Using Sensor Network for Android gaze control*

Jani Even, Carlos Toshinori Ishi, Hiroshi Ishiguro
Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International, Japan.
even@atr.jp *

Abstract

This paper presents the approach developed for controlling the gaze of an android robot. A sensor network composed of RGB-D cameras and microphone arrays is in charge of tracking the person interacting with the android and determining the speech activity. The information provided by the sensor network makes it possible for the robot to establish eye contact with the person. A subjective evaluation of the performance is made by subjects that were interacting with the android robot.

1 INTRODUCTION

The eyes of a human convey a considerable amount of information during interaction. For this reason, it is important to implement a human like gaze behavior in robots that communicate with humans. In [1], the authors followed the example of the human visual system to develop the gaze of their humanoid robot. Their robot, called Kismet, controls his eyes and neck to look at target detected by four cameras located in the eyes and on the face.

The ability to perform eye contact is important but gaze also plays an important role in the mutual attention [1, 2] (and reference herein) and pointing [3]. In [4], a reactive gaze implementation for mutual attention and eye contact is presented for a humanoid robot in an explanation setting. A motion capture system is used to get the head orientation of the human. Only the robot's head is actuated. Another example of robot head with human-style gaze ability is the system presented in [5].

This paper presents the gaze control of the android robot developed for the ERATO Ishiguro Symbiotic Human-Robot Interaction Project [6]. This android is called Erica which stands for ERATO Intelligent Conversational Android. Erica was designed to have a realistic human like appearance, see Fig.1.

Figure 1: Close-up of Erica the android robot of the ERATO Ishiguro Symbiotic Human-Robot Interaction Project.

The goal of this paper, is to investigate the ability of Erica to look at a given direction in the environment. This is done by using a sensor network for finding and tracking the point of interest and controlling Erica to look at this point.

Erica is sitting on a chair, but contrary to the robots in [5, 1, 7], Erica has a complete body. Consequently, the gaze implementation presented in this paper actuates not only the eyes and the neck but also the waist of Erica.

2 SENSOR NETWORK

Before describing the gaze control, let us present how the system determine the point of interest. The basic idea is that a sensor network provides information on the context around Erica and depending of the intended interaction, a point of interest is determined.

In the current state, the sensor network main role is to track human [8, 9, 10] and determine who is talking [11, 12]. For this purpose a human tracking system is combined with a sound localization system. Figure 2

*Research supported partly by the JST ERATO Ishiguro Symbiotic Human-Robot Interaction Project and partly by the Ministry of Internal Affairs and Communications of Japan under the Strategic Information and Communications R&D Promotion Programme (SCOPE).

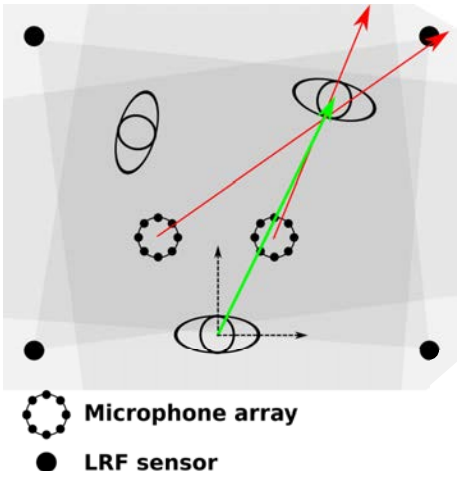

Figure 2: Example of possible sensor network configuration.

shows one example of configuration with four laser range finders (LRFs) for tracking humans and two microphone arrays for performing sound localization. During the experiments, the human tracker system was not using LRFs but RGB-D cameras attached to the ceiling of the room [13]. Using the sound localization (the red arrows in Fig.2) it is possible to determine who is talking. Then the goal is to have Erica pays attention to that person (the green arrow). Namely, the sensor network gives a point of attention that can vary. This point of attention is referred to as the *focus point* in the remainder.

3 KINEMATIC CHAIN

In this section, we describe the kinematic chain to be controlled for setting the gaze of Erica at a given focus point.

Figure 3 shows the joints involved in the gaze control. The kinematic chain controlling the eyes direction has 7 degrees of freedom (DOF):

- yaw and pitch for the eyes,
- yaw pitch and roll for the neck,
- yaw and pitch for the waist.

However, the current implementation does not use the neck roll.

Pneumatic actuators are used to move the joints. These actuators are controlled by on board PID controllers. The commands are sent to the robot at a frequency of 20 Hz. The robot provides a feedback measured by potentiometers also at the frequency of 20 Hz. The on board PID are tuned to favor smoother movements which results in a lesser control accuracy. Consequently, it is necessary to rely on the feedback to get the achieved positioning.

Using the specifications of Erica, a computer model of the kinematic chain was implemented. The posture of the model is updated when the feedback from the actuators is received. Namely, the model provides an estimate of the current posture of Erica.

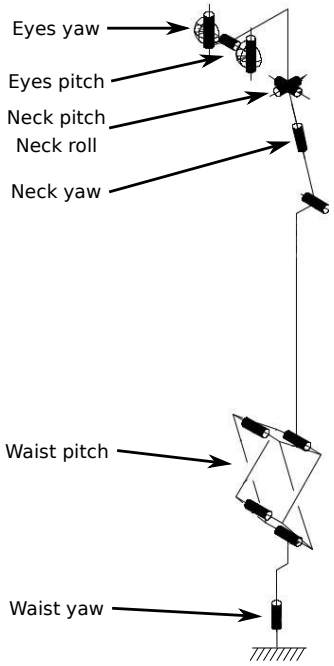

Figure 3: Kinematic chain for the gaze.

4 GAZE CONTROL

The kinematic model presented in the previous section provides the current gaze direction of Erica's eyes.

This is illustrated in Fig.5 that depicts a part of the graphical user interface (GUI). In the left view, the current gazing direction of Erica is shown by the pink line. The green line shows the direction of the focus point (the red box). At this moment Erica is not requested to look at the focus point. When asked to look at the focus point, the look direction (pink) is aligned to the focus point direction (green) as in the right part of Fig.5. The goal of the gaze control is to send command to move the joints of Erica in order to perform this alignment.

Figure 4 shows the flow chart of the algorithm. The control sequence is as follows:

1. the position of the focus point $f(k)$ is given to the gaze control,
2. the gaze control requests the current gaze direction (i.e. the current orientation of the joints $\widetilde{\theta(k)}$) to the kinematic model,
3. if the gaze direction is close enough to the focus point direction go to 9 otherwise go to 4,

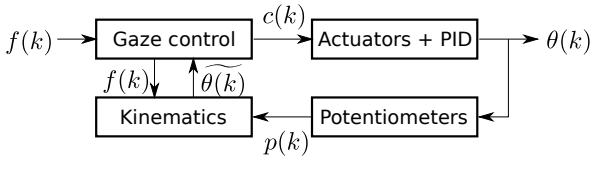

Figure 4: Flowchart showing the different blocks of the gaze control.

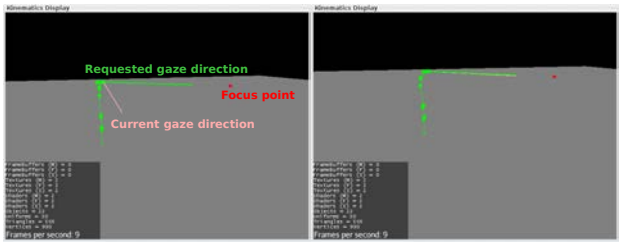

Figure 5: Visualization of the kinematic chain and the focus point.

4. the gaze controller determines the commands $c(k)$ to send to the joints,
5. the actuators move the joints,
6. the potentiometers give the feedback $p(k)$,
7. the kinematic model is updated,
8. loop to 2,
9. gaze control completed.

The step 4 is the most important ones. Given the direction of the focus point and the current direction of the gaze, the controller has to determine the commands to send to the different joints.

Figure 6 illustrates the procedure for the yaw command of the waist, the neck and the eyes. In Fig.6-a, Erica has a posture determined by the waist yaw, neck yaw and eye yaw and is requested to look in the set gaze direction. All these directions are represented by the colored arrows. The controller determines the desired angles for the joints. These angles are represented by the dash arrows in Fig.6-b. For the waist (red) and neck (orange), the desired angles are converted in absolute commands defined as a fraction of the set angles. For the eyes (yellow), the desired angle is converted to a command relative to the current position of the eyes which is represented by the black arrow.

Let us denotes the set angle by $\theta(k)$ then the waist and neck angles are

$$\begin{aligned} \theta_{\text{waist}}(k) &= \alpha_{\text{waist}}\theta(k) \\ \theta_{\text{neck}}(k) &= \alpha_{\text{neck}}\theta(k) \end{aligned} \tag{1}$$

where α_{waist} and α_{neck} control the amount of rotation distributed to the waist and neck.

For the eye angle, the relative value is

$$\theta_{\text{eyes}}(k) = \theta(k) - \widetilde{\theta_{\text{eyes}}(k)} \tag{2}$$

where $\widetilde{\theta_{\text{eyes}}(k)}$ is the estimated eye angle given by the kinematic model.

Only the eyes are controlled in a closed loop because the accuracy on the eye movement is greater than on the waist and neck.

When the joints have started to move, as in Fig.6-c, the absolute angles for the waist and neck do not change whereas the relative one for the eye is updated.

Figure 6-d shows an example of gaze control completion. In this case, a small error still exists for the neck that did not reach the absolute angle (the double black arrow). However, the gaze direction is reached as the relative angle computed for the eyes compensated the residual error on the neck.

In practice, for all the joints, the angles are converted in command values that are in the range $[0, 255]$ before sending them to the robot. The feedback values received are also in the range $[0, 255]$. The conversion is a simple linear mapping. For example for the eyes

$$c_{\text{eyes}}(k) = \theta_{\text{eyes}}(k) * \frac{255}{\theta_{\text{eyes, max}} - \theta_{\text{eyes, min}}}$$

$$\widetilde{\theta_{\text{eyes}}}(k) = \theta_{\text{eyes, min}} + p_{\text{eyes}}(k) * \frac{\theta_{\text{eyes, max}} - \theta_{\text{eyes, min}}}{255}$$

where $\theta_{\text{eyes, min}}$ and $\theta_{\text{eyes, max}}$ are the angles corresponding to the command or the potentiometer values 0 and 255.

5 EXPERIMENTAL RESULTS

5.1 objective evaluation

In this experiment, the focus point was set to subject tracked by the sensor network. This subject was walking in front of Erica for four minutes. The direction of the subject (the focus point) and the estimated gaze direction given by the kinematic model were recorded. The command and potentiometer values were also recorded. The goal of this experiment is to check if Erica is able to track a moving focus point using the proposed gaze control approach.

The top of Fig.7 shows the yaw of the focus direction (solid line) and the yaw of the gaze direction given by the kinematic model (dashed line). The three other graphs are showing the command values (solid lines) and the potentiometer values (dashed lines) for the control of the waist, neck and eyes yaw. The focus direction is well tracked by the gaze direction except for the period between the two vertical red dashed lines.

Figures 8 and 9 respectively show a good tracking period (the green vertical dashed lines in Fig. 7) and the bad tracking region. The top graph of Fig.8 clearly shows that the gaze direction closely follows the focus direction. We can note a slight delay, which is expected, and some overshoots. However, the graph for the neck control shows some large errors and the one for the waist some small errors. These two graphs are by construction scaled version of the focus angle, see Eq.(1). Then, we can see on the graph for the eyes that the command is different and it compensated for the error as expected.

The tracking error that appears in Fig.9 is explained by the fact that the large error on the neck angle could not be corrected by the eyes because they saturated (the command reached 0). This is due to the fact that the subject was at a large focus angle.

Figure 10 shows the cumulative density functions (CDFs) for the errors on the yaw (left) and the pitch (right). The horizontal black dashed lines indicate the 90% quantiles. For the yaw, 90% of the errors are smaller

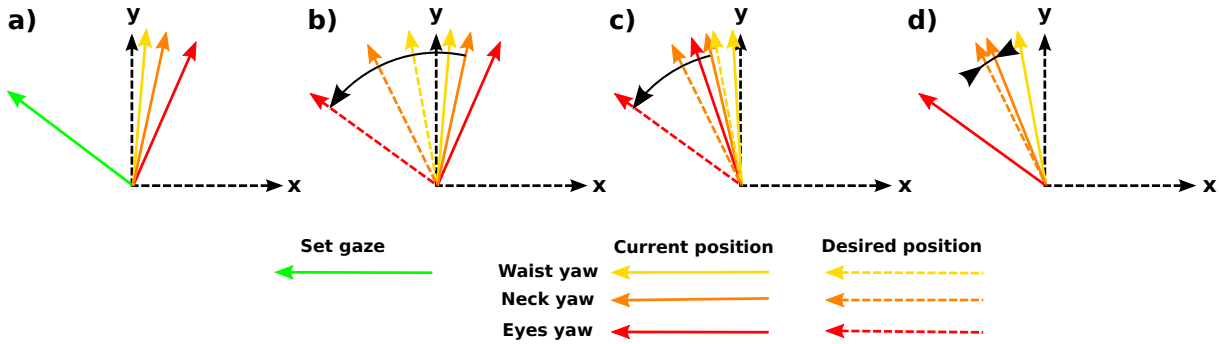

Figure 6: Ratio for the different body parts during gaze setting. Note the remaining error on the neck in d.

than 11 degrees and for the pitch smaller than 5 degrees. The larger error on the yaw is due to the fact that while the person was moving in front of Erica, the pitch did not vary much whereas the yaw presented large variations. The error showed in Fig.9 created the small bump around 45 degrees in the CDF of the yaw. Note that these errors are computed while tracking a moving person. Then the small tracking delays contribute to the error for the lower values of the CDFs.

Figure 11 shows the cumulative density functions (CDFs) for the errors on the yaw command for the waist (left), the neck (center) and the eyes (right). As expected, the 90% quantile is significantly higher for the neck.

This experiment showed that the proposed approach is able to accurately track a moving focus point. The performance was measured on the feedback given by the potentiometers. This means that some bias may be present if the calibration is not done properly. Namely, the measured focus direction and the true focus direction may differ.

A finding is also that most of the error comes from the neck. In particular, for some large angles, Erica could not look at the desired directions because of the error on the neck positioning. These situations correspond to cases where the human would also turn on themselves to look. This is due to friction forces that prevent the neck actuator to achieve the desired positioning while moving smoothly. To solve this problem, a low level controller that is aware of the friction will be implemented.

5.2 Subjective evaluation

The subjective evaluation of the gaze control is performed by setting a focus point and asking a subject to position herself/himself where she/he feels Erica is making eye contact with her/him. This is done for several focus points in front of Erica, see Fig.12. Then for each of the focus points, the position where eye contact is felt the best is recorded using the human tracker. The height of the subject eyes is measured to set the height of the focus points. For the focus points, the yaw angle θ is computed and for the corresponding position of perceived eye contact, the yaw angle $\widehat{\theta}$ is also computed. For the selected focus point, in green in Fig.12, the yaw angle θ is represented by the green arrow and the yaw angle $\widehat{\theta}$ of the perceived eye contact is represented by a red arrow.

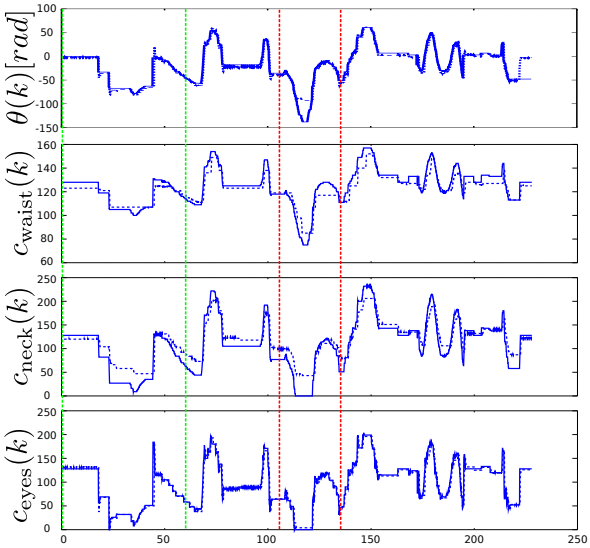

Figure 7: Axis command (dashed) and potentiometer feedback (solid) for the yaw.

Figure 13 is a plot of the perceived angles versus the focus point angles. The data points for two different subjects are plotted (circles and crosses). The black line is $\widehat{\theta} = \theta$ and the red line is the linear fit:

$$\widehat{\theta} = 0.79\ \theta + 7.92 \tag{3}$$

the RMSE is 5.27.

The angle $\widehat{\theta}$ of the perceived eye contact does not correspond to the set angle θ. Meaning that the subjects did not feel the eye contact at the exact set position.

However, a linear fit of the data is possible. The bias of 7.92 degrees and the scaling error of 0.79 could be explained by calibration errors. The ranges $\theta_{\text{XXX, max}}$ and $\theta_{\text{XXX, min}}$ (where XXX is for waist, neck or eyes) have to be adjusted.

Without re-calibration of the ranges, the linear fit could be used to select the set angle to look at a position given by the human tracker:

$$\theta = 1.21\ \widehat{\theta} - 9.47 \tag{4}$$

the RMSE is 6.54. Figure 14 shows this linear fit.

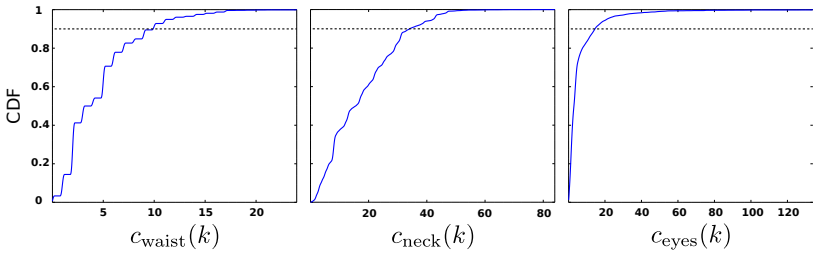

Figure 11: Cumulative density functions for the command errors for the waist (left), the neck (center) and the eyes (right).

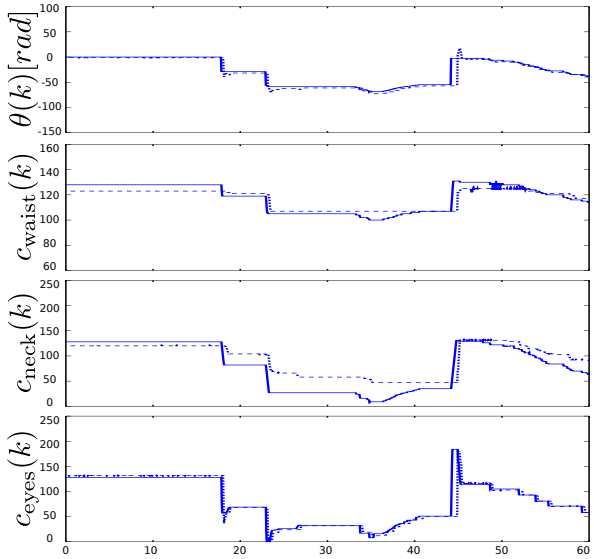

Figure 8: Close-up of the axis command (dashed) and potentiometer feedback (solid) for the yaw.

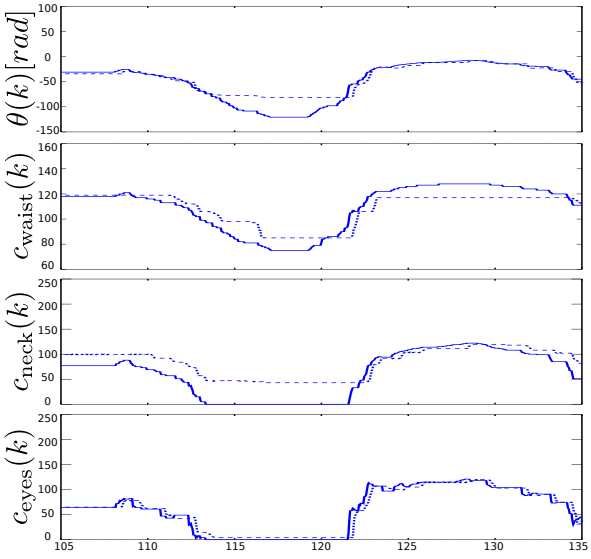

Figure 9: Close-up of the axis command (dashed) and potentiometer feedback (solid) for the yaw.

6 CONCLUSIONS

This paper presented the low level gaze function of Erica. The objective experiment showed that the gaze control is behaving as expected. The system is able to compensate the measured error. However, the subjective evaluation suggests that there is still a calibration to be done in order to obtain eye contact. An alternative way would be to use the linear mapping between perceived gaze angle and set angle.

In addition to the ability to look at a given point, a humanoid robot should also reproduce a human like behavior [14, 15]. Human like features of the gaze are implemented at a higher level in Erica's control architecture. The integration of these higher level features with the low level control will be the focus of future research.

References

[1] C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati, "Active vision for sociable robots," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 31, no. 5, pp. 443–453, 2001.

[2] B. Scassellati, "Investigating models of social development using a humanoid robot," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, 2003, vol. 4, pp. 2704–2709 vol.4.

[3] Sotaro Kita, *Interplay of gaze, hand, torso orientation and language in pointing, in Pointing: Where Language, Culture, and Cognition Meet*, Lawrence Erlbaum Associates, 2003.

[4] Y. Mohammad and T. Nishida, "Reactive gaze control for natural human-robot interactions,"

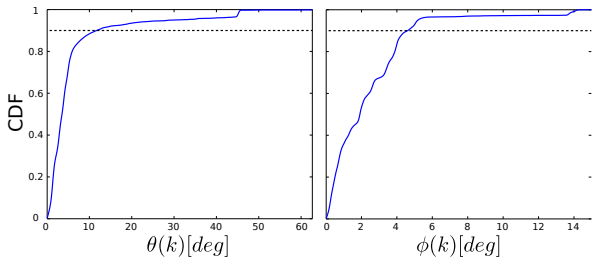

Figure 10: Cumulative density functions for the angular errors for the yaw (left) and the pitch (right).

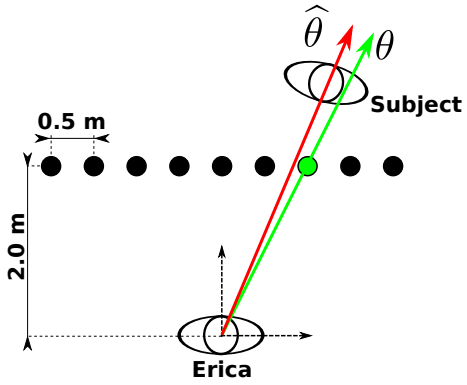

Figure 12: Settings for the subjective test showing the focus points.

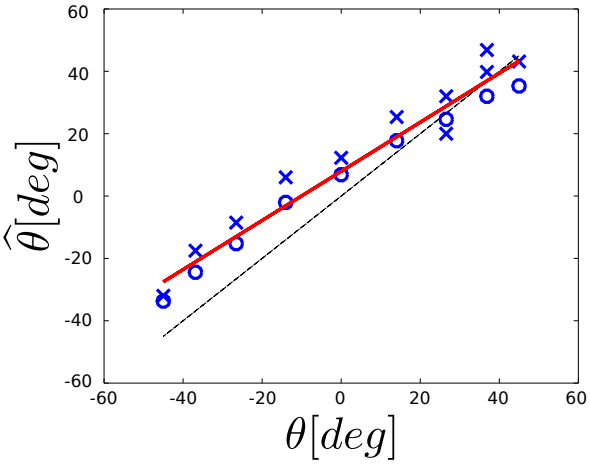

Figure 13: Subjective angle $\widehat{\theta}$ versus set angle θ. The red line is a linear fit of the data points.

in *Robotics, Automation and Mechatronics, 2008 IEEE Conference on*, 2008, pp. 47–54.

[5] A. Takanishi, H. Takanobu, I. Kato, and T. Umetsu, "Development of the anthropomorphic head-eye robot we-3rii with an autonomous facial expression mechanism," in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, 1999, vol. 4, pp. 3255–3260 vol.4.

[6] Hiroshi Ishiguro et al., "Erato ishiguro symbiotic human-robot interaction project," `http://www.jst.go.jp/erato/ishiguro/en/index.html`, 2015.

[7] D. Hanson and The University of Texas at Dallas, *Humanizing Interfaces: An Integrative Analysis of the Aesthetics of Humanlike Robots*, University of Texas at Dallas, 2007.

[8] Jae Hoon Lee, T Tsubouchi, K Yamamoto, and S Egawa, "People tracking using a robot in motion with laser range finder," 2006, pp. 2936–2942, Ieee.

[9] D.F. Glas et al., "Laser tracking of human body motion using adaptive shape modeling," *Proceedings of 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 602–608, 2007.

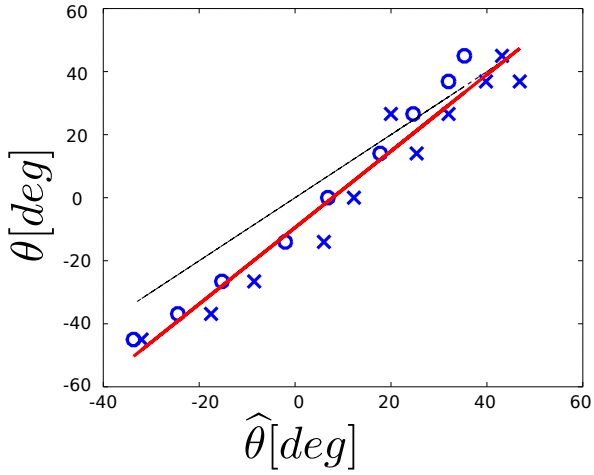

Figure 14: Set angle θ versus subjective angle $\widehat{\theta}$. The red line is a linear fit of the data points.

[10] L. Spinello and K. O. Arras, "People detection in rgb-d data.," in *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2011.

[11] C.T. Ishi et al., "Evaluation of a music-based real-time sound localization of multiple sound sources in real noisy environments," *Proceedings of 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2027–2032, 2009.

[12] C.T. Ishi, J. Even, and N. Hagita, "Using multiple microphone arrays and reflections for 3d localization of sound sources," *Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3937–3942, 2013.

[13] D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-d range sensors," *Human-Machine Systems, IEEE Transactions on*, vol. 43, no. 6, pp. 522–534, 2013.

[14] J.M. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonomic Bulletin & Review*, vol. 1, no. 2, pp. 202–238, 1994.

[15] R Weidner, J Krummenacher, B Reimann, H Muller, and G Fink, "Sources of top-down control in visual search," *Cognitive Neuroscience, Journal of*, vol. 21, no. 11, pp. 2100–2113, 2009.