Non-invasive Gaze Direction Estimation from Head Orientation for Human-Machine Interaction

Zhi Zheng[1(✉)], Yuguang Wang[2], Jaclyn Barnes[2], Xingliang Li[2], Chung-Hyuk Park[3], and Myounghoon Jeon[2]

[1] University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA
Zheng36@uwm.edu
[2] Michigan Technological University, Houghton, MI 49930, USA
{jaclynb,xli17,mjeon}@mtu.edu
[3] George Washington University, Washington, DC 20052, USA
chpark@gwu.edu

Abstract. Gaze direction is one of the most important interaction cues that is widely used in human-machine interactions. In scenarios where participants' head movement is involved and/or participants are sensitive to body-attached sensors, traditional gaze tracking methods, such as using commercial eye trackers are not appropriate. This is because the participants need to hold head pose during tracking or wear invasive sensors that are distractive and uncomfortable. Thus, head orientation has been used to approximate gaze directions in these cases. However, the difference between head orientation and gaze direction has not been thoroughly and numerically evaluated, and thus how to derive gaze direction accurately from head orientation is still an open question. In this article, we have two contributions in solving these problems. First, we evaluated the difference between people's frontal head orientation and their gaze direction when looking at an object in different directions. Second, we developed functions that can map people's gaze direction using their frontal head orientation. The accuracy of the proposed gaze tracking method is around 7°, and the method can be easily embedded on top of any existing remote head orientation method to perform non-invasive gaze direction estimation.

Keywords: Gaze direction estimation · Head orientation
Human-machine interaction

1 Introduction

Gaze direction is one of the most useful interaction cues in human-machine interactions (HMI). Gaze tracking has been widely applied in fields such as human–computer interaction [1–3], human-robot interaction [4–6], and virtual reality [7, 8]. Many existing gaze tracking technologies use invasive methods [9], such as wearing eye tracking glasses [10] and attaching electrooculogram sensors around the eyes [11]. However, many people cannot use such methods due to their sensitivity to body

M. Kurosu (Ed.): HCI 2018, LNCS 10903, pp. 380–389, 2018.
https://doi.org/10.1007/978-3-319-91250-9_30

attached hardware [12]. In addition, wearing hardware limits people's activities and may cause discomfort, too. While there are some methods to avoid direct contact with the hardware, such as using eye trackers that can be place in front of the participant [13], the calibration process of these devices are time consuming and may not work for some participants such as very young children. In addition, participants need to hold their head still during tracking, which is not applicable in HMI studies that require large head movements [12, 14]. In addition, most of these devices are expensive and only available to professionals. Thus, gaze tracking methods using invasive hardware and/or requiring head to be held do not satisfy many HMI scenarios.

Since people tend to turn their head to a target when looking at it, head orientation can also be used to indicate gaze direction. Previous research has used frontal head orientation to approximate gaze direction directly [15]. However, literature [16, 17] and common sense tell us that differences exist between people's frontal orientation and their gaze direction. Therefore, other studies modeled gaze direction as an uncertainty with assumed probability distribution on top of head orientation [18] or captured clear eye images (usually requires cameras positioned closely to the participant) and ensemble gaze angles on top of head orientation [19, 20]. Nevertheless, only a few works [16, 17] have studied the difference between head orientation and gaze direction carefully, while most of them focused on this difference for attention tracking, instead of the explicit relation between gaze and head orientation.

We aim to solve these problems in this paper. ***First, we conducted a detailed evaluation on the difference between frontal head orientation and gaze direction***. Based on this difference, we ***introduced a novel gaze direction estimation method that accommodates the rotation of head, without the need of invasive hardware or capturing clear eye images.***

This new method approximates gaze direction of a person using his/her head orientation information. Head orientation can be detected remotely and precisely using computer vision techniques [21]. In this paper, we chose Microsoft Kinect, which is a low cost device available to the public. Seven adults participated in a data collection experiment, where their synchronized gaze direction data and head orientation data were recorded. These data were used to train mapping functions that use head orientation as input to calculate the corresponding gaze direction.

Comparing to previous gaze direction estimation methods, the proposed method can be implemented directly on top of any existing head orientation method and does not require any extra hardware for gaze tracking. Since the computational cost of the mapping functions is low, real-time gaze tracking is possible if the functions are applied on real-time head orientation estimation methods. As remote head orientation estimation methods [21], such as the CSIRO Face Analysis SDK [22] and Microsoft Kinect, have been well developed, applying the gaze mapping functions on top of these methods results in non-invasive gaze tracking easily.

This paper is structured as follows: Sect. 2 shows the data collection and processing of this study. Section 3 presents the difference between gaze direction and frontal head orientation. Section 4 introduces how the gaze direction estimation functions were designed accordingly and their accuracies. Finally, Sect. 5 concludes this article and discusses future works.

2 Data Collection and Processing

Functions that map head orientation to gaze direction were derived by data fitting. We conducted a data collection experiment with seven adults recruited as participants.

In each experiment session, a participant was asked to look at a moving marker (a red spot) projected on a wall (a plane). The position of the marker was used to indicate the ground truth of the participant's gaze direction. Before the experimental trials, the marker was at the center of the display region, and the participant's head orientation when his/she was looking at the marker was recorded as a baseline value. This value was used to calibrate the data for each participant as discussed in the later part of this section. The marker's movement was arranged in multiple trials. In each trial, the marker started from a random position and moved horizontally, vertically, and diagonally following a random order. The participant may unconsciously anticipate the motion of the marker if it moved in the same direction for a long period. To solve this problem, we kept each trial short, which lasted from a few seconds to about a minute. Meanwhile, the moving speed of the marked was adjusted so that participants could follow the motion of the marker easily. After each trial, the marker disappeared and showed up in another position to start the next trial until the end of the experiment session. The combined path of the marker in all the trials covered a region of 360 cm × 117 cm. Figure 1 illustrates an example of the path that the moving marker followed.

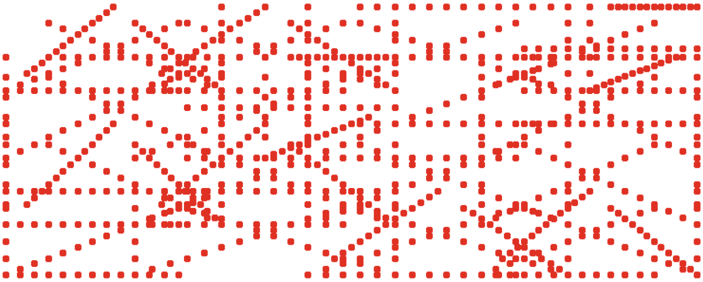


Fig. 1. Moving marker path example (Color figure online)

During the experiment, the participant was seated facing the central part of the display region. The center of the head was approximately 150 cm from the ground and 160 cm from the display region. Therefore, in order to look at the whole display region, the participants' gaze needed to shift from −48.37° to 48.37° in horizontal direction and from −18.81° to 21.34° in vertical direction, respectively, as shown in Fig. 2. This simulates a normal gaze range when people are communicating with other agents or paying attention to objects in front of them. When a participant was looking at the marker, his/her head orientation was estimated using Microsoft Kinect. As shown in Fig. 2, the Kinect was placed at the bottom of the display region's central part, facing the participant. Thus, the participant's head was in the view of the Kinect to estimate the head pose, while the Kinect did not block the participant's vision towards the

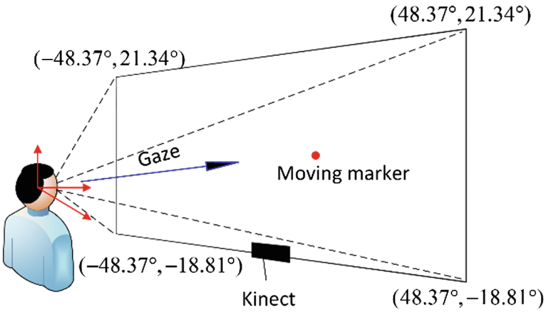


Fig. 2. Data collection experiment configuration

display region. The head orientation data were recorded and synchronized with the participant's ground truth gaze direction, and thus data pairs can be used to calculate the difference between gaze direction and frontal head orientation as well as fit mathematical functions that map head orientation to gaze direction.

The frontal head orientation (a vector) $\vec{f}$ was computed from the head rotation quaternions given by the Kinect for Windows Software Development Kit 2.0. The term $\vec{f}$ was projected horizontally and vertically, and thus generated two angles, α and β, which represented the head rotation horizontally and vertically from the frontal direction, respectively. As mentioned in the last section, before the experimental trials, each participant's baseline head orientation was recorded when they were instructed to look at the center of the display region, which was approximately right in front of the participant. The baseline orientation was used to calibrate a participant's head orientation. This is necessary since different participant tended to look at the same spot in the display region with different head orientation. For example, some people may raise head a little more than others, and some people may turn their head to one side a little more than the other side. Therefore, we recorded a participant's head orientation at the calibration point as (α_c, β_c), and all the head orientation data was subtracted by this pair to eliminate the baseline differences, i.e., $\alpha_f = \alpha - \alpha_c$, $\beta_f = \beta - \beta_c + \Delta\beta$. The calibration point is slightly higher than the exact horizontal direction. $\Delta\beta$ is the difference between the horizontal direction and the vertical direction of the calibration point, which is approximately 1.43° in this study.

The calibrated head orientation angles α_f and β_f are used to fit the mapping functions. As shown in Fig. 3, we define the position of the moving marker as (x, y), which is associated with the ground truth gaze direction (α_g, β_g). Based on the geometry of the experimental setup,

$$\alpha_g = \tan^{-1}(x/160), \tag{1}$$

$$\beta_g = \tan^{-1}(y + 4/160). \tag{2}$$

As illustrated in Fig. 3, the gaze direction does not overlap with the head orientation exactly.

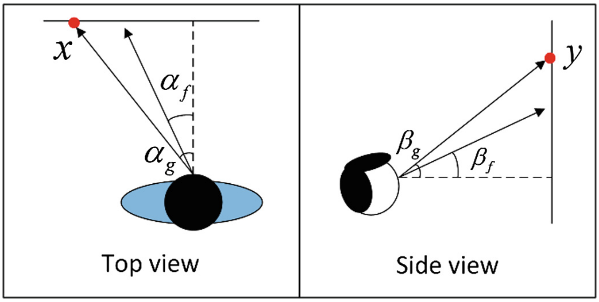


Fig. 3. Illustration of the gaze angles and head orientation angles

Data streams of (α_g, β_g) and (α_f, β_f) were synchronized so that all the ground truth gaze direction data were paired with their synchronized and calibrated head orientation data. From the 7 participants, 7281 pairs of data were collected initially. Even though the participants were instructed to stare at the moving marker as much as possible, we still observed occasional unconscious gaze shift and its corresponding head rotation shift during the experiment. Therefore, to eliminate outliers, we divided the display region in to 300 grids, and each grid was 12 cm (length) by 11.7 cm (height). For each grid, the following steps were executed:

Step 1: All the (α_f, β_f) that correspond to (α_g, β_g) within this grid were used to calculate their mean values $(\bar{\alpha}_f, \bar{\beta}_f)$ and the standard deviation $(\sigma_f^{\alpha}, \sigma_f^{\beta})$.

Step 2: For a head orientation data point (α_f^i, β_f^i), if $\left|\alpha_f^i\right| > \bar{\alpha}_f + 2\sigma_f^{\alpha}$ or $\left|\beta_f^i\right| > \bar{\beta}_f + 2\sigma_f^{\beta}$, this point was removed from the dataset along it's paired gaze point.

After these two steps, 554 pairs (7.6%) were removed from the initial synchronized data, leaving 6727 pairs for fitting the mapping functions.

3 Difference Between Gaze Direction and Frontal Head Orientation

Figure 4 plots α_g and β_g against (α_f, β_f) in a 3 dimensional space. We can see that the frontal head orientation and gaze direction are related but not coincident. The distributions of the points are slightly non-linear.

In addition, from the data, we found that the further a participant was looking away from his/her frontal direction, the larger the difference between his/her head orientation and gaze direction was. This is an important behavioral phenomenon that needs to be considered when approximating gaze direction using head orientation.

We divided the data into 21 sets in both horizontal and vertical directions, based on the data's distance from the center of the display. Thus, each horizontal set covers a range of about 4.61°, and each vertical set covers a range of about 1.91°. Then, we

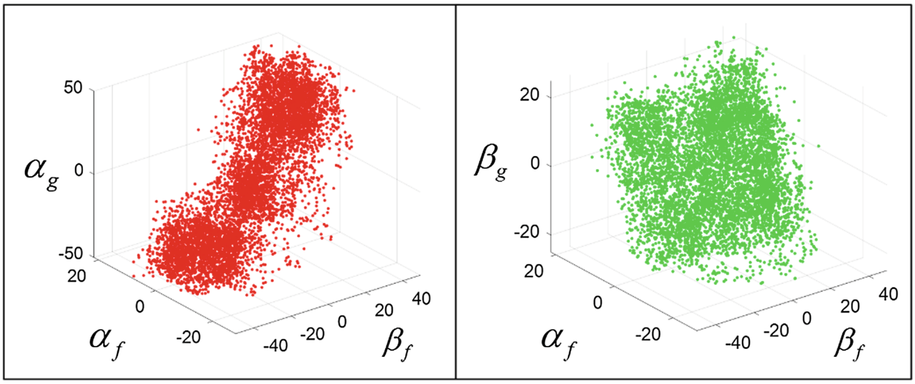
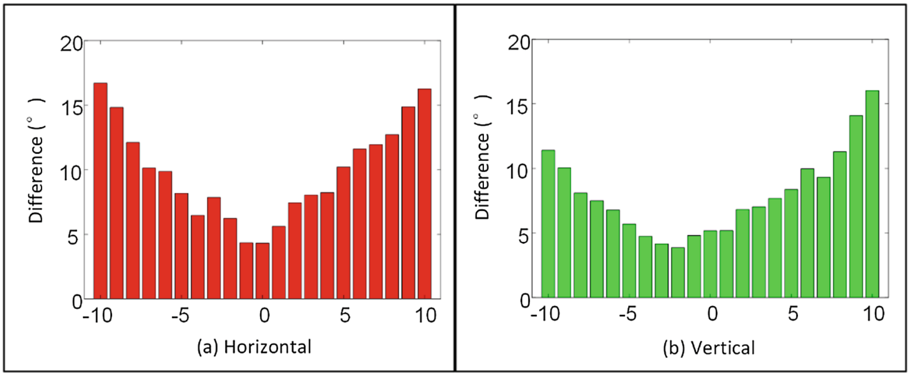


Fig. 4. Plots of α_g and β_g against (α_f, β_f)

calculated the difference between the frontal head orientation and the gaze direction in each direction within each set. Figure 5 shows the patterns clearly, where each bar indicates the difference in one set (marked from −10 to 10). Subfigures (a) and (b) illustrate the average values of $|\alpha_g - \alpha_f|$ and $|\beta_g - \beta_f|$, respectively. Larger x-coordinates represent longer distances from the center of the display. In the horizontal direction, bar at 0 indicates the difference in the center set ([−2.31°, 2.31°]), bars at negative x coordinates are the differences on the left plane, and bars at positive x coordinates are the differences on the right plane. In the vertical direction, bar at 0 indicates the difference in the center set ([−0.96°, 0.96°]), bars at negative x coordinates are the differences on the lower plane, and bars at positive x coordinates are the differences on the upper plane.



(a) Horizontal

(b) Vertical

Fig. 5. Differences between gaze direction and frontal head orientation

From Fig. 5, we can see that the differences between gaze direction and the frontal head orientation became larger when the participants looked further away from the center in both horizontal and vertical directions. In horizontal direction, the difference in the range of [−20.75°, 16.14°] (bar −4 to bar 3) was lower than 8°. The minimum difference (4.31°) was at [−2.31°, 2.31°] (bar 0). The difference increased as the participants' gaze shifted to the sides. At bar −10 and 10, the differences are 16.71° and 16.26°, respectively. In the vertical direction, the difference in the range of [−10.51°, 8.60°] (bar −5 to bar 4) was lower than 6°. The minimum difference (3.86°) was at [−4.78°, −2.87°] (bar −2). The difference increased as the participants' gaze shifted upwards and downwards. At bar −10 and 10, the differences are 11.41° and 16.03°, respectively.

4 Derivation of the Mapping Functions for Gaze Direction Estimation

Two mapping functions F_1 and F_2 were derived as $\alpha_g = F_1(\alpha_f, \beta_f)$ and $\beta_g = F_2(\alpha_f, \beta_f)$ by fitting 2D surfaces using the points clouds. Linear and polynomial regressions were conducted with linear interpolation. The average error of linear regression for F_1 and F_2 was 7.81° and 7.83°, respectively. The RMSDs decrease slightly in polynomial regressions, however, high orders beyond the second cause overfitting and thus are not appropriate. Therefore, we choose the second-order polynomial regression, and the average error of F_1 and F_2 was 7.74° and 7.63°, respectively. The form of the mapping functions are as follows, with the coefficients listed in Table 1:

Table 1. Coefficients of F_1 and F_2 (rounded to 6 decimals)

	P_{00}	P_{10}	P_{01}	P_{11}	P_{20}	P_{02}
F_1	1.492150	1. 320722	0.082641	−0.007427	−0.001641	−0.002192
F_2	4.338289	−0.024300	1.066991	−0.003938	−0.004744	0.000685

$$F = P_{00} + P_{10}\alpha_f + P_{01}\beta_f + P_{02}\alpha_f^2 + P_{11}\alpha_f\beta_f + P_{02}\beta_f^2. \tag{3}$$

Therefore, F_1 and F_2 can be used to approximate the participants' gaze direction using their frontal head orientation within the range of the experimental setup. Figure 6, shows the fitted surface that represent F_1 and F_2, respectively. The nonlinearity of the two functions can be observed in the graph, especially in F_2. These two functions can be easily embedded to any existing head orientation estimation methods for gaze tracking.

The mapping functions were embedded into the existing Kinect head orientation estimation program. Due to the low computational cost, this gaze tracking method can work in real-time (about 30 frames per second).

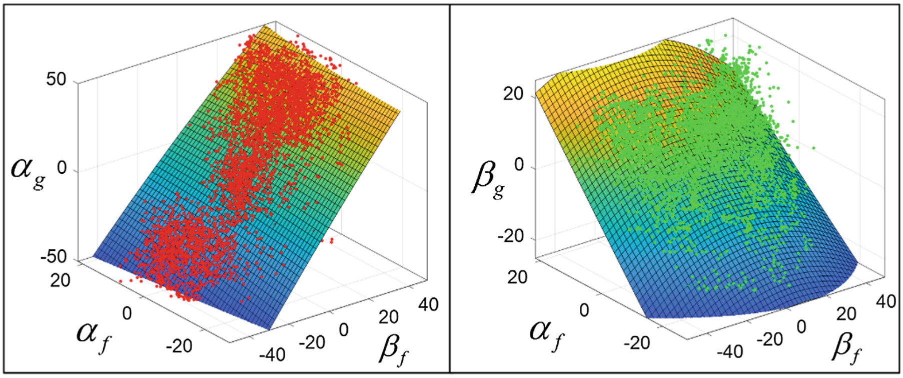


Fig. 6. Plots of mapping functions

5 Conclusion and Discussion

Since people tend to turn their head towards the target when looking at it, we can extract gaze direction using frontal head orientation. In this work, we first studied the relation, especially the difference, between people's gaze direction and their frontal head orientation. Then, we proposed a gaze direction estimation method based on frontal head orientation. Seven participants were recruited and instructed to look at a moving marker that indicated their ground truth gaze direction. Meanwhile, their head orientation when looking at the moving marker was recorded using Microsoft Kinect. We found that people's gaze direction deviate from their head orientation when they look away from frontal direction. In order to use head orientation to estimate gaze direction, mapping functions could be fitted using the synchronized gaze and frontal head orientation data. Two second-order polynomial mapping functions were derived through regression. The average error of the proposed method is below 8°. The functions are with low computational cost and are independent of the particular devices/hardware for head orientation detection. Therefore, it can be applied on any existing head orientation estimation techniques, such as using a Microsoft Kinect, to achieve real-time gaze tracking with low cost. Comparing with other gaze tracking methods, the proposed method does not require any other hardware/sensors particularly for gaze tracking except those for head orientation estimation. Since head orientation has been studied for decades and there are many available methods and products for it, extending them for gaze tracking using the proposed method is easy and handy.

However, there are a few limitations of the current work that need to be addressed in the future. First, the accuracy (around 7°) is low compared with that of using commercial eye trackers under careful calibration and withholding head pose. Therefore, the proposed method may not be applied to studies that require very accurate gaze tracking, such as tracking the exact point that a participant is looking at. Instead, the proposed method can be used in scenarios where a general and quick referring of people's gaze is enough, such as distinguishing what large objects a participant looks at and the gaze switching between these objects [16]. Since the objects in many human-machine interaction studies are separated much more than 8° visually, the proposed method has a great potential to provide a quick solution for referring gaze in these studies. Examples include: (1) human-robot interaction studies, where the system needs to distinguish whether a participant is looking at a robot or another object positioned far away from the robot [14]; and (2) human-computer interaction studies, where participants need to look at different monitors during the interaction [12].

The second limitation of the current study is the small sample size. Only seven participants' data were used, and thus the fitted functions are tuned for this small group. In order to develop a general function that works well for larger population, more data need to be collected from a large sample in the future.

Another way to improve the current work is to study how the combination of head orientation and body gestures impacts the gaze direction. The current work studied participants' looking behaviors when they were facing the object (i.e., the display region). However, people's gaze direction with respect to the head orientation may be influenced by body posture. For example, when a person is called from back, he/she may turn the upper body half way and rotate the head to look back, or this person may turn the whole body to the back completely without turning the head with respect to the body. These are more complex cases compared with the current study and have not been thoroughly and symmetrically addressed in previous research.

In summary, this paper introduced an easy and quick way to estimate gaze from frontal head orientation. The proposed method gives a coarse gaze direction estimation and can be applied on human-machine interactions that require rough gaze direction. This method can be embedded into any existing head orientation estimation methods and does not require extra hardware. Although limitations exist, we believe the current work is an important step towards more accurate, cost effective, and non-invasive gaze tracking technologies.

References

1. Hutchinson, T.E., et al.: Human-computer interaction using eye-gaze input. IEEE Trans. Syst. Man Cybern. **19**(6), 1527–1534 (1989)
2. Jacob, R., Karn, K.S.: Eye tracking in human-computer interaction and usability research: ready to deliver the promises. Mind **2**(3), 4 (2003)
3. Lazar, J., Feng, J.H., Hochheiser, H.: Research Methods in Human-Computer Interaction. Morgan Kaufmann, San Francisco (2017)
4. Rich, C., et al.: Recognizing engagement in human-robot interaction. In: 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE (2010)
5. Lallée, S., et al.: Cooperative human robot interaction systems: IV. Communication of Shared Plans with Naïve Humans Using Gaze and Speech. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2013)
6. Moon, A., et al.: Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction. ACM (2014)
7. Ohshima, T., Yamamoto, H., Tamura, H.: Gaze-directed adaptive rendering for interacting with virtual space. In: Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium. IEEE (1996)

8. Lee, E.C., Park, K.R., Whang, M.C., Park, J.: Robust gaze tracking method for stereoscopic virtual reality systems. In: Jacko, Julie A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 700–709. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73110-8_76
9. Holmqvist, K., et al.: Eye Tracking: a Comprehensive Guide to Methods and Measures. OUP, Oxford (2011)
10. Ye, Z., et al.: Detecting eye contact using wearable eye-tracking glasses. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. ACM (2012)
11. Manabe, H., Fukumoto, M., Yagi, T.: Direct gaze estimation based on nonlinearity of eog. IEEE Trans. Biomed. Eng. **62**(6), 1553–1562 (2015)
12. Zheng, Z., et al.: Design of an autonomous social orienting training system (ASOTS) for young children with autism. IEEE Trans. Neural Syst. Rehabil. Eng. **25**(6), 668–678 (2017)
13. Harezlak, K., Kasprowski, P., Stasch, M.: Towards accurate eye tracker calibration–methods and procedures. Procedia Comput. Sci. **35**, 1073–1081 (2014)
14. Zheng, Z., et al.: Design, development, and evaluation of a non-invasive autonomous robot-mediated joint attention intervention system for young children with ASD. IEEE Trans. Hum. Mach. Syst. (2017). Preprint. https://doi.org/10.1109/THMS.2017.2776865
15. Bekele, E.T., et al.: A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. IEEE Trans. Neural Syst. Rehabil. Eng. **21**(2), 289–299 (2013)
16. Massé, B., Ba, S., Horaud, R.: Tracking gaze and visual focus of attention of people involved in social interaction. IEEE Trans. Pattern Anal. Mach. Intell. (2017)
17. Sheikhi, S., Odobez, J.-M.: Combining dynamic head pose–gaze mapping with the robot conversational state for attention recognition in human–robot interactions. Pattern Recogn. Lett. **66**, 81–90 (2015)
18. Mukherjee, S.S., Robertson, N.M.: Deep head pose: Gaze-direction estimation in multimodal video. IEEE Trans. Multimed. **17**(11), 2094–2107 (2015)
19. Lu, F., et al.: Learning gaze biases with head motion for head pose-free gaze estimation. Image Vis. Comput. **32**(3), 169–179 (2014)
20. Valenti, R., Sebe, N., Gevers, T.: Combining head pose and eye location information for gaze estimation. IEEE Trans. Image Process. **99**(1), 1 (2011)
21. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–626 (2009)
22. Cox, M., et al.: CSIRO face analysis SDK. In: 10th IEEE International Conference on Automatic Face and Gesture Recognition (2013)