Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Development of a Pointing-and-Calling Monitoring and Feedback System to Enhance Safety at Pedestrian Crosswalks

Ryonosuke Nakagawa[1†], Wen Liang Yeoh[1], Kenjiro Sakado[2]
Tatsuya Hamamoto[2], Takeshi Sano[2], Osamu Fukuda[1]

[1]Department of Science and Engineering, Saga University, Saga, Japan
(E-mail: ryonosuke1515@gmail.com, {wlyeoh, fukudao}@cc.saga-u.ac.jp)
[2]Toyota Motor Kyushu, Inc, Fukuoka, Japan
(E-mail: {kenjiro_sakado, tatsuya_hamamoto, takeshi_sano}@toyota-kyushu.co.jp)

Abstract: Traffic accidents in industrial environments have increased in recent years. One possible measure to increase safety is through pointing-and-calling to raise safety awareness of employees on pedestrian cross- walks. However, there remains many challenges introducing this in organizations, such as the lack of immediate feedback, per- ceived social awkwardness, and lack of reminders. In this study, we propose that the use of a real-time monitoring and feedback system, which automatically evaluates the pointing-and-calling movement of employees, which can be used to promote this habit, thus improving safety at crosswalks. We performed two experiments evaluating the robustness of the detection systems and the accuracy of the scoring system for the movements.

Keywords: Human-Machine Systems, Factory Automation, Safety, Environment and Eco-Systems

1. INTRODUCTION

1.1. Background

According to Japan's Ministry of Health, Labour, and Welfare, the number of fatalities and injuries in the manufacturing industry in 2023 reached 27,194, the highest among the industries surveyed [1]. Although there is a decreasing trend for workplace accidents such as entanglements (-6.1%) and falls/crashes (-12.8%), traffic accidents in the workplace have increased by as much as +14.7%.

One measure taken to address this is the introduction of a pointing-and-calling safety procedure at pedestrian crosswalks. The pointing-and-calling procedure, originating from the former Japanese National Railways, has been demonstrated to be an effective means of preventing human error in wide-ranging fields, including railways, aviation, construction, medical care [2], [3]. By raising the safety awareness of employees on pedestrian crosswalks through this procedure, it is expected that collisions at crosswalks will be reduced

However, there remain many challenges in instilling this pointing-and-calling habit in organisations. Among them is the lack of immediate feedback on whether they are performing the action correctly or consistently, perceived social awkwardness about performing exaggerated gestures in public, and a lack of reminders. To address this, we propose that the use of a strategically placed real-time monitoring and feedback system, which evaluates the pointing and calling movement of employees, can be used to promote this habit, thus improving safety at crosswalks.

Therefore, in this study, we developed a system which utilises a 2D and 3D pose estimation AI to evaluate the movements of employees with respect to their point-and-

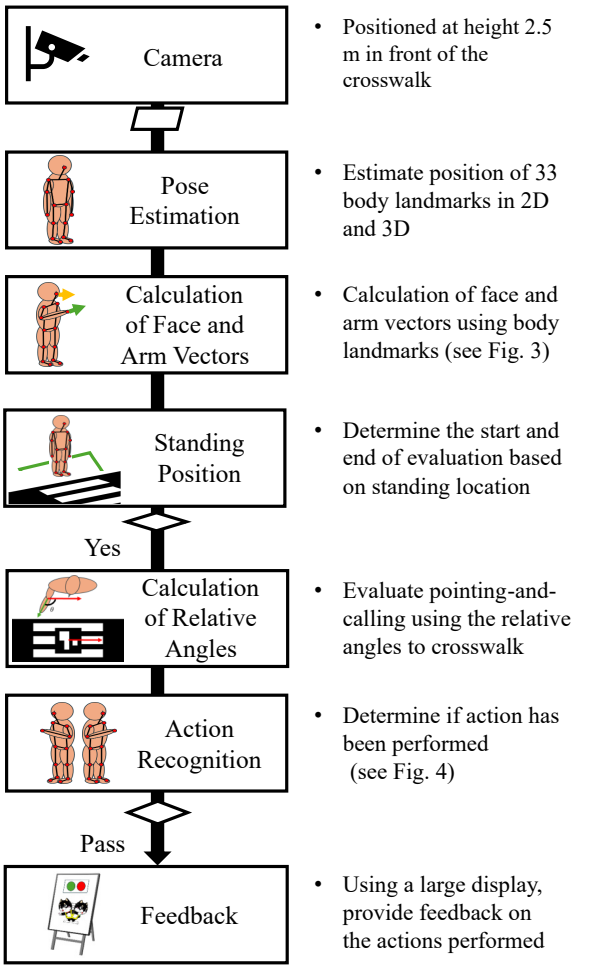


Fig. 1 Flowchart showing the automated evaluation and feedback process of pointing-and-calling action at a crosswalk

† Ryosuke Nakagawa is the presenter of this paper.

calling actions and provides feedback on the correctness of those movements. This can act as a reminder, as well as provide immediate feedback and lessen the social awkwardness of performing the point-and-calling actions, helping promote safety awareness and habits in the organisation. Additionally, we perform two verification experiments to understand the limits of the system and ensure that the system is working as expected.

1.2. Related Works

Today, monitoring technologies are routinely used in industrial environments, but they are rarely used to promote safety [4]. Among the main uses of these monitoring technologies is in the understanding of the causes of various accidents. For example, Mutlu *et al.* used an algorithm to discover sequential patterns that can lead to an accident using the data collected using monitoring technologies [5].

Another possible use of these monitoring technologies in proactive approaches to safety, such as behaviour-based safety [6]. A key component of behaviour-based safety is in observing the actions of employees and providing timely feedback and encouragement to change the habit or behaviour. This can be challenging to do in-person, therefore, an automated feedback system can be effective. Iwasaki *et al.* used a wearable system to observe the pointing-and-calling actions of employees [7]. However, such technologies have drawbacks, including limited battery life and the requirement to equip all employees with the devices.

Therefore, a non-contact monitoring approach is used in this study, where a pose estimation AI is used to observe the actions of an employee. In recent years, the performance of pose estimation AIs has improved substantially [8], [9] and has been shown to be an effective means of observing human movement in fields ranging from sports [10] to rehabilitation [11].

2. SYSTEM

2.1. System Overview

The developed automatic pointing-and-calling valuation and feedback system is shown in Fig. 1. The system consists of a CMOS colour camera (Microsoft LifeCam Studio Q2F-00021), a computer to perform the pose estimation and evaluation, and a large digital signage to provide feedback to the pedestrian. Firstly, a CMOS colour

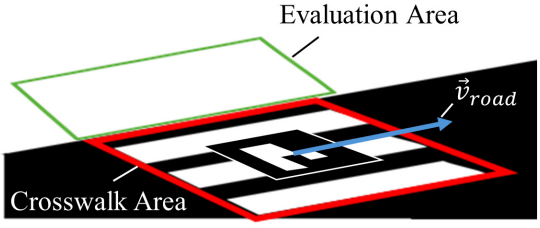


Fig. 2 Information to be provided about the environment from the camera view during system setup

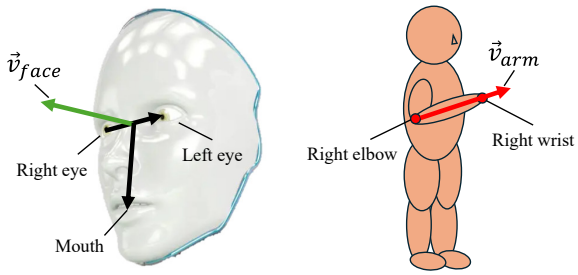


Fig. 3 Body landmarks used to calculate the face vector (the direction the pedestrain is looking), and the arm vector (the direction the pedestrian is pointing)

camera is expected to be placed at a height of 2.5 m and within 30 degrees from the direction facing the crosswalk. The frames obtained by the camera are then processed by a pose estimation AI. Using the estimation results of the pose estimation AI, we are able to estimate the direction the pedestrian's face is facing and the direction the arm is pointing. These directions are then compared to the direction of the road to evaluate if the pedestrian has performed the pointing-and-calling action. Once the pedestrian has performed the pointing-and-calling action or if they attempt to cross the road before completing the actions, feedback will be provided using a large digital signage.

The four actions required to be performed is "Look Left", "Look Right", "Point Left", and "Point Right".

2.2. System Setup

Before operating the system, the relative position and orientation of the camera relative to the crosswalk needs to be obtained. This was done by placing an ArUco on the crosswalk, with its direction aligned with the road, such that a road vectors, $\vec{v}_{road}$ designating the direction of the road relative to the camera can be obtained, as shown in Fig. 2.

Additionally, the 2D area in the camera view which a pedestrian or employee is expected to perform the pointing-and-calling action is designated. This area is labelled as "Evaluation Area" in Fig. 2. The location of the crosswalk, labelled "Crosswalk Area" in Fig. 2, is also designated to determine when the pedestrian or employee attempts to cross the road before performing the pointing-and-calling action.

2.3. Pose Estimation

MediaPipe Pose [8] was used as the pose estimation AI in this system. It is capable of estimating the 3D coordinates of 33 body landmarks of pedestrians within the view of the camera. The position estimated body landmarks can be expressed as normalised coordinates, showing the position of the pedestrian in the image, as well as in terms of world coordinates where the real-world position relative to the hip in meters is estimated. The estimated body landmarks are represented by x-coordinates and y-coordinates in the image, as well as z-coordinates, which are depth information. The model was able to per-

form the pose estimation at a frame rate of 30 fps from a camera frame resolution of 1920 ×1080 pixels.

2.4. Standing Position

The state of the system changes based on the position of the pedestrian. There are three main states, namely, a (1) No person state, where there was no one in both the evaluation area and the crosswalk area, a (2) Evaluating state, where the pedestrian is in the evaluation area and the system is checking if pointing-and-calling is being performed, and a (3) Crossing road state, where the pedestrian is in the process of crossing or is starting to cross the road. The states are determined by the area in which the pedestrian is standing. This is detected using the 2D coordinates of both the pedestrian's feet estimated using the pose estimation AI described in Section 2.3, and compared to the evaluation area and crosswalk area described in Section 2.2.

In the (1) No person state, the system waits for a pedestrian and performs no actions. In the (2) Evaluating state, the systems keeps track of whether the pedestrian has performed each of the required pointing-and-calling actions. In the (3) Crossing road state, the tracking parameters are reset and the evaluated performance is fed back to the pedestrian using the large digital signage.

2.5. Face and Arm Vectors

To determine whether the pedestrian has performed the pointing-and-calling action, we first need to obtain the direction the face is facing and the direction the arm is pointing. This is calculated using the body landmarks estimated in Section 2.3. The landmarks used for the face vector, $\vec{v}_{face}$, and the arm vector, $\vec{v}_{arm}$, are shown in Fig. 3.

The face vector, $\vec{v}_{face}$, the direction the pedestrain is looking is calculated using four points: left eye, right eye, left mouth, and right mouth. First, a horizontal vector from the right eye to the left eye and a vertical vector from mid-eye (mid-point of left eye and right eye) to mid-mouth (mid-point of left mouth and right mouth) is calculated. The outer product of these two vectors is the denoted the face vector, $\vec{v}_{face}$, as shown in Eq. 1.

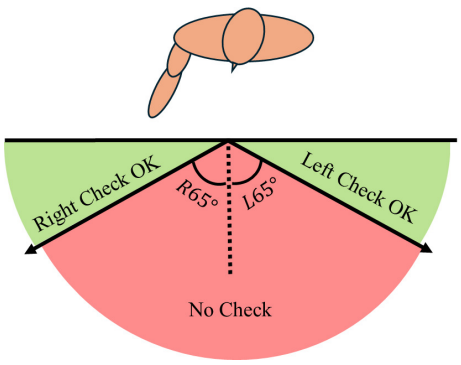


Fig. 4 The angle at which the pedestrain is deemed to have performed the pointing and looking action appropriately (set at 65°in this study)

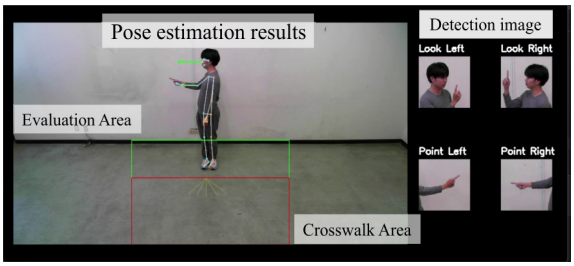


Fig. 5 Camera view and the four pointing-and-calling actions recognised

$$\vec{v}_{face} = \left(\vec{v}_{left\,eye} - \vec{v}_{right\,eye}\right) \times \left(\vec{v}_{mid\,mouth} - \vec{v}_{mid\,eye}\right) \tag{1}$$

We assume that the pedestrian will perform the pointing-and-calling action using their right hand. Therefore, the direction in which the pedestrian is pointing at can be calculated using two body landmarks, the right elbow and the right wrist. The arm vector, $\vec{v}_{arm}$, is calculated using Eq. 2.

$$\vec{v}_{arm} = \vec{v}_{right\,wrist} - \vec{v}_{right\,elbow} \tag{2}$$

2.6. Relative Angles

To determine if the pedestrian has performed the pointing-and-calling action appropriately, we refer to the relative angle between the face and the arm vectors, and the road vector. These angles are calculated using the inner product as shown in Eq. 3.

$$\theta = \arccos\left(\frac{\vec{v} \cdot \vec{v}_{road}}{\|\vec{v}\|\,\|\vec{v}_{road}\|}\right) \cdot \frac{180}{\pi} \tag{3}$$

2.7. Action Recognition

In every frame, the relative angle between the face and the arm vectors, and the road vector is compared to a preset threshold angle. For each action, if the relative the angle exceeds the threshold value, as shown in Fig. 4, the pedestrian is deemed to have performed that action.

This threshold angle can be varied based on the circumstance and environment the system is being used in. The effective visual field range is known to vary depending on the individual and the surrounding environment. In this study, the threshold of 65°was chosen based on a previous study that showed the limits of the effective field of view during standing and walking to be between 25°and 30°in the lateral direction [12].

An interface was created to enable the system user to easily check whether the pedestrian performed the pointing-and-calling action. The screen of the developed interface is shown in Fig. 5. The large image to the left shows the camera view with the pose estimation results and the face and arm vectors overlaid. Additionally, the evaluation area and crosswalk area are shown

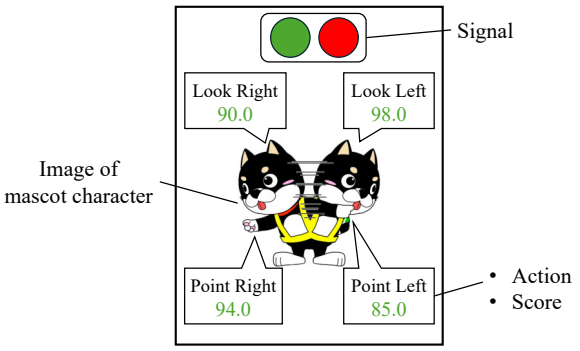


Fig. 6 The feedback information of the pointing-and-calling evaluation to be displayed on the large digital signage

using a green and a red bounding box respectively. If the pedestrian is deemed to have performed the pointing-and-calling action specified when in the evaluation area, a cut-out of the action is shown on the right. If no pointing or looking action is performed, nothing is displayed cut-out area. As described in Section 2.4, the tracking of the actions will be reset once the pedestrian enters the crosswalk area. The interface allows us to easily check, at a glance, which of the four actions (Look Left, Look Right, Point Left, Point Right) has been detected.

2.8. Feedback

The results of the system evaluations are fed back to the pedestrian using a large 65-inch digital signage (JapanNext, JN-IPS60UHDR-M). Fig. 6 shows one version of the feedback screen to be displayed. In the centre of the screen, an illustration of a mascot character performing a left-right visual check and pointing check is displayed. Around it, the scores for each judgment item are displayed. The colour of the score is initially displayed in red but changes to green when the set threshold is exceeded. At the top of the screen, there is a circular signal indicating the judgment result, which is normally lit red. However, when the scores for all check items exceed the threshold value, the red signal is switched off, and the green signal is switched on. This allows the safety checking behaviour of the evaluation target to be visualised in real time and can be used to provide behavioural feedback and educational support for the employees. It is expected that this will not only provide feedback on their pointing-and-calling actions, but also help remind and encourage the employees to perform the actions, as well as normalising the behaviour to minimise social stigma. The score is calculated using Eq. 4.

$$\text{score} = \begin{cases} 100 & \text{if } \theta \geq \text{threshold} \\ \dfrac{100}{\text{threshold}} \cdot \theta & \text{if } 0 < \theta < \text{threshold} \\ 0 & \text{if } \theta \leq 0 \end{cases} \quad (4)$$

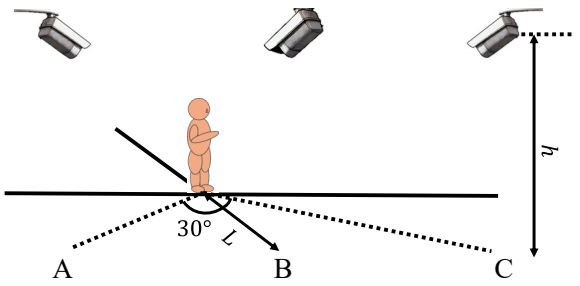


Fig. 7 Experimental conditions for experiment 1: Robustness of evaluation results to camera installation position.

3. VERIFICATION EXPERIMENTS

Two experiments were performed to verify the robustness of the systems and the evaluation approach used.

Firstly, depending on the installation position of the camera, the pose estimation accuracy and the accuracy of the relative angle calculated might vary due to visibility and depth estimation issues. To determine if this could lead to problems with our evaluation system, we investigated the robustness of the system with respect to the camera's installation position (Experiment 1).

Secondly, to provide feedback with respect to the correctness of their movements, it is important to not only determine whether the action has been performed, but be able to provide a score on how well it has been performed. Based on that score, able to instruct the pedestrian or employee on areas and ways to improve. Experiment 2 investigated whether the scores evaluated matched the quality of the actions performed.

3.1. Experiment 1

In this experiment, we evaluated the robustness of the evaluation results to camera installation position.

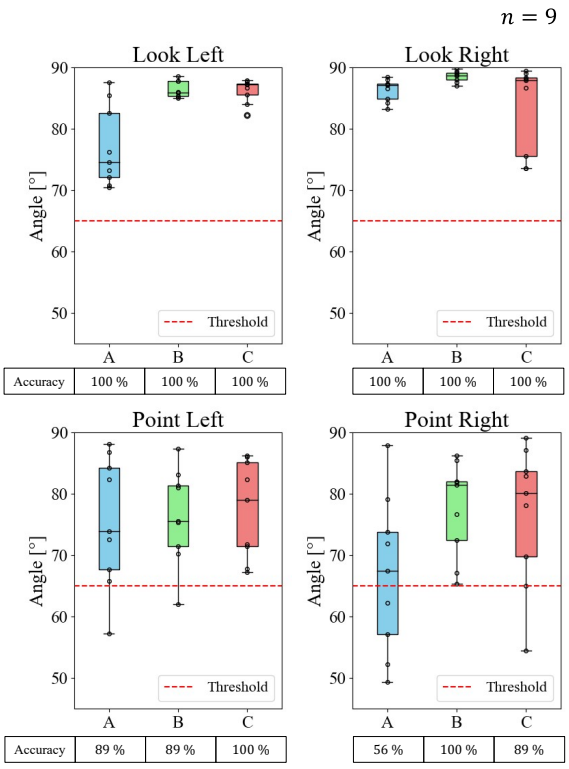
3.1.1. Methods

This was performed by using three identical cameras to simultaneously record participants performing the pointing-and-calling action. All cameras were installed at a distance of 3.5 m in the direction of travel (L in Fig. 7) and at a height of 2.5 m (h in Fig. 7) from the location where the pointing-and-calling action would be performed. The position of the three cameras representing the experimental conditions was positioned at: 30 °to the left of the direction of travel (A in Fig. 7), in line with the direction of travel (B in Fig.7) and 30 °to the right of the direction of travel at (C in Fig. 7).

Three young healthy male participants (22.3 ±0.47) were instructed to perform two types of actions, 'with pointing and looking before crossing' and 'without pointing and looking before crossing' for three trials each.

3.1.2. Results

When participants were instructed to use the crosswalk 'without pointing and looking before crossing' , all three cameras had accuracy of 100% for all trials and participants. The system was able to correctly determine when



* 100% accuracy for all cases of
'without pointing and looking before crossing'

Fig. 8 Boxplot with data points showing the results of experiment 1. The maximum angle inside the evaluation area when performing the actions when the participants were instructed to use the crosswalk 'with pointing and looking before crossing'

the participant did not perform the pointing and looking actions, and there was no false positive result.

Fig. 8 shows the maximum relative angles to the direction of the road obtained by the three camera installation positions for each action. For visual check, that is 'looking before crossing' , with the parameter obtained using the face vector, the accuracy for 100% for all installation locations. However, for 'pointing before crossing' , there were occasions where the system evaluated the participant to not have performed the actions even when they had. Particularly at the installation location A (30°to the left), the accuracy was low, with only five out of nine attempts for pointing confirmation evaluated correctly.

3.2. Experiment 2

The aim of experiment 2 was to investigate whether the scores evaluated matched the quality of the actions performed.

3.2.1. Methods

In this experiment, the participants were the same as those shown in Section. 3.1.1, and the camera was set up at position B in Fig. 7. The extent of head turning and arm pointing to the left and right is used as a proxy for the quality of the pointing and looking action performed

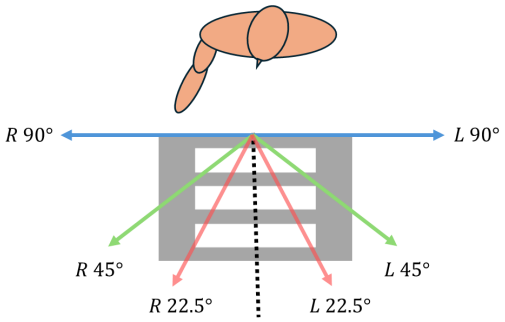


Fig. 9 Experimental conditions for experiment 2: Quality of the actions performed by participants (instructions to given to participants in regard to the extent of head turning and arm pointing to the left and right)

by the participants. As shown in Fig. 9, the participants were instructed to perform the pointing and looking action to three different extents, as well as a 'No Check' condition where they neither look nor pointed left and right. Six different conditions were investigated, namely, 90 °, 45 °, 22.5 °, and 'No Check' . Similar to the previous experiment, the participants performed three trials for each condition. The direction from which the participant enters to evaluation area was randomised.

3.2.2. Results

Fig. 10 shows the evaluated scores for different qualities of the looking and pointing action. It was confirmed that the scores for all the judgment items were evaluated as smaller in the order of 90 °, 45 °, 22.5 °and No Check.

If the participants approach the crosswalk from the left, the systems evaluate them to have looked right because that is the direction they are facing. The same applies to when the participant approaches from the right. Therefore, the scores for 90°and No Check conditions were in line with the expected scores.

4. DISCUSSION

From experiment 1, it was confirmed that the accuracy of the pointing evaluation was low when the camera was installed in the left 30°. This might be because all participants used their right hand to perform the pointing. This can make it difficult for the pose estimate AI to accurately estimate the depth of the right hand landmarks from the left side.

From experiment 2, the distribution of the scores was not as even as expected. Particularly for pointing left as shown in Fig. 10, there is little to no difference in the scores for 45°and 22.5°. This might be due to when the participants were using their right hand to point left, they have to twist and close their body, which can affect the visibility and depth estimation of the landmarks.

5. CONCLUSION

We developed a pointing-and-calling monitoring and feedback system to enhance safety at pedestrian crosswalks. In this study, we demonstrated that we are able

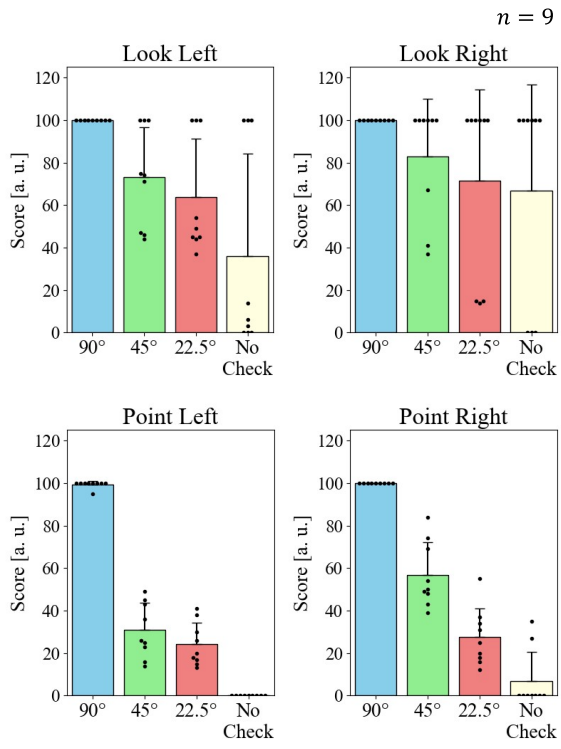


Fig. 10 Bar plots with data points showing the results of experiment 2: The evaluated score at different quality of action

to reliably determine whether a pedestrian performed the pointing and looking actions required before crossing a road. Nonetheless, some challenges remain. Specifically, accuracy can drop when the camera is placed on the opposite side of the pointing hand, and the left-right difference needs to be considered when scoring the movement due to posture. As future prospects, we aim to improve the accuracy by improving the pedestrian sensing method, comparing the accuracy with other posture estimation models, and introducing a model suitable for the application of this system. The usefulness of the system will be confirmed by actually introducing the system in a factory.

REFERENCES

[1] L. Ministry of Health and Welfare, *Occupational accidents (in japanese)*, https://www.mhlw.go.jp/content/11302000/001099504.pdf, 2023.

[2] H. Watanabe, J. Kawaguchi, Y. Nakai, T. Tsukada, M. Hikono, and H. Nakamura, "Cognitive psychological approach for an effective shisa-kosyo in a field depended on human memory function," *The Japanese journal of ergonomics*, vol. 41, no. 4, pp. 237–243, 2005.

[3] S. Haga, "Effect of finger pointing on eye movement," *The Japanese journal of ergonomics*, vol. 43, no. 2 Supplement, pp. 140–141, 2007.

[4] G. Kortuem et al., "Sensor networks or smart artifacts? an exploration of organizational issues of an industrial health and safety monitoring system," in *UbiComp 2007: Ubiquitous Computing*, J. Krumm, G. D. Abowd, A. Seneviratne, and T. Strang, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 465–482.

[5] N. G. Mutlu, S. Altuntas, and T. Dereli, "The evaluation of occupational accident with sequential pattern mining," *Safety Science*, vol. 166, p. 106 212, 2023.

[6] S. Carra, E. Bottani, G. Vignali, M. Madonna, and L. Monica, "Implementation of behavior-based safety in the workplace: A review of conceptual and empirical literature," *Sustainability*, vol. 16, no. 23, p. 10 195, 2024.

[7] M. Iwasaki and K. Fujinami, "Recognition of pointing and calling for industrial safety management," Apr. 2012.

[8] C. Lugaresi et al., "Mediapipe: A framework for perceiving and processing reality," 2019.

[9] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *CoRR*, vol. abs/2006.10204, 2020.

[10] X. Xi, C. Zhang, W. Jia, and R. Jiang, "Enhancing human pose estimation in sports training: Integrating spatiotemporal transformer for improved accuracy and real-time performance," *Alexandria Engineering Journal*, vol. 109, pp. 144–156, 2024.

[11] Y. Qiu, J. Wang, Z. Jin, H. Chen, M. Zhang, and L. Guo, "Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training," *Biomedical Signal Processing and Control*, vol. 72, p. 103 323, 2022.

[12] D. Saito, "Change in effective visual field using smartphone with walking," *Journal of Biomedical Fuzzy Systems Association*, vol. 23, no. 1, pp. 63–68, 2021.