2016 IEEE International Conference on Automation Science and Engineering (CASE)
Fort Worth, TX, USA, August 21-24, 2016

Tracking and Recognition of a Human Hand in Dynamic Motion for Janken (rock–paper–scissors) Robot

Koichiro Ito†, Tomohiro Sueishi†, Yuji Yamakawa†, and Masatoshi Ishikawa†

Abstract— **In this paper, we focus on a human hand recognition system for human–robot interaction. The motion of the human hand is so dynamic that the hand needs to be tracked at high speed, and non-blurred images are required to achieve stable and correct recognition. Moreover, a wide recognition area is necessary for natural interaction. To achieve these aims, we constructed a new active sensing system that can actively track and recognize the human hand by using a high-speed vision system. An application that clearly requires high-speed recognition and actuation is the rock–paper–scissors game between a human and a robot, and we focused on this task to demonstrate the proposed system. In this situation, we track a human hand at high speed and detect its sign (rock, paper, or scissors) correctly in every frame. Our experiments showed that the robot hand formed its sign to beat the human opponent in natural interaction thanks to its high speed and wide recognition area.**

I. INTRODUCTION

In recent years, many studies have been conducted in the field of human–computer interaction (HCI) and human–robot Interaction (HRI). In these fields, robot systems recognize human motion in order to estimate human intentions and emotions, and the robot systems react appropriately according to the recognition result. Gestures are now widely adopted as inputs for many systems. In particular, the hand can convey a lot of meaning, and recognition of hand shapes and gestures has been widely studied [1], [2].

Hand gestures are generally recognized by using a glove-based method or by a vision-based method. In the glove-based method [3], the human user wears a glove on which many sensors are attached to measures the joint angles, movement, rotation etc. On the other hand, the vision-based method does not use any wearable devices and instead recognizes a bare hand using a camera. These days, many studies are being conducted utilizing Kinect sensors to detect hand gestures [4], [5], [6].

In this paper, we focus on two problems that are important in achieving natural interaction: the recognition area and latency.

First, regarding the problem of the recognition area, when we consider vision-based hand recognition, it is necessary to note that the human hand and fingers have highly dynamic characteristics. The hand movement area is so large that vision systems cannot completely capture the motion within a limited angle of view. Cameras are usually fixed, which results in a small recognition area and restricts the range of detectable human motion. If we consider using a wide-angle lens, there is a trade-off relation between the recognition area and the image resolution. Therefore, it is difficult to maintain sufficient resolution to recognize a hand with fixed cameras.

Second, latency, which humans perceive, results in poor operability. In general, humans can recognize latency of about 6 ms in continuous interaction [7] and 30 ms in impulsive events [8]. Latency does not matter in applications where humans simply click on a button. However, latency is critical in dynamic environments, for example, in tele-manipulation requiring immersion or video games requiring continuous input with gestures. In addition, current vision systems can only recognize a hand at a video frame rate of 30 fps, and are thus affected by blurring of the image. Therefore, an interaction system that recognizes hand gestures and operates an actuator in dynamic interactions requires a vision system having a high frame rate because stable input is required for system control, and latency should not be perceptible to humans.

We constructed a new active sensing system based on a high-speed tracking system [9], a specially designed tracking technique [10], and high-speed image processing for hand recognition in dynamic motion. Active sensing means that the system actively tracks and recognizes the target at the same time. Our system can always capture the hand at the center of the image, even when the human user swings his or her arm violently. We also introduce a fast sign (hand shape) detection algorithm for recognizing a hand in motion. As a result, we achieved

1) high-speed active hand tracking,
2) fast sign recognition,
3) and a combination of 1) and 2) in dynamic situations.

Moreover, to evaluate the performance our active sensing system, we applied to a Janken (rock-paper-scissors) robot [11] and controlled the robot's hand according to human hand signs. The existing Janken robot system can only recognize a human hand when it appears directly in front of the vision system, which restricts the motion of the human, thereby impairing the range of recognizable hand gestures. In our demonstration, we achieved a wide recognition area and high-speed response (recognition and motion) between the human and the robot, which allowed more natural interaction.

II. PROPOSED SYSTEM

Fig.1 shows a schematic diagram of the proposed system. Our system consists of a recognition system for tracking and recognition of a hand and an actuation system. The recognition system comprises a high-speed gaze direction

†K. Ito, T. Sueishi, Y. Yamakawa, and M. Ishikawa are with Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo, Tokyo, Japan koichiro_ito@ipc.i.u-tokyo.ac.jp.

978-1-5090-2409-4/16/$31.00 ©2016 IEEE

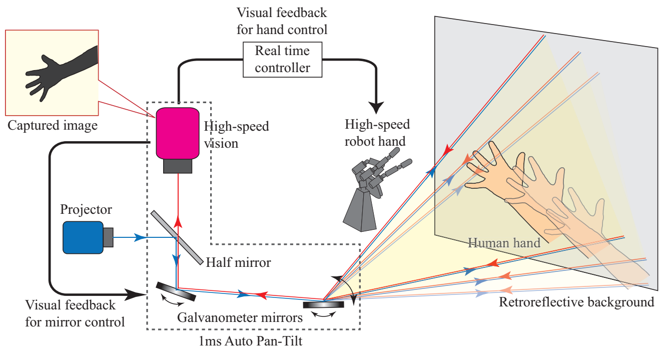


Fig. 1. The proposed system

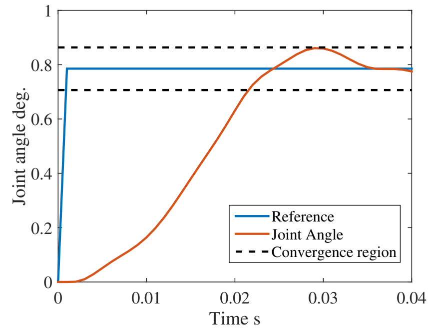


Fig. 2. Robot hand response: The blue plot is the reference joint angle. The red plot is the actual joint angle. The black plot is the convergence region, ± 10 percent around the reference.

controller (1 ms Auto Pan-Tilt) [9], a projector, and a retroreflective background. The actuation system comprises a high-speed, multi-fingered hand [12] and a real-time controller for controlling the robot hand motion.

The 1ms Auto Pan-Tilt tracks a human hand at high speed and captures non-blurred images of the hand to enable sign recognition. The robot hand is then controlled to display a sign that beats the human at such high speed that the human cannot recognize its motion. Our goal was to construct a system that can be actuated at high speed in response to a highly dynamic human gesture. We demonstrated stable tracking in hand recognition and the high-speed performance of the system through a rock–paper–scissors game between a human and a robot.

A. Recognition System

The dynamic behavior of a human hand in the rock–paper–scissors game is very fast, and a vision system needs to recognize its shape in motion without blurs. To achieve this, high-speed vision and active sensing are required.

We introduced a high-speed optical axis controller called 1 ms Auto Pan-Tilt [9] to achieve this. This system consists of a vision system with a high frame rate (500 fps), two small galvanometer mirrors, and pupil shift lenses. The galvanometer mirrors change the gaze direction by $60°$, and $40°$ change in angle can be completed in 3.5 ms. The pupil shift lenses enable a wide viewing angle even with the small mirrors and very quick changes in the gaze direction, resulting in a system that is suitable for human hand tracking and acquisition of non-blurry images in a wide recognition area.

When we apply the system to the tracking of an object, strong illumination is required to achieve high-speed stable tracking because of the short exposure time. However, such levels of illumination are not desirable for humans because they cause stress on the eyes. Moreover, the distance over which light can sufficiently illuminate an object is limited. We therefore introduce a robust high-speed tracking technique [10] using a projector and a retroreflective background. The optical axis of the vision system is set to that of the projector, which then illuminates the object at low lightning levels. The consequent reflection from the retroreflective background allows the system to track a human hand. The reflection setup can provide a larger recognizable distance than simple illumination.

B. Actuation System

The robot hand needs to complete its movement before the hand forms a sign, requiring a high-speed actuation system. In our research, we use a high-speed multi-fingered robot hand [12] with three fingers and a wrist which has two degrees of freedom (DOF). Each finger is divided into a root link and a top link, and each joint consists of a small harmonic drive gear® and a high-power mini-actuator. The motors are brushless and are highly responsive because the gears have no backlash. Consequently, the hand can close its joints at a rate of $180°/0.1$ s. This high responsiveness was also confirmed in our previous system [11], and 30 ms was needed to finish forming the signs (rock, paper, or scissors) after recognizing the human's hand sign. In this research, the actuation performance of the system is better owing to parameter tuning, and a sign-forming time of only 20 ms was needed, as shown in Fig.2.

III. Hand Tracking and Recognition

In this section, we introduce our overall strategy for achieving the rock–paper–scissors game between a human and a robot. First, the human hand is tracked at high speed and with high accuracy by the 1 ms Auto Pan-Tilt system. A high-speed algorithm is used to classify the hand sign as rock, paper, or scissors based on the premise that the vision system can consistently capture the human hand with high-speed robust tracking. Finally, we control the robot's motion in relation to the human's motion. The robot is controlled to form the sign that beats the human's sign based on the recognition result.

A. Preprocessing

The algorithm processes mainly binarized images for fast recognition. The image moment is used to derive the center position of the human hand and to classify the hand sign. A captured image is shown in Fig. 1, at the upper left, in which the shadow produced by the hand is represented by the black part, and the rest, shown as the white part, results from retroreflective illumination. We first binarize the image as follows:

$$f(i,j)=\begin{cases}0 & \text{if } \mathrm{src}(i,j)\geq k\\ 1 & \text{otherwise}\end{cases}, \tag{1}$$

where $\mathrm{src}(i,j)$ and $f(i,j)$ are input and binarized images, respectively, and (i,j) are the coordinates of the image. This results in an image in which the human hand is white. The retroreflective illumination saturates the background, and a sufficiently bright background provides high contrast. As a result, the noise level is small even with simple binarization. In general, a $(p+q)$-order moment of the binarized image $f(i,j)(=0 \text{ or } 1)$ is calculated by

$$m_{pq}=\sum_i\sum_j i^p j^q f(i,j), \tag{2}$$

In this case, m_{00} is an object's area. The center position of the object is derived as:

$$u_{\mathrm{g}}=m_{10}/m_{00},\ v_{\mathrm{g}}=m_{01}/m_{00} \tag{3}$$

The principal axis of inertia can be obtained by:

$$\theta=\frac{1}{2}\arctan\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right), \tag{4}$$

where μ_{pq} is a $(p+q)$-order moment around the center of gravity and is defined by the following equation:

$$\mu_{pq}=\sum_i\sum_j (i-u_{\mathrm{g}})^p(j-v_{g})^q f(i,j) \tag{5}$$

B. High-Speed Hand Tracking

We need to track the highly dynamic motion of the human hand. When forming a gesture, the human hand can change its shape in the course of the motion, and its center depends on the shape. Sueishi et al. [10] achieved a robust tracking technique that can be applied to a hand, although it is a deformable target. We briefly describe the method here.

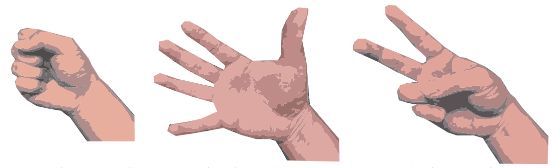
We need to capture the palm as a target at the center and to exclude the wrist. We set an adaptive window around the center of the image whose radius changes with S_{prev}, which is the size of an object in the window in the previous frame. The window is defined as follows:

$$B(u,v)=\begin{cases}0 & (f(u,v)>\alpha, (u-u_{\mathrm{g}})^2+(v-v_{\mathrm{g}})^2>R_{\mathrm{win}}^2)\\ 1 & (\text{otherwise})\end{cases} \tag{6}$$

R_{win} denotes the radius of the adaptive window and is expressed with a certain margin R_{margin} given by:

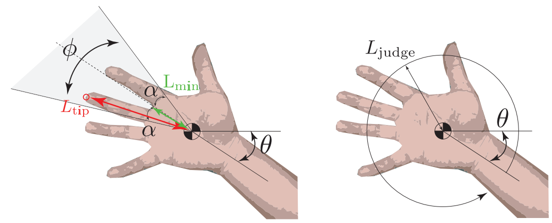
$$R_{\mathrm{win}}=\sqrt{\frac{S_{\mathrm{prev}}}{\pi}}+R_{\mathrm{margin}} \tag{7}$$

The center of gravity $\boldsymbol{u}_{\mathrm{g}}$ is calculated inside the window. Defining the center of the image as $\boldsymbol{u}_{\mathrm{c}}$, the two-axis small galvanometer mirrors are controlled to reduce the difference $\boldsymbol{u}_{\mathrm{g}}-\boldsymbol{u}_{\mathrm{c}}=\boldsymbol{0}$ by using high-speed visual feedback. Thus, the center of the hand is captured at the center of vision unless the shape of the hand changes dramatically. The tracking process is completed within 2 ms in every frame.



(a) Sign 1 (Rock) (b) Sign 2 (Paper) (c) Sign 3 (Scissors)

Fig. 3. Signs in the Japanese style.



(a) (b)

Fig. 4. Sign classification: (a) the range of the fingertip and its length; (b) the determination circle for classifying the sign.

C. Sign-classifying algorithm

Next, we need to develop a sign-classifying algorithm utilizing the captured images. Although the hand is always captured at the center of the image, the orientation changes frame-by-frame in response to arm motion. Thus, the algorithm needs to be robust against changing orientation. The Japanese rock–paper–scissors signs used in this research are shown in Fig. 3. We label "rock" as sign 1, "paper" as sign 2, and "scissors" as sign 3. In the Japanese rock–paper–scissors game, signs 2 and 3 can be classified by counting the number of extended fingers. Sign 1 is detected when the hand remains closed.

We count the fingers on a circle whose radius changes according to the human's fingertip position. This algorithm is robust against changing hand orientation because the fingers are counted from angle θ. Here, we present the proposed algorithm:

1) Detect the fingertip and measure the length L_{tip} between the center and the tip.
2) Judge whether the hand is open or not on the basis of the lengths L_{tip} and L_{min}.
3) Count the fingers on the circle when the hand is open.

We assume that the center of the hand is captured at the center of the image in the high-speed tracking. First, the hand contours are detected. Since the contours include the wrist near the elbow, the point farthest from the center cannot simply be treated as a fingertip. We make use of the angle of the principal axis of inertia, θ, because it inclines toward the wrist, and the fingertip is located opposite the wrist. As shown in Fig. 4, the fingertip is usually at $\theta+\pi-\alpha<\phi<\theta+\pi+\alpha$, and the farthest contour in the range can be regarded as the fingertip. The length from the center to the fingertip is defined as L_{tip}, and the minimum distance in the range is stored as L_{min}.

When the equation $L_{\mathrm{tip}}>\beta\times L_{\mathrm{min}}$ is satisfied, the hand is regarded as open, and the number of the fingers is counted with β as a threshold value. The circle with the radius of L_{judge} is shown in Fig. 4. Here, $L_{\mathrm{judge}}=L_{\mathrm{tip}}\times\gamma$, where

(a) Sign 1 (Rock) (b) Sign 2 (Paper) (c) Sign 3 (Scissors)

Fig. 5. Signs formed by the robot hand.

γ is a constant.

The fingers are counted on the circle, starting from θ in order to be robust against changing orientation. The number of extended fingers is counted when the pixel value of the image on the circle changes from 0 to 1. The counting works only when the pixel value "1" is consecutive in order to eliminate the effect of noise. Sign 2 is detected when the number of fingers is two, and sign 3 is detected when the number is 4 or more. The reason the threshold is set to four is to avoid classification failure when a thumb is not on the circle.

D. Robot Hand Control

The real-time controller receives the classification result and θ, and then controls the robot to form the sign necessary to beat the human opponent. Each sign is shaped by controlling motors in the joints. The step input of the reference joint angle is sent to each motor of the high-speed robot hand to form the signs, and the joint angles are controlled by proportional and derivative (PD) control. Joint angle control is conducted every 1 ms. Since the robot hand in our system has only three fingers, we define each sign as shown in Fig. 5.

In Section V, we demonstrate that the robot forms its sign just when the human does. Due to the 1 ms Auto Pan-Tilt system, the classification algorithm can detect a sign even before the arm of the human subject swing downwards. In order to synchronize the robot with the human's motion, it is necessary to prevent the robot hand from forming its sign before the human does. The angle of the principal axis of inertia θ is the value that changes with the arm's swinging motion and can be assumed to reflect the angle of the arm in a rock–paper–scissors gesture. We assume that the sign is formed by the hand when the arm is swinging downwards, and we make use of the angle θ to detect the moment at which the human hand forms its sign and the robot is allowed to form its sign. We define the angle θ_{thresh} as the threshold value for detecting the timing at which the human forms a sign and allow the robot to produce its the sign only when the angle θ satisfies the equation $\theta > \theta_{\mathrm{thresh}}$.

IV. Experimental Evaluation

We conducted experiments to evaluate hand tracking performance and the computational speed of the sign-classifying algorithm.

A. Experimental System

For the high-speed vision system, we employed a Photron IDP-Express R2000 high-speed camera which can capture a 512×512 pixel Bayer image at 500 fps. An EPSON EH-TW7200 (with a resolution of $1,920 \times 1,080$) projector was used as illumination. The vision system and projector were set in the 1ms Auto Pan-Tilt system so that the optical axis of the vision system and that of the projector matched. The controller PC of the tracking system was a DELL PRECISION T7610 [Windows 7 64-bit, Intel Xeon CPU 2.60 GHz (2 processors) with 32 GB RAM]. The retroreflective background sheet was composed of Ref-lite 9301. The robot hand was described in Section II-B. It included a real-time controller (dSPACE) whose sampling rate was 1 ms. The robot hand could complete formation of a sign approximately 20 ms after the controller received the reference joint angles.

As shown in Fig. 1. The retroreflective background was hung behind the viewing scene, and the robot hand was located 2 m away from the 1 ms Auto Pan-Tilt system. A human participant also stood 2 m away from the vision system and swung his or her hand to allow the vision system to recognize the shape of the palm. The subject then formed a sign in front of the robot. The angle θ and the human hand sign detected by the vision system were sent to the real-time controller. In the experiment, we evaluated only the recognition part; therefore, the robot hand was removed. The robot hand was used in Section V. In order to evaluate tracking and recognition performance, the retroreflective markers were attached to the human hand. We used data acquired online and in the captured video. The video was used to analyze the performance in offline processing.

B. Hand Tracking Performance

We initially evaluated the hand tracking performance. We manually set $k = 200$, $\alpha = \pi/9$, $\beta = 0.7$, $\gamma = 2.0$, and $R_{\mathrm{margin}} = 12$. In the experiment, the participant's hand was swung upwards and downwards at high speed, maintaining the sign throughout the motion. The experiment was conducted for sign 2 (paper) which is the largest sign, in order to confirm that the vision system can fully capture the hand in the image without the lack of displayed fingers.

We evaluated the tracking performance in offline processing using the video captured by high-speed vision during the experiment. A marker was attached, in advance, to the back of the hand, as shown in Fig. 6(e). Fig. 6(b) shows the difference between the center of the image and the marker in the video as time proceeds. (The marker is used only for the evaluation and not for tracking.) The location of the marker in the image was varied depending on the wrist orientation and was not equal to the center of gravity in the image. However, tracking was achieved while keeping the difference below 22 pixels in each frame.

Fig. 6(c) shows the size of the margin as time proceeded, which shows that the hand could be fully captured in the image. The margins were defined as the minimum distance from each side of the image to the hand, as in Fig. 6(h). Here, M_{left} is the margin from the left end of the image to the hand. The vertical margin was affected by the arm orientation in the image and was therefore defined using θ

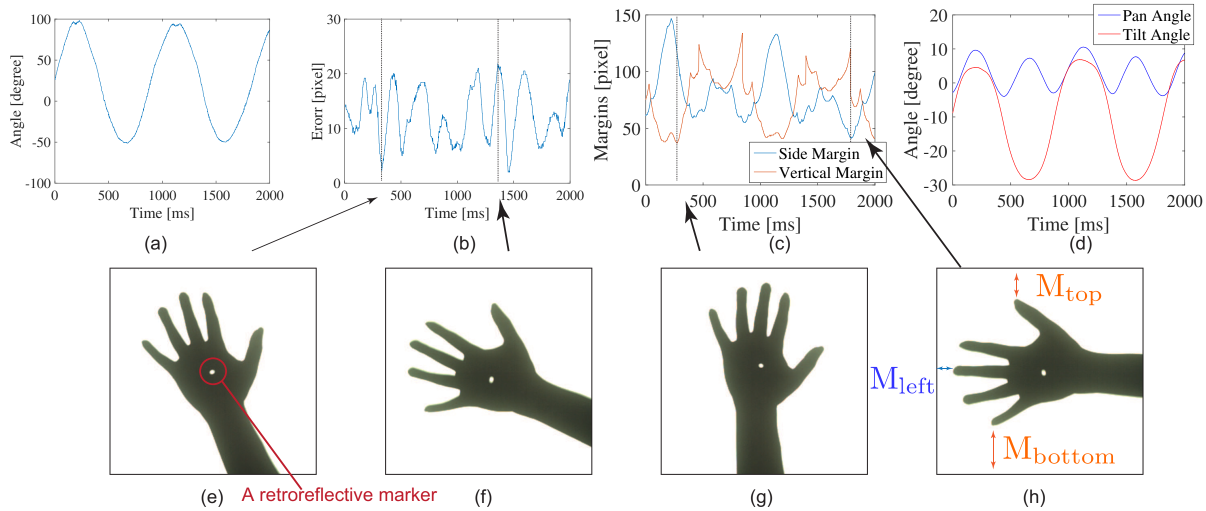


Fig. 6. Tracking evaluation: the sampling rate was 2 ms. The image resolution was 512 × 512 pixels. (a) shows θ as time proceeds. The domain is modified to avoid discontinuity. This figure indicates that the subject swung his/her arm upwards and downwards three times in 2 s. (b) is the difference between the center and markers in the image. (c) shows horizontal and vertical margins, respectively. (d) shows the angles of the two galvanometer mirrors. The captured images are also shown below. (e)-(g) in sequence are the captured images at 330 ms, 1358 ms, 272 ms, and 1792 ms in the video.

as follows:

$$M_{\text{vertical}} = \begin{cases} M_{\text{top}} & \theta > 0 \\ M_{\text{bottom}} & (\text{otherwise}) \end{cases} \tag{8}$$

The minimum margin shown in Fig. 6(c) is 40 pixels and the maximum error in (b) is under 22 pixels. These results show that the tracking is successful enough for the hand to be detected correctly without a lack of displayed fingers in an image.

As shown in Fig. 6(d), our system could track the hand in highly dynamic motion. The galvanometer mirrors were actuated according to the motion, and the hand was captured at the center of the image. This contributed to successful hand tracking and image capturing in rapid motion. The tilt mirror could be actuated to a maximum angle of about $-30°$. The 1 ms Auto Pan-Tilt system has been shown to be capable of tracking an object through a $60°$ angle of view [9], allowing the system to track more dynamic motion. The non-blurry images in Fig. 6 were captured owing to the high-speed vision system, which also resulted in stable recognition.

C. Validity and Computational Cost of Sign Classification

In this experiment, the human user moved his/her hand at random, and changed its sign at random. During sign recognition processing, we converted the 512 × 512-pixel Bayer images into 128 × 128 images in order to reduce the computational cost. Markers were attached to each fingernail in order to evaluate the classification. The number of markers detected was assumed to be equal to the number of fingers extended. We defined the fingers as being extended when the markers were detectable. The classification results and the number of markers were compared in offline processing using video captured during the experiment.

Fig. 7 shows the number of markers and the classification results as time proceeded. We defined sign 1 as the hand shape observed when the fingers were all closed, sign 2 as the hand shape observed when two fingers were detected, and sign 3 as the hand shape observed when four fingers were detected, as described in Section. III-C. The figure shows that classification (blue plot) worked successfully as the number of extended fingers (red plot) changed. The classification could detect the signs more quickly compared with just counting the fingernails.

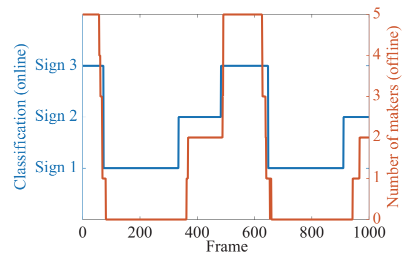


Fig. 7. Evaluation of sign classification: The sampling rate was 2 ms. The classified result (online) is shown in the blue plot, and the number of markers is shown in the red plot (offline). The number of markers corresponded to the number of fingers in opening motion.

Fig. 8 shows the computational cost of the algorithm for each sign. Signs 2 and 3 took slightly longer because the hand was open and the fingers were counted. The algorithm judged the signs while the hand was being opened; therefore, the radius changed in each frame, which explains the error bar. However, the recognition time of less than 0.25 ms shows that we achieved low-cost classification.

V. Demonstration

The movie [13] shows the human and the robot hand playing the rock–paper–scissors game. It can be seen that the robot formed its sign simultaneously with the human and won the game in every trial, demonstrating that the system accurately tracked the human hand and detected its sign.

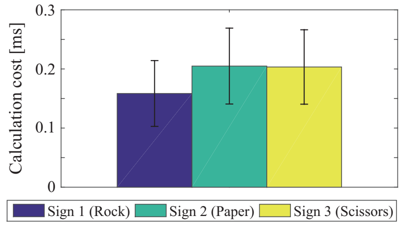


Fig. 8. The calculation cost of the classification: the average calculation cost of sign 1 was 0.16 ± 0.06 ms, that of sign 2 was 0.20 ± 0.06 ms, and that of sign 3 was 0.20 ± 0.06 ms.

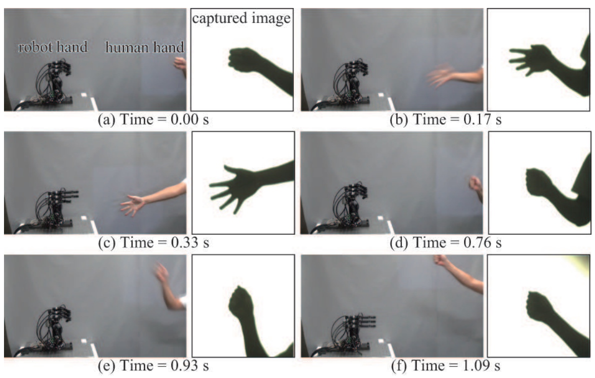


Fig. 9. Wide recognition area of the system: the left images were captured by a video camera (30 fps), and the right images by the high-speed camera. (a) is the scene when the robot hand was waiting, and (b) shows scene when the human hand was about to form a sign. The robot detected the sign and formed one that beat the human's sign in (c). The human pulled the hand away for the next game in (d), and was about to form a sign at a different place in (e). However, the system tracked the human hand and formed the sign to beat the human, as shown in (f).

In the demonstration, the robot hand usually waited for the control signals, forming its hand into the "rock" sign. The real-time controller instructed the robot to form its sign when θ showed that the human began to form the sign. The movie also shows that the robot formed its sign when the human swung his or her arm down and drew back the sign when the human swung his or her arm up according to θ. The game was played dynamically, and low latency was demonstrated in the last part of the movie.

Fig. 9 shows the human forming signs in various orientations and the robot correctly reacting to the signs. In this demonstration, the robot was controlled to show its sign when the human hand approached the robot. The system was able to track the dynamic human hand motion in every frame.

VI. Conclusions

In the work described in this paper, we achieved a system for natural interaction between a human and a robot with low latency and a wide recognition area. Our contribution was to develop an active sensing system based on high-speed hand tracking and classification. In the experiment, we showed that the system was able to capture a hand at the center of the image even when the human user swung his or her hand violently. Fast classification was also confirmed through the experiment.

As an application that requires high-speed performance, we applied our active sensing system to a Janken (rock–paper–scissors) robot system. In the demonstration, our system actively tracked and detected the human hand at the same time, and the robot was able to form its sign to beat the opponent with low latency. Our system took 2 ms for image processing. Taking the 20 ms for robot actuation into account, the response time was shorter than 30 ms [8], which is the minimum time resolution of the human eye. Therefore, the human user could not notice the latency, and the robot was able to react according to the human's sign, achieving natural interaction.

In this paper, we proposed a classification algorithm specially designed for the rock–paper–scissors game. Many studies on hand recognition have been conducted [1], [2], and in future work, we plan to combine these other studies with our system. In addition, we will also apply active sensing to tele-manipulation [14], which requires immersion and other HRI interfaces.

References

[1] A.R. Sarkar, G. Sanyal, and S. Majumder. Hand gesture recognition systems: A survey. *International Journal of Computer Applications*, 71(15):25–37, 2013.

[2] S.S. Rautaray and A. Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artifical Intelligence Review*, 43(1):1–54, 2015.

[3] Ng.Y.Y. Kevin, S. Ranganath, and D. Ghosh. Trajectory modeling in gesture recognition using cybergloves and magnetic trackers. In *Proc. 2004 IEEE The Institute of Electrical and Electronics Engineers Region 10*, pages 571–574, 2004.

[4] M. Van den Bergh, D. Carton, R. De Nijs, N. Mitsou, K. Kuehnlenz C. Landsiedel, D. Wollherr, L. Van Gool, and M. Buss. Real-time 3d hand gesture interaction with a robot for understanding directions from humans. In *RO-MAN*, pages 357–362, 2011.

[5] H. Wu, M. Su, S. Chen, Y. Guan, H. Zhang, and G. Liu. Kinect-based robotic manipulation: From human hand to end-effector. In *Proc. 2015 IEEE Industrial Electronics and Applications*, pages 806–811, 2015.

[6] V. Frati and D. Prattichizzo. Using kinect for hand tracking and rendering in wearable haptics. In *Proc. 2011 IEEE World Haptics Conference*, pages 317–321, 2011.

[7] A. NG, J. Lepinski, D. Wigdor, S. Sanders, and P. Dietz. Designing for low-latency direct-touch input. In *Proc. the 25th annual ACM symposium on User interface software and technology*, pages 453–464, 2012.

[8] W.H. Marshall, S. A. Talbot, and H. W. Ades. Cortical response of the anaesthesized cat to gross photic and electrical afferent stimulation. *Nerophysiology*, 6:1–15, 1943.

[9] K. Okumura, K. Yokoyama, H. Oku, and M. Ishikawa. 1 ms auto pan-tilt video shooting technology for objects in motion based on saccade mirror with background subtraction. *Advanced Robotics*, 29(7):457–468, 2015.

[10] T. Sueishi, H. Oku, and M. Ishikawa. Robust high-speed tracking against illumination changes for dynamic projection mapping. In *Proc. IEEE Virtual Reality Conference*, pages 97–104, 2015.

[11] Y. Katsuki, Y. Yamakawa, and M. Ishikawa. High-speed human / robot hand interaction system. In *Proc. 2015 ACM/IEEE Int. Conf. on Human-Robot Interaction*, pages 117–118, 2015.

[12] A. Namiki, Y. Imai, M. Ishikawa, and M. Kaneko. Development of a high-speed multifingered hand system and its application to catching. In *Proc. 2003 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 2666–2671, 2003.

[13] Ishikawa Watanabe Laboratory. Janken (rock-paper-scissors) robot with 100 % winning rate: 3rd version. `https://www.youtube.com/watch?v=Qb5UIPeFClM`.

[14] Y. Katsuki, Y. Yamakawa, Y. Watanabe, M. Ishikawa, and M. Shimojo. Super-low-latency telemanipulation using high-speed vision and high-speed multifingered robot hand. In *Proc. 2015 ACM/IEEE Int. Conf. on Human-Robot Interaction*, pages 45–46, 2015.