure 4: Hand segmentation

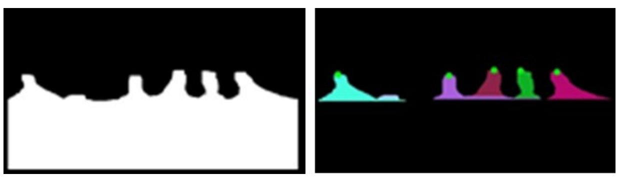


Figure 5: Log-polar transformation

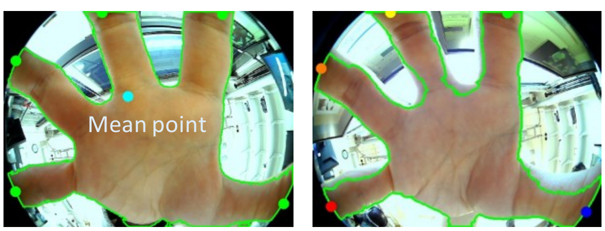


Figure 6: Finger detection

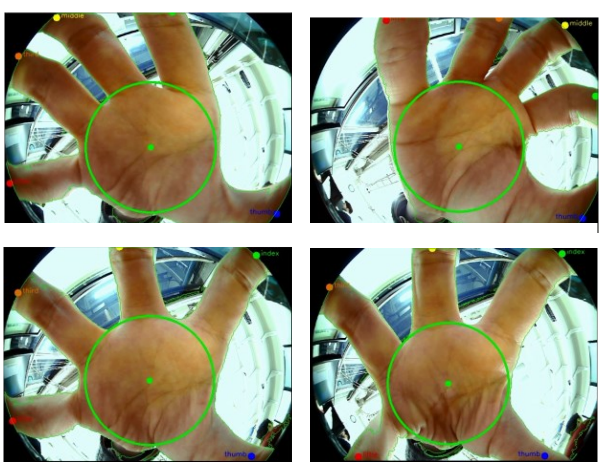


Figure 7: Finger detection when a jar is grasped in various ways

coordinates. The fingertips are then placed in order, based on the position of the index of the thumb, which can be labeled as the fingertip farthest from the mean position of all the fingertips (as shown in Figure 6). Figure 7 shows each fingers are detected correctly as a jar is grasped in various way. However, in current system, it is difficult to find each fingers correctly when fingers are bunched together or only a few fingers are used.

3.3. Calculating 3D hand coordinates

From the 2D fingertip coordinates, angle θ is calculated. It is calculated from the middle of the image, as shown on the left of Figure 8. As fisheye camera is attached to the center of the cap of the object, the 3D fingertip coordinates are determined from θ and the radius of the cap. Figure 8 shows its overview.

Hand–cap distance is calculated from the radius of a circle inscribed on the contours of the hand. We measured the correlation between the radius and the distance from the object to the hand, determining the hand–cap distance from the radius of the circle.

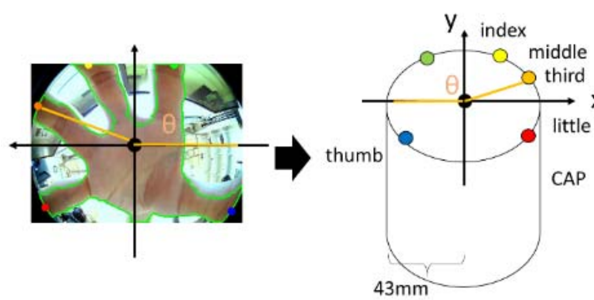
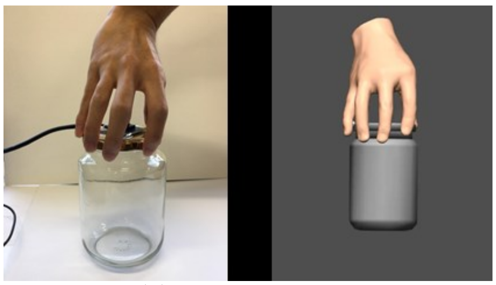


Figure 8: Determining 3D fingertip coordinates

Figure 9: Digital hand models with feature points

(a) View from left



(b) View from front

(c) View from back

Figure 10: Reconstructed digital hand models

Figure 11: Simultaneous measurement experiment conducted using the proposed system and an optical motion-capture system

Figure 12: 28 markers

3.4. Reconstructing Hand Grasping Posture

The hand's grasping posture is reconstructed using 3D fingertip coordinates, the hand–cap distance, and the digital hand model.

We use DhaibaWorks [15], a platform software application developed for assisting product design for digital human models. A model of a subject's hand can be created by manipulating a generic hand model in a way that satisfies the hand dimensions required [16]. We can estimate a full set of hand dimensions from a few hand measurements using regression analysis derived from our hand dimension database of more than 500 Japanese adults. In this research, we use four hand measurements (hand length, hand width, middle finger width, and thickness).

Six feature points are attached to the hand model, as shown in Figure 9. The grasping posture is determined, by minimizing the position error between the calculated 3D hand position and the feature points. The results are shown in Figure 10.

4. Comparison with motion-capture system

To check the accuracy of the reconstructed postures created using the system described above, the grasping postures of three participants (one man and two women) were captured simultaneously via the proposed system and an optical motion-capture system, as shown in Figure 11.

The motion tracker system used was the Vicon MX [2]. The postures captured by the motion-capture system were reconstructed using DhaibaWorks.

A comparison of the resulting reconstructed postures is given in Figure 13. The system can reconstruct the grasping posture approximately, but there are some shortcomings, especially concerning the curve of the fingers, because our system detects the fingertips only – it is difficult to detect each joint with CV technique.

To check the accuracy of fingertip points, we compared the fingertip position determined via the proposed system with those captured via an optical motion-capture system. The results are shown in Figure 14. The fingertip positions determined via our system are approximately the same as those given by the optical motion-capture system.

The mean distances between the finger position shown with the optical motion tracker and that given with the proposed system were 4.46 mm ±2.54 (mean ±s.d.) for Participant 1, 5.65 mm ±3.00 for Participant 2, and 9.73 mm ±6.09 for Participant 3.

5. Application

The technique proposed here for capturing the hand's grasping posture can be applied in other scenarios. In this section, we present one of its other possible applications: use as part of a grip introduction for baseball players.

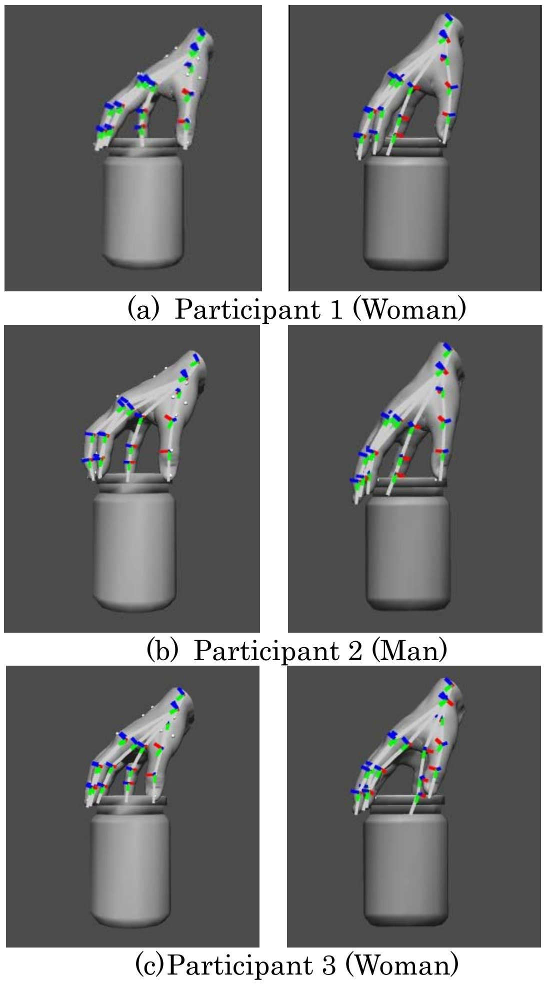


Figure 13: Comparison of reconstructed hand models generated via an optical motion-capture system and using the proposed system
(Left: Motion-capture System Right: Proposed System)

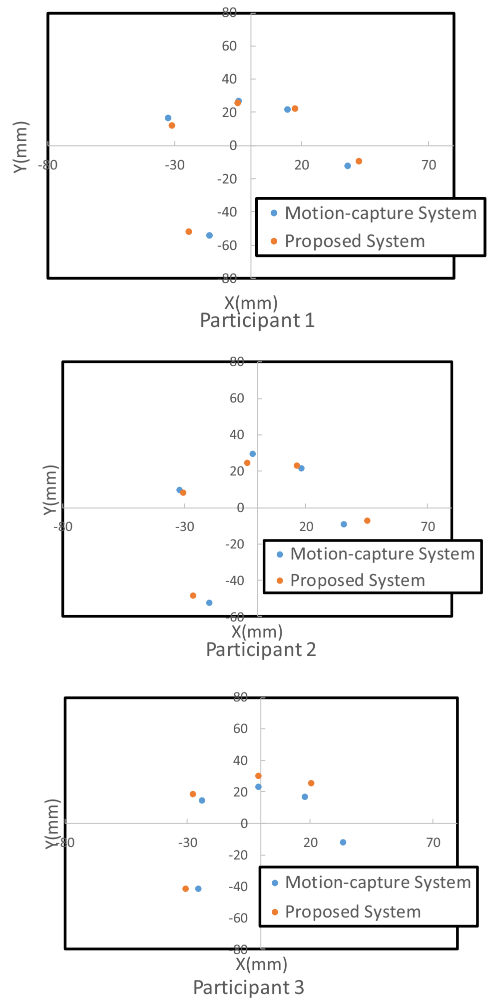


Figure 14: Comparison of fingertip positions determined via an optical motion-capture system and using the proposed system

Our system can be utilized to determine a baseball player's current hand posture, as shown in Figure 15. If one created an image of a goal hand (a certain kind of grasp in baseball), it could be used as a tool to guide a trainee in achieving the correct posture. Such visualizations could be useful for helping self-taught players to learn complicated grasp definitions. It would be a more interactive method than simply looking at 2D images of a hand. This approach could also be used to show how professional players perform various motions, if grasp data was provided.

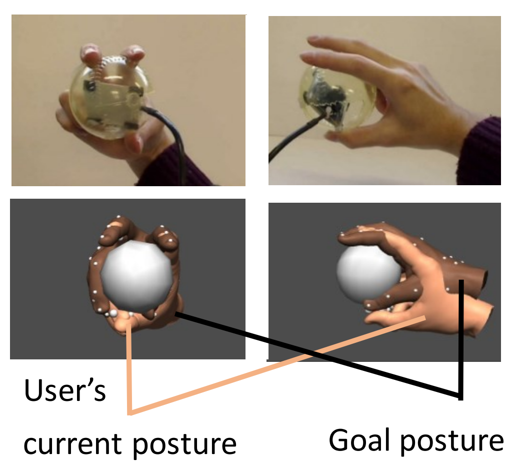


Figure 15: Reconstructing hand-grasp postures of baseball players

6. Public demonstration

We demonstrated the stability of the system through a demonstration in a public space at the middle of December 2016 (Figure 16). The system worked well for various male and female visitors.

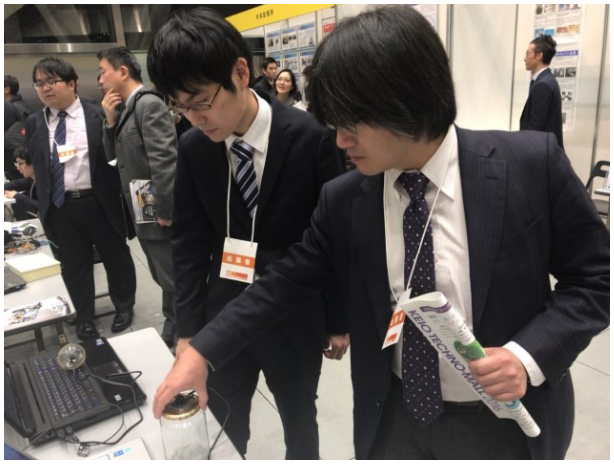


Figure 16: Public demonstration of proposed system

7. Limitations and future work

Grasp Type: While our current system detects grasping posture when fingers are widely open, it is difficult to estimate grasping posture if fingers are bunched together or only a few fingers are used. If we were able to categorize grasping postures via the images captured, our system would be able to analyze a wider range of postures.
Hand Size: Our current system cannot be used to detect the fingertips if the hand of the user is too small – the fingertips cannot be detect when the hand is too close to the lenses of the camera because it cannot detect 5 fingers clearly. This problem could be overcome by embedding a camera inside the jar or a bottle, instead of on top.
Other Objects: In this paper, we showed only baseball. To show how accurately and robustly the proposed method works in other cases, our approach needs more example of grasping other objects such as a doorknob or a bowl.

8. Conclusion

In this paper, we have proposed a grasp observation system for the bare hand. The system functions via a fisheye camera attached to the cap of a cylindrical object, which detects the fingers. The whole hand posture is then reconstructed by utilizing models of the hand and the object. Preliminary experimental results show that the system works stably to reconstruct the grasping posture of various users. Further discussion of an alternative application demonstrates the potential of the proposed system to be used in sports training.

Our future work include an extension of this approach to more various types of grasp and other objects such as a doorknob or a bowl.

Acknowledgments

This work was supported in part by JSPS Grant-in-Aid for Scientific Research(S) 24220004.

References

[1] Leap-Motion Homepage, https://www.leapmotion.com/, accessed Jan 2017.
[2] Vicon Homepage, http://www.vicon.com/, accessed Jan 2017.
[3] Marsico M., Levialdi S., Nappi M., and Ricciardi S. Figi: Floating Interface for Gesture-based Interaction. Journal of Ambient Intelligence and Humanized Computing (2012), 1–14.
[4] Wang R.Y., and Popovi, J. Real-time Hand-tracking with a Color Glove. ACM Transactions on Graphics, 28: 3 (2009).
[5] Lin J.W., Wang C., Huang Y. Y., Chou K.T., Chen H.Y., Tseng W.L., and Chen M.Y. BackHand: Sensing Hand Gestures via Back of the Hand. In Proc. UIST ’15, ACM (2015), 557–564.
[6] Yohsimoto S., Kawaguchi J., Imura M., and Oshiro O. Finger Motion Capture from Wrist-electrode Contact Resistance. In Proc. EMBC ’15 (2015), 3185–3188.
[7] Zhang Y., and Harrison C. Tomo: Wearable, Low-Cost, Electrical Impedance Tomography for Hand Gesture Recognition. In Proc. UIST ’15, ACM (2015), 167–173.
[8] Kim D., Hilliges O., Izadi S., Butler A., Chen J., Oikonomidis I., and Olivier P. Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor. In Proc. UIST ’12, ACM (2012), 167–176.
[9] Lee T., and Höllerer T. HandyAR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking. In Proc. ISWC ’07 (2007), 1–8.
[10] Chan L., Chen Y.L., Hsieh C.H., Liang R.H., Chen B.Y. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions through a Fisheye Ring. In Proc. UIST ’15, ACM (2015), 549–556.
[11] Gong N.W, Steimle J., Olberding S., Hodges S., Gillian N., Kawahara Y., and Paradiso J.A. PrintSense: A Versatile Sensing Technique to Support Multimodal Flexible Surface Interaction. In Proc. CHI ’14, ACM (2014), 1407–1410.
[12] Dementyev A., Kao H.L., Paradiso J.A. SensorTape: Modular and Programmable 3D-Aware Dense Sensor Network on a Tape. In Proc. UIST ’15, ACM (2015), 649–658.
[13] Ono M., Shizuki B., and Tanaka J. Touch & Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing. In Proc. UIST ’13, ACM (2013), 31–40.
[14] Miyata N., Honoki T., Maeda Y., Endo Y., Tada M., and Sugiura Y., Wrap & Sense: Grasp Capture by a Band Sensor. In Proc. UIST ’16 Adjunct, ACM (2016), 87–89.
[15] DhaibaWorks, http://www.dhaibaworks.com/, accessed Jan 2017.
[16] Miyata N., Shimizu Y., Motoki Y., Maeda Y., Mochimaru M., Hand MoCap Using an Individual Model with a Skeleton and a Surface Skin. International Journal of Human Factors Modelling and Simulation, 3:2 (2012), 147–168.