2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)
Beijing, China, November 6-9, 2018

Development of Movable Binocular High-Resolution Eye-Camera Unit for Humanoid and the Evaluation of Looking Around Fixation Control and Object Recognition

Tasuku Makabe[1], Kento Kawaharazuka[1], Kei Tsuzuki[1], Kentaro Wada[1], Shogo Makino[1], Masaya Kawamura[1], Ayaka Fujii[1], Moritaka Onitsuka[1], Yuki Asano[1], Kei Okada[1], Koji Kawasaki[2], Masayuki Inaba[1]

***Abstract*— Nowadays, studies have been conducted on humanoid robots with human mimetic structures. However, in the field of recognition, there is still much difference between humans and ordinary humanoid robots. For example, humans have movable eyes and use the degrees of freedom (DOF) of eyes in several ways: extension of the field of view, immediate changing of sight direction, focusing on objects to observe thoroughly. This DOF enables humans to adjust input images by aiming the direction of sight to desired objects, even when the body movement is limited. On the other hand, ordinary humanoid robots tend to have the cameras fixed to the head. Therefore, the humanoid has to move the head as a whole in order to change the sight direction of the camera.**

In this study, we developed both a movable binocular high-resolution eye-camera unit, which is small enough to be installed in the humanoid head, and a system to change input images according to the environment.

The developed unit contains cameras, actuators, and motor control boards, being a stand-alone unit that enables the use of the movable eye. A small high-resolution auto-focus camera is used for this eye-camera unit. The developed system is used to adjust the images to the environment, controlling the recognition area by changing the direction of sight, size of input images and resolution.

I. Introduction

Nowadays, a study has been conducted on the humanoid which has the human mimetic structure. Those humanoids are made for studying on the advantages of specific human body design or sensor mechanism. As preceding study, musculoskeltal humanoid "Kengoro" [1] has the tendon-driven human mimetic skeleton body and the force sense. However, in the field of recognition, there is the difference between human and an ordinary humanoid robot. For example, humans have movable eyes and use degrees of freedom of eyes in several ways: extension of the field of view, immediate changing of sight direction, fixation of objects to observe thoroughly. This DOF enables human to adjust input images by aiming the direction of sight to object even under a condition that the body movement is limited by some actions such as driving.

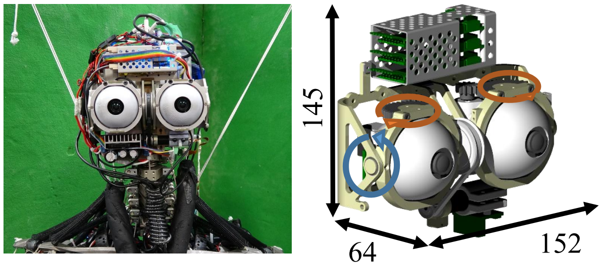


Fig. 1: Newly developed eye-camera unit.

However, cameras of the ordinary humanoid, whose functions are similar to that of human, are fixed to the head. The humanoid has to move the head to change the sight direction of the camera.

There are plenty proceeding studies on robots with movable eyes. Some examples are the Child sized robot with gear driven movable eye [2]; emotion expressing robot to change sight direction and facial expression [3]; upper body robot which has a tendon-driven single camera [4]; robot with binocular eyes which are equipped with two digital color cameras (wide and narrow-angle) [5]; and movable binocular camera robot for researching on neural learning of embodied interaction dynamics of the human eye [6].

However, there seems to be no life-sized humanoid robot with movable eyes which can carry on human-like movement while touching the environment positively. Under such situations, the humanoid movement and sight direction is often limited by narrow space or other environment fixations, such as the car driver being restricted by the seat belt.

In this study, we developed a movable binocular high-resolution eye-camera unit in Fig. 1, which is small enough to be installed in the head of the humanoid robot "Kengoro" [1], and system to change input images according to an environment. This unit contains cameras, actuators, and motor control boards and we can use movable eye only this single unit. A small high-resolution auto-focus camera is used for this eye-camera unit. In addition, to adjust image to necessity image, we developed the control system of recognition area to change the direction of sight, the size of input images and the resolution. We installed the developed eye-camera unit in the musculoskeletal humanoid Kengoro and achieved looking around motion as a movement of changing input images as needed.

As important comparison points of the looking around motion, we emphasize the wide range of the field of view and quick shift of the sight direction. Both of these are improved

[1] Authors are with Department of Mechano-Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. [makabe, kawaharazuka, tsuzuki, wada, makino, kawamura, a-fujii, onitsuka, asano, k-okada, inaba]@jsk.t.u-tokyo.ac.jp

[2] TOYOTA MOTOR CORPORATION[koji_kawasaki@mail.toyota.co.jp]

978-1-5386-7282-2/18/$31.00 ©2018 IEEE

by the movable eye unit, which enables quick eye movement, faster and more flexible than the neck movement.

In section I, we explained the background and problems of the recognition system for humanoid robots compared with that of the human and the importance of the looking around motion. In section II, we explain the design of the eye-camera unit that we developed. In section III, we explain the control method of the recognition area. In section IV, we explain the evaluation of the developed eye-camera unit and control method of the control area. In section V, we explain the experiments of the looking around motion by Kengoro. In section VI, we state the conclusion of this study and future works.

II. Three Degrees of Freedom of Movable Binocular Eye-camera Unit

A. Overview

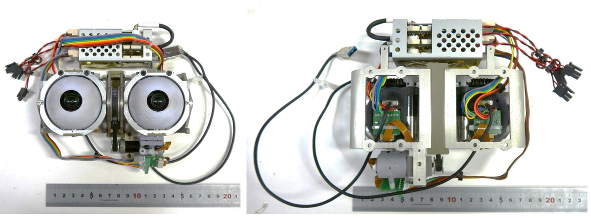


Fig. 2: Overview of eye-camera unit.

The newly developed movable binocular eye-camera unit is shown in Fig. 2. This unit is small enough to be contained in the humanoid head. Both eyes have common tilt axis and separated pan axis, allowing the sight to be changed in various directions. High-resolution and auto-focus cameras are used for the eye-cameras. By using such high-resolution cameras, it is possible to get images which are big enough to process even with cropped images. In addition, the changing-focus function makes it possible to focus again after moving sight direction and changing the fixation point.

B. Imitation of Degrees of Freedom of The Human Eye

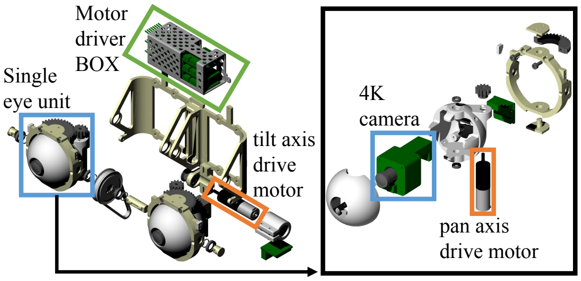


Fig. 3: The arrangement of camera and motor in unit.

TABLE I: The mechanical parameters of eye unit.

axis	Min [deg]	Max [deg]	resolution [deg]
tilt	-20 (downstream)	6 (upstream)	0.15
pan (both side)	-22 (abduction)	21 (abduction)	0.15

Human eyes have six muscles and corresponding DOF, allowing us to change the sight direction upstream, downstream, right side and left side. That DOF is used for recognition of environments in several ways. For example, we can recognize the position of objects by using binocular convergence. This is a movement of looking at one position using both eyes and occurs mainly when looking to a nearby point. When humans recognize the environment, information regarding the horizontal position of an object tends to be more important than the vertical position since we live on flat land. This is thought to be one reason for many animals, including humans, to have two eyes side by side. Those eyes move together changing sight direction upstream or downstream, and move independently changing sight direction right side or left side. To imitate human eye DOF, we need twelve actuators for driving tendon. However, it is difficult to imitate those many eye DOF since the size of the humanoid head is small. So we designed this eye-camera unit which has common tilt axis and separated pan axis for both eyes. Exploded view of the movable binocular eye-camera unit is shown in Fig. 3. This unit is composed of 4 links. Base link is fixed to the humanoid head and motor driver circuits are fixed to base-link. Carrier link is connected to base link via tilt axis and right single eye unit and left single eye unit are connected to carrier link via independent pan axis. Each axes are driven by EC brushless motor and rotate angle is measured by hall sensor of motor. The movable range of the eye-camera unit which is developed in this research and the resolution of measuring rotate angle are shown in Table I.

C. Selection of Camera

We used DFK-AFUJ003 camera and low-distortion megapixel mini-lens set (ImagingSource Inc.), which is shown in Fig. 4 for each eye in the movable binocular eye-camera unit. Main spec is shown in Table II.

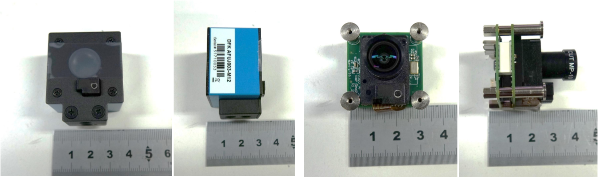


Fig. 4: The appearance of camera-lens set.

TABLE II: DFK-AFUJ003 & TBN 5.4C mini lens

size [mm]	35 × 35 × 40
sensor size [size, mm×mm]	1/2.3 inch (6.2 × 4.2)
resolution [pixel], frame rate [FPS]	3872 × 2764, 7 1920 × 1080, 30 640 × 480, 90
focus focus length [mm]	auto-focus & changeable 5.4
lens	low distortion mini lens
MAX horizontal FOV [deg] MAX vertical FOV [deg]	57.29° 44.33°
shutter	rolling shutter

We selected this camera and lens set for the following three reasons:

1) It has a high-resolution image sensor.
2) It can change the lens position.
3) The size of the camera is small.

At first, we can get the 3872 × 2764 pixel image for 57.29° horizontal field of view (FOV) and 44.33° vertical

FOV by using this high-resolution single camera. We can enlarge FOV using a wide-angle lens, but an image is distorted and resolution becomes low in the rid of the image. So, we used a low-distortion lens to get a less-distorted and high-resolution image in this research.

Next, the function of changing the lens position can be used to focus again after moving the sight direction. Since sight direction changes frequently, the distance between the point of fixation and the eye-camera unit changes, being necessary to focus again to make the image on the image sensor of the camera. Thus, we selected this camera to change the position of a focal point during robot motion according to embedded auto-focus function or control signal from a user and made it possible to get clear input image even when sight direction is changing.

Finally, small camera size is important for reducing the eye ball unit size and with that enlarging the movable range of the eye-camera. This camera is used because it has high-resolution, adjustable focus and lens position, and its size is extremely small (35mm × 35mm × 40mm).

III. Control of Recognition Area

In this research, we control the recognition area to recognize environments selectively by changing the input image from the camera for looking around motion. This is done by performing the control of modulating camera sight direction, range of vision and resolution, according to an input of the robot pose and visual information from the environment.

A. Overview of Control System of Recognition Area

Overview of the control system of recognition area is shown in Fig. 5.

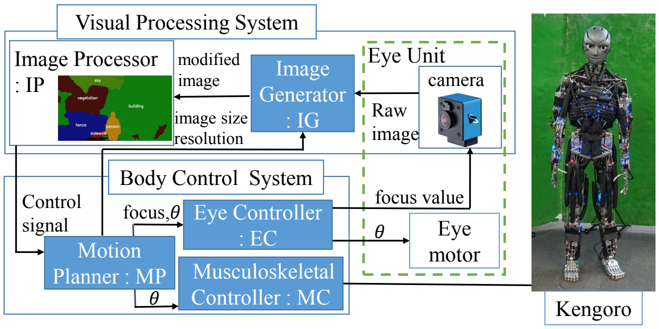


Fig. 5: Control system of recognition area.

At first, raw input image from environment is processed in a module called "Image Generator" (IG). The raw image is then cropped and the resolution of the image is changed in this module.

Next, the cropped image with modified resolution is then processed in "Image Processor" (IP) module with an existing image processing filter or detectors such as a color filter or detector using DNN. In "Motion Planner" (MP) module, many commands to several modules are generated according to an input of image processed result from the IP module and input of robot pose. For example, command of joint angle for "Musculoskeletal Controller" (MC) module which controls the whole robot body, command of sight direction angle and position of the lens for "Eye Controller" (EC) which controls the eye-camera unit and command of image crop size and resolution for IG module are generated in this module. Control system of recognition area is composed of the above modules. This system also has a feedback loop consisted of image-processing, driving sight direction and changing input image size and/or resolution.

B. Modifying Range of Vision (Image Generator: IG)

In this module, we imitated modifying the range of vision by cropping and modifying the resolution of the image using software. This function is implemented by image-proc package using ROS and we can use this function dynamically during robot motion.

C. Processing Input Image (Image Processor: IP)

This module is separated into two sections, an image processing section and interpreting the processed result section. Image processing section is composed of existing image processing filter or detector such as the color filter or segmentation detector using DNN and the processed image is sent to interpreting processed result section. Interpreting the processed result section has a function such as calculating the center of recognized object or ratio of specific object image area in the input image and those calculated values are inputted in MP.

D. Decision of Motion and Several Commands for Other Modules (Motion Planner: MP)

Motion Planner module is the most upper layer module generating commands for several modules (IG, EC, MC) from an input of IP module result or robot body pose.

E. Body Control Interface Module (Musculoskeletal Controller: MC, Eye Controller: EC)

These two modules are used for an interface of moving real robot body. MC module receives command of a joint angle from Motion Planner and controls the body by adjusting the wire (tendon) length based on an internal robot model. Those commands are passed to downstream layer module controlling the real robot in wire length level. EC module, which receives commands of sight direction angle and position of the lens, generates commands of a servo control for the eye-camera unit DOF to drive the motor and commands regarding the lens position for the camera. We can control the position of the lens by using the camera embedded auto-focus function or manual signals from users.

F. Explanation of Controllers

In this section, there are descriptions about controllers which are composed of IP module and MP module which accords to several tasks using control of recognition area method.

1) Normal Controller: This controller is implemented in MP module and interface of making commands for IG module and EC module and MC module from values predefined by the user. This controller enables switching sight direction, range and resolution of an image according to the state of the task.

2) Object Tracking Controller: This controller generates commands of sight joint angle for the EC module from results of object position in the input image in the IP module, being able to track the recognized object. Red ball, AR-marker or other objects can be used as objects to be tracked in this research. In the IP module, *hsv-color-filter* or *ar-track-alvar* ROS packages are used for image processing and calculation of the center of the recognized object. Those center positions are then used as a fixation point and sight direction is controlled to reconcile the fixation point with the center of the range of vision. This tracking system is controlled by P control, which is the simplest control scheme. When the fixation point coordinate is (x_r, y_r) in right camera-coordinate system and (x_l, y_l) in left camera coordinate system, both having input image size of $width \times height$, we define target convergence position height as $(x_{ref}, y_{ref}) = (\frac{width}{2}, \frac{height}{2})$ and P control gain vector which converts position difference into the joint angle as K_θ. Controlling equation is showed below.

$$\boldsymbol{r} = (x_{rref}, x_{lref}, y_{ref}) \tag{1}$$

$$\boldsymbol{y} = (x_r, x_l, \frac{y_r + y_l}{2}) \tag{2}$$

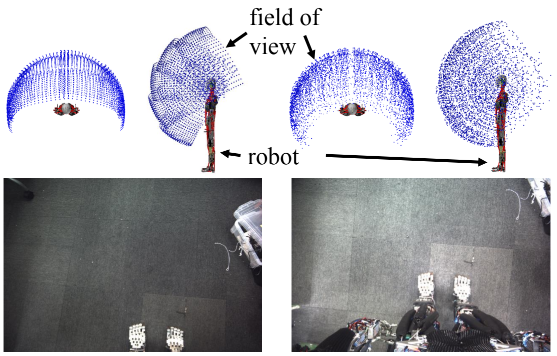
$$\boldsymbol{K}_\theta = (k_{\theta x_r}, k_{\theta x_l}, k_{\theta y}) \tag{3}$$

$$\boldsymbol{\theta} = \boldsymbol{K}_\theta(\boldsymbol{r} - \boldsymbol{y}) \tag{4}$$

IV. Evaluation

In this section, the evaluation of the movable binocular eye-camera unit and of the control system of the recognition area is shown.

A. Checking Field of View Range



only neck DOF   neck + eye DOF

Fig. 6: Difference of field of view using the neck and eye DOF.

Experiments were performed to check the field of view range was checked with both neck only and neck + eye DOF using Normal Controller. We simulated the field of view range on robot model using FOV calculated from camera spec and DOF of eye and neck. In this simulation, we made the field-of-view range visible by displaying blue dots 1 m away from the center of eyes. Comparison of a field of view range is shown in Fig. 6 between using only the neck DOF and using both the neck and eye DOF. Since tendon driven humanoid Kengoro's neck has multiple movable membersand wide movable range, it has wide field-of-view range. The additional eye DOF furtherly widens the field of view range and enables to look at the abdominal region, the side of the body and the back of the body. In an experiment, we checked the difference of FOV range between using eye DOF and not using it when looking downward. When using only the neck DOF, we could only look the tips of robot toes. However, using the eye DOF enabled to look regions closer to the robot trunk, such as the abdominal region.

B. Measurement of Moving Speed

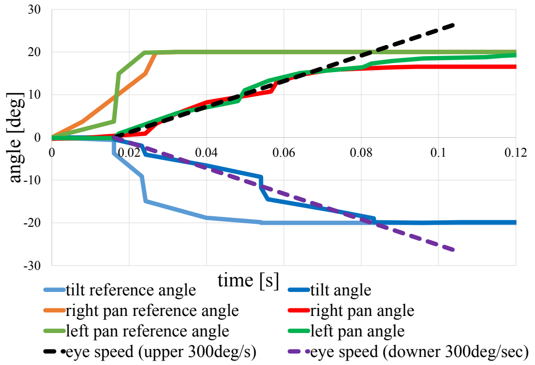


Fig. 7: The joint angle follow-up ability of eye-camera unit (tilt right-pan left-pan).

To test the joint angle follow-up ability of the eye-camera unit, we measured the moving angular speed of the unit using Normal Controller. The graph of Fig. 7 shows the joint angle follow-up ability of eye-camera unit from a step-like signal input of sight direction according to common tilt axis and right and left pan axis. Since only one motor drives common tilt axis and two motors drives separated pan axes of a binocular eye, we expected that joint angle follow-up ability of pan axis is better than that of tilt axis. Contrary to expectations, the angular velocity was about 300 [deg/s] for all axes and there was no difference between axes. In comparison, it has been observed that human eyes can move with angular speed of 440 [deg/sec] [7]. The developed eye-camera unit does not move as fast as the human eyes, but it can reach reasonable speed.

C. Fixation

In this experiment, at first the low-resolution input image of the entire field of view is segmented and then used to decide the fixation point. In the next experiment, sight direction is moved toward the fixation point and robot segments high-resolution input image of the cropped field around fixation point again. Kengoro changes sight direction and range of the field of view using the Normal Controller and the Object Tracking Controller. We used *pspnet* by Zhao [8] learned using cityscapes data sets as a detector of segmentation for calculation of the center of the area which is segmented as a person. The size and resolution of the image are defined by the user in a former experiment. The result of this experiment

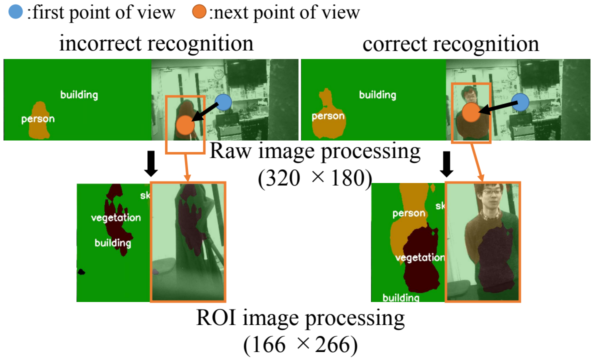


Fig. 8: Segmentation results of fixation experiment.

is shown in Fig. 8. In the left result, a tripod hanging a jacket was segmented as a person in a low-resolution image at first, but it was segmented not as a person after high-resolution image of fixation. In the right result, a real person was segmented as a person in both low-resolution image at first and in high-resolution image after fixation. This result shows fixation makes it possible to correct the identification error of unknown objects and change the sight direction, input image size and resolution accordingly.

D. Faraway detection

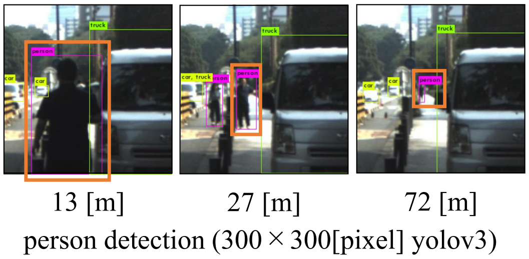


Fig. 9: Detection results of faraway segmentation.

We carried faraway detection to evaluation the performance of the high-resolution camera. Experiment results are shown in Fig. 9.

In this evaluation, we cropped an image from 3872 × 2764 pixel to 300 × 300 pixel to fixate and detected person in a small sized image using yolo v3 as a detector of person [9]. We focused on a distant place using camera embedded auto-focus function. With the above, we succeeded on detecting a person who is at a distance of about 72[m] from the camera. In this result, a cropped image has 300 × 300 pixel, but it is indistinct image since we could not focus for that range. So we can guess it is necessary to focus on minutely when fixating on a distant place using the processing result of a cropped image like contrast or edge of an image.

V. Experiment

TABLE III: Conditions of looking around, starting driving and stopping experiment.

image size [pixel]	field of view [deg]
640 × 360	56.9 × 35.3

In this section, experiment using both eye-camera DOF and the control system of recognition area for looking around motion is shown. A small electric vehicle was stopped on the roller which enabled the wheel of that rotating freely inside of the room and we took musculoskeletal humanoid Kengoro in that and Kengoro looked around the vehicle. We used *open-pose* by Cao [10] as a detector of person. The image from the left eye camera was used as an input image. The conditions of this experiment are shown in Table III.

A. Quick Sight Changing Experiment

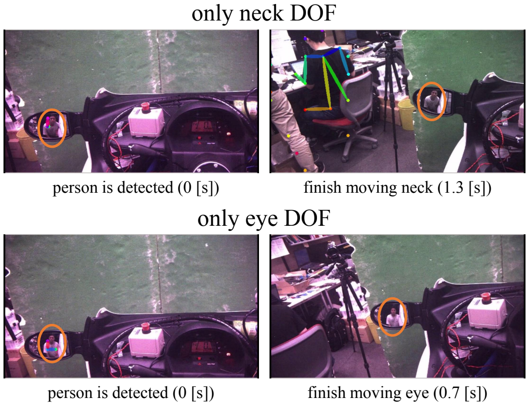

only neck DOF
person is detected (0 [s])
finish moving neck (1.3 [s])
only eye DOF
person is detected (0 [s])
finish moving eye (0.7 [s])

Fig. 10: The result of sight changing experiment.

In this experiment, we compared the time of movement of changing sight between that using neck DOF and that using eye DOF. Kengoro changed sight direction toward side mirror in the edge of sight when the person is detected in sight and we measured the time of that movement. The movement using only neck DOF took 1.3[sec] and that using the only eye DOF took 0.7[sec]. As it is shown in Fig. 10, using the DOF of eye enabled quick change of sight direction. Since the moment of inertia of robot head is bigger than that of a small eye, this result does not contradict.

B. Looking Around Driving Experiment

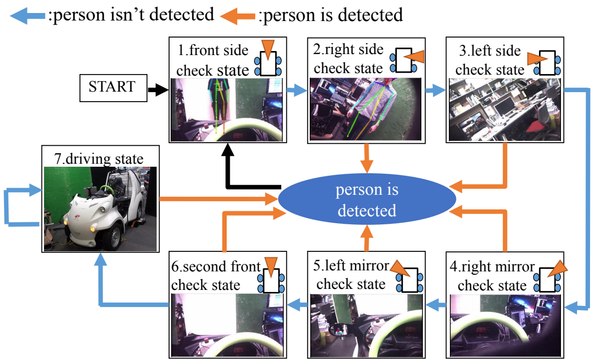
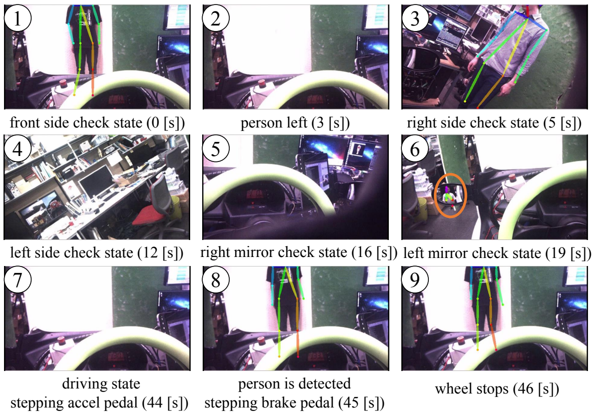


Fig. 11: The sequence of looking around detection.

We carried a driving experiment as one application of looking around motion. Kengoro sitting in the vehicle changed sight direction depending on the presence of a person in the sight. If there is no person during last sight direction, Kengoro finishes safety check around the vehicle, steps the accel pedal and the driving wheel starts rotating.



front side check state (0 [s]) person left (3 [s]) right side check state (5 [s])

left side check state (12 [s]) right mirror check state (16 [s]) left mirror check state (19 [s])

driving state stepping accel pedal (44 [s]) person is detected stepping brake pedal (45 [s]) wheel stops (46 [s])

Fig. 12: The result of looking around detection.

If a person is detected while driving, for instance when someone rushes out in front of the vehicle, Kengoro steps the brake pedal and stops the car. In this experiment, we set the image resolution to low for fast recognition, and set the sight direction for checking broad area as the control system of the recognition area. The sequence of this experiment is shown in Fig. 11. We defined five sight directions, front side, right side, left side, right mirror, left mirror. The result of this experiment is shown in Fig. 12.

The experiment started with detecting the person in front of the vehicle ① and Kengoro started looking around motion after the person left ②. Next, Kengoro detected person in right side ③ and moving small person in left mirror ⑥ with quick sight change and started looking around motion again from the beginning. Finally, Kengoro finished safety check and started driving ⑦, detected rushing out person ⑧, slammed the brakes on and the driving wheel stopped ⑨. In this experiment, we achieved a wide horizontal range of sight on the right and left side of the vehicle, quick sight change using only eye DOF and recognition of the person in all sight direction. From this result, we showed movable binocular eye-camera unit and the control system of recognition area enabled human-like recognition with a wide range recognizable area and quick sight direction changing.

VI. CONCLUSION

In this study, we developed the movable binocular eye-camera unit, which is small enough to be installed in the humanoid head, and the system to change the input image according to the environment. In addition, we achieved the looking around motion to adjust input image using that camera unit and control system.

We proposed three characteristics of developing the eye-camera unit and the control system of the recognition area.

- Small sized so that can be installed in the humanoid head
- Image with enough range of view and recognition and adjustable focus
- An integrated system of image processing and eye moving

The movable eye widened the range of the field of view and made it possible to look areas that could not be seen with only the neck, such as the abdominal region. By using the eye DOF, the range of the field of view was widened even during driving task, where body movement is limited by the seat belt. In addition, Kengoro achieved using side mirror for checking person presence in the looking around motion during driving.

In future works, there are two problems in order to improve this system. First, it is necessary to widen the movable range of eyes by using another actuating method such as a tendon-driven method. In this study, a movable range of eye is narrow compared with human eye due to intervention between eye carrier link and robot mask. To drive eye using tendon, intervention between eye carrier link and robot mask may decrease since tendon wire is slim and flexible.

Second, an invention of original three-dimensional reconstruction software is needed for this movable eye unit. In this research, we could not use ordinary reconstruction software, such as stereo matching, since the relative posture of the two camera dynamically changes during eye movement. It is essential to make new three-dimensional reconstruction software such as human binocular stereopsis, which uses images and angles from both eyes as input information from the environment.

REFERENCES

[1] Y. Asano, T. Kozuki, S. Ookubo, M. Kawamura, S. Nakashima, T. Katayama, Y. Iori, H. Toshinori, K. Kawaharazuka, S. Makino, Y. Kakiuchi, K. Okada, and M. Inaba, "Humanmimetic musculoskeletal humanoid kengoro toward real world physically interactive actions," in *Proceedings of the 2016 IEEE-RAS International Conference on Humanoid Robots*, 2016, pp. 876–883.

[2] R. Beira, M. Lopes, M. Praca, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Saltaren, "Design of the Robot-Cub (iCub) Head," in *Proceedings of The 2006 IEEE International Conference on Robotics and Automation*, 2006, pp. 94–100.

[3] T.Kishi, H.Futaki, G.Trovato, N.Endo, M.Destephe, S.Consentino, K.Hashimoto, and A.Takanishi, "Development of a Comic Mark Based Expressive Robotic Head Adapted to Japanese Cultural Background," in *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 2608–2613.

[4] H. G. Marques, M. Jantsch, and S. Wittmeier, "ECCE1:The first of a series of anthropomimetic musculoskeletal upper torsos," in *Proceedings of the 2010 IEEE-RAS International Conference on Humanoid Robots*, 2010, pp. 391–396.

[5] T. Asfour, J. Schill, H. Peters, C. Klas, J. Bücker, C. Sander, S. Schulz, A. Kargov, T. Werner, and V. Bartenbach, "Armar-4: A 63 dof torque controlled humanoid robot," in *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*, 2013, pp. 390–396.

[6] Y. Kuniyoshi and L. Berthouze, "Neural learning of embodied interaction dynamics," in *Neural Networks*, 1998, pp. 1259–1276.

[7] T. Tsutsumi, "Gap of saccadic eye movement characteristics between adduction and abduction." *Equilibrium Research*, vol. 67, no. 2, pp. 95–100, 2008, in japanese.

[8] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2881–2890.

[9] A. F. Joseph Redmon, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[10] Cao, Zhe and Simon, Tomas and Wei, Shih-En and Sheikh, Yaser, "Realtime multi-person 2d pose estimation using part affinity fields," in *CVPR*, vol. 1, no. 2, 2017, p. 7.