OXFORD CDE

Journal of Computational Design and Engineering, 2024, **11**, 158–173

DOI: 10.1093/jcde/qwae042
Advance access publication date: 6 May 2024
Research Article

Integration of eye-tracking and object detection in a deep learning system for quality inspection analysis

Seung-Wan Cho[1,†], Yeong-Hyun Lim[1,†], Kyung-Min Seo[2,*] and Jungin Kim[3,*]

[1]Department of Industrial & Management Engineering, Hanyang University, 222, Wangsimni-ro, Seong-dong-gu, Seoul 04763, Republic of Korea
[2]Department of Industrial & Management Engineering, Hanyang University ERICA, 55, Hanyangdaehak-ro, Sangnok-gu, Ansan 15588 Gyeonggi-do, Republic of Korea
[3]AI Research, Hankyong National University, 283, Samnam-ro, Pyeongtaek 17738 Gyeonggi-do, Republic of Korea
*Correspondence: kmseo@hanyang.ac.kr (K.-M.S.); jik@hknu.ac.kr (J.K.)
†Equal contribution

Abstract

During quality inspection in manufacturing, the gaze of a worker provides pivotal information for identifying surface defects of a product. However, it is challenging to digitize the gaze information of workers in a dynamic environment where the positions and postures of the products and workers are not fixed. A robust, deep learning-based system, ISGOD (Integrated System with worker's Gaze and Object Detection), is proposed, which analyzes data to determine which part of the object is observed by integrating object detection and eye-tracking information in dynamic environments. The ISGOD employs a six-dimensional pose estimation algorithm for object detection, considering the location, orientation, and rotation of the object. Eye-tracking data were obtained from Tobii Glasses, which enable real-time video transmission and eye-movement tracking. A latency reduction method is proposed to overcome the time delays between object detection and eye-tracking information. Three evaluation indices, namely, gaze score, accuracy score, and concentration index are suggested for comprehensive analysis. Two experiments were conducted: a robustness test to confirm the suitability for real-time object detection and eye-tracking, and a trend test to analyze the difference in gaze movement between experts and novices. In the future, the proposed method and system can transfer the expertise of experts to enhance defect detection efficiency significantly.

Keywords: quality inspection, eye-tracking, object detection, deep learning, system integration

1. Introduction

In the manufacturing field, gaze data, such as eye-tracking points captured by workers' gazes, are pivotal during quality inspection tests (Mark *et al.*, 2021; Zheng *et al.*, 2022). Such gaze data facilitate the efficient identification of defects and play a crucial role in determining the sequence of assembly and machinery operations (Lušić *et al.*, 2016). For this reason, many researchers have analyzed the pattern and sequence of human eye movements and collected gaze data such as gaze fixation, dwell time duration, and fixation count (Cristino *et al.*, 2010; Kanan *et al.*, 2015; Ooms *et al.*, 2012; Wang *et al.*, 2022). The collected data have been utilized for the transfer of skills and know-how among workers, contributing to more efficient manufacturing operations; however, this transmission has often been informal and unstructured between individuals (Nakamura *et al.*, 2019; Ye *et al.*, 2023).

Recently, to overcome these shortcomings, researchers have shifted their focus toward digitalizing gaze data (Ahrens *et al.*, 2023; Borgianni *et al.*, 2018; Ghanbari *et al.*, 2021; Ramachandra *et al.*, 2021; Ren *et al.*, 2023; Takahashi *et al.*, 2018). For example, Sadasivan *et al.* (2005) utilized eye movement to pre-train operators in the aircraft inspection process. Lusčić *et al.* (2016) delved into the distinctions between static and dynamic contexts by analyzing eye-tracking data during manual product assembly processes. However, in the dynamic realm of manufacturing, merely tracking an individual's gaze is insufficient in the following two ways. First, effective data collection requires eye-tracking and the identification of specific points on the object. Next, the collection and analysis of gaze data must consider the changing positions and postures of both the workers and objects. These two problems hinder the immediate utility of data and complicate real-time and dynamic applications.

An effective analysis of eye-tracking data must encompass the detection of moving objects and the changing conditions present in manufacturing settings (Praveen *et al.*, 2010). To this end, this study proposes a robust deep learning system named ISGOD (Integrated System with worker's Gaze and Object Detection), which performs object detection with eye-tracking to determine which part of the object is viewed. ISGOD is composed of the following four modules: (i) collecting the module of image and gaze data for an eye-tracking device, (ii) the detection module of objects from the captured images, (iii) the integration module between eye-tracking data and object detection data, and (iv) an analysis module based on the proposed evaluation matrix.

For the eye-tracking module, Tobii Pro Glass 3 was used to collect human eye movements, encompassing gaze points, fixations, viewing duration, head movement, and orientation (T. H. Li *et al.*, 2020). The object detection module facilitates the six-dimensional (6D) pose estimation algorithm, which achieves high accuracy

Received: January 29, 2024. **Revised:** April 29, 2024. **Accepted:** April 29, 2024

Table 1: Comparison of literary review.

Related research	Type	Characteristics	Limitations
Niemann *et al.* (2019)	Eye tracking	Analysis of quality inspection in automotive manufacturing using gaze data	Necessitates subsequent human visual analysis
Ulutas *et al.* (2020)	Eye tracking	Analysis of eye-tracked data of quality assurance workers using Hidden Markov Models	Limited to experimentation in static settings
Jonas *et al.* (2021)	Eye tracking	Analysis of the impact of cleanliness in aircraft part visual inspection using gaze data	Limited to experimentation in static settings
Bukschat *et al.* (2020)	Object detection	Development of a model to estimate position and orientation of objects in 3D space	Requirement for large volumes of training data.
Sampaio *et al.* (2021)	Object detection	Development of a systematic method to effectively train object detection models	Requirement for detection of dynamic entities

across objects of varying sizes and offers relatively swift computational efficiency, rendering it suitable for real-time detection applications (Jamie *et al.*, 2013). The algorithm is also well suited for analyzing gaze data on objects by estimating their states, including their orientation, rotation, and location. An integrated module was developed that consolidates the eye-tracking module with the object detection module, addressing any arising integration challenges. The real-time application feasibility can be enhanced by resolving latency issues through algorithm warm-up and frame sampling methods. Finally, an analyzer module was designed to handle objects detected at varying sizes and rotational angles under dynamic conditions. A perspective transformation algorithm was employed to streamline the analysis process, unifying the detection of objects of varying sizes within a dynamic environment.

The evaluation matrix comprised metrics such as the gaze score, accuracy score, and concentration index for comprehensive analysis. The gaze score evaluates how far the gaze is from the point of interest, and the accuracy score is the average of the gaze scores across multiple points of interest. If the gaze aligns precisely with the point of interest, it receives a high score, and the score decreases as the distance from the point of interest increases. The concentration index measures the ratio of the gaze falling within the region of interest (ROI) of the point of interest to the overall gaze.

Two experiments were conducted to demonstrate the robustness of the proposed system and to evaluate the discrepancy among workers. The first experiment, conducted to evaluate the robustness of the system, organized the assessment into six scenarios with four subjects, structured around the presence of either one or four defect points. Each scenario was compared quantitatively using an evaluation matrix. In this matrix, the gaze score is the primary measure, and the robustness of the system is determined by analyzing the average and variation in this score. For example, these evaluation matrices did not show significant differences, such as variances of 0.01, 0.05, and 0.07 for each case in Scenario 1, and it was confirmed that there was no significant difference even on the visualized heat map.

The next experiment analyzed the differences in gaze patterns between four novice and four professional workers. The position of the novice and professional gazes was expressed as a graph over time, resulting in a difference in gaze according to skill level. In practical applications, significant disparities exist in the defect detection methods employed by novices and professional workers. This variation can be attributed to professionals learning more efficient and accurate gaze routes over time compared to novices. Through such gaze data collection, the strategies of professionals can be effectively communicated to novices (Nakamura *et al.*, 2019). Therefore, stable gaze data collection in dynamic environments is essential.

These experiments ensured a stable collection of gaze data in dynamic manufacturing environments, enabling gaze analysis through object detection. By facilitating the real-time analysis of gaze points without the need for post-processing, two key issues were addressed: the inability to detect dynamic objects and the challenge of synchronizing gaze tracking with object location. This advancement not only resolves these existing problems but also paves the way for further analysis and application of expert gaze data.

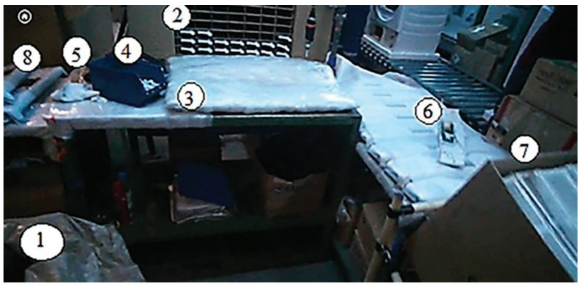
The remainder of this paper is structured as follows: Section 2 reviews related works in the manufacturing field, and Section 3 describes the proposed system architecture, which is segmented into four modules. Section 4 discusses the experiments and results, showcasing the robustness and adaptability of the system. Section 5 concludes the paper.

2. Related Work

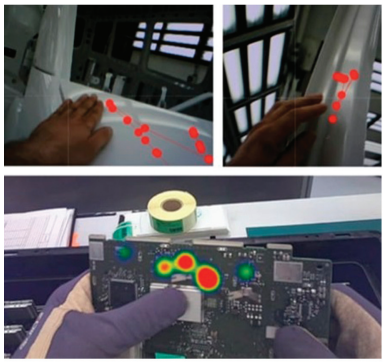
Below, Table 1 summarizes important results from major studies on gaze tracking and object detection, and it reviews the features and drawbacks of these studies. Additionally, Fig. 1 shows a photograph related to the research.

In the manufacturing domain, one of the primary applications of gaze data and eye-tracking technology is quality inspection. Niemann *et al.* (2019) leveraged gaze data to enhance inspection procedures during the painting stage of production. Their research enabled workers to optimize the sequence of operations by examining the fixation order during inspections, as depicted in Fig. 1a. Additionally, this study facilitates the identification and improvement of inspection processes, highlighting areas commonly overlooked by inspectors. However, the gaze data collection process requires subsequent manual verification to ascertain which parts of the object have been observed. This complicates the analysis of continuous data streams and requires substantial human intervention.

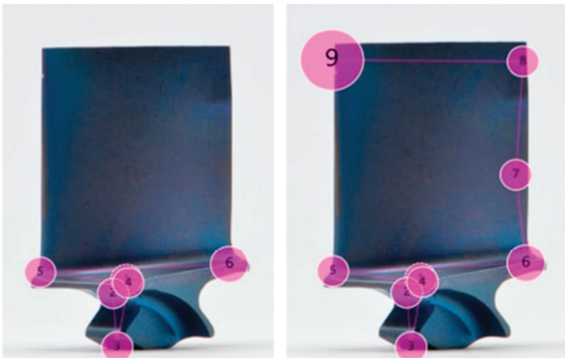
Ulutas *et al.* (2020) analyzed eye-tracking data gathered from quality inspection personnel. The data were collected using an eye-tracking device during the inspection of various plastic control panels of the tumble dryer, as shown in Fig. 1b. The study delineated the differences between novice and professional inspectors by evaluating recorded eye-movement patterns. Furthermore, it was confirmed that there is a clear difference in eye-movement patterns between experts and novices. The analysis of visual engagement within specified areas of interest (AOIs) employs sophisticated methods, such as the Hidden Markov Model. However, a limitation of the experiment was the



(a) Quality control area from the point of an operator's view



(b) Employee's eye movement in assembly line



(c) Engine blade fault detection process

Figure 1: Quality inspection cases using eye-tracking skills.

static design and testing of AOIs, which presents challenges for extrapolation to dynamic real-world scenarios in which both objects and individuals are in motion.

Jonas *et al.* (2021) implemented factor analysis utilizing eye-tracking technology to examine the process of visually inspecting aircraft parts. This investigation involved 50 professionals from the industry who scrutinized the images and assessed the depicted features, as illustrated in Fig. 1c. This study specifically focused on the variations in visual attributes observed during the inspection of both clean and dirty blades. It was posited that cleaning blades prior to inspection significantly influenced the visual inspection outcomes. However, note that this experiment was conducted using static images. This limitation raises the possibility that the findings may differ in the dynamic context of visual inspection. To effectively gather eye-tracking data in the dynamically changing conditions of a manufacturing site, it is crucial to discern not only the object being observed but also the specific part of the object under scrutiny. Previous research has predominantly focused on *post hoc* analysis; however, integrating object detection can address this challenge.

Object detection algorithms must possess the capability not only to discern the location of objects but also to determine their orientation and angle of rotation accurately. Bukschat *et al.* devised an EfficientPose algorithm that estimates the position and orientation of objects within 3D spaces (Bukschat *et al.*, 2020). However, this algorithm requires extensive training data for effective learning. The insufficient amount of data from the Linemod benchmark dataset was supplemented using data augmentation (Hinterstoisser *et al.*, 2011). However, if the amount of data cannot be increased in this manner, a new dataset must be created manually, including labeling the orientation of objects, which is a challenging task requiring considerable time and workforce. Sampaio *et al.* employed Computer-Aided Design (CAD) models to produce synthetic images that were dynamic in the real world, thereby streamlining the training process for object detection models and simplifying data acquisition across diverse fields (Sampaio *et al.*, 2021).

This study introduces an integrated object detection and eye-tracking system designed to utilize worker gaze data directly, eliminating the requirement for additional post-processing. For object detection, the system employs a 6D pose estimation algorithm that can detect the direction, rotation, and positions. Furthermore, Unity creates a comprehensive dataset of training image backgrounds that reflect real-world environments (Lee *et al.*, 2021). This method allows the proposal of a system capable of reliably collecting and analyzing gaze data even in dynamic manufacturing environments.

3. Proposed System Architecture

This section describes the overall structure of the proposed system for analyzing worker gazes. The system dynamically detects moving objects and integrates the worker's gaze coordinates to determine the specific part of the object on which the gaze is focused. The overall structure of the proposed system is illustrated in Fig. 2. The proposed system incorporates four distinct modules for digitizing the position of the target object and the operator's gaze information. The integration module introduces a method for minimizing latency. In addition, the analysis module describes an algorithm designed to compare and analyze the positions of various objects in 3D space.

First, the target object was observed using a wearable device, Tobii Pro Glasses3. The device is responsible for storing and transmitting real-time video and acts as an eye-tracking module. The eye-tracking results are transmitted to the integration module, whereas the real-time video data are forwarded to the object detection module to identify objects within the video stream.

The object detection module employs a 6D pose estimation algorithm, which is a modification of the EfficientPose algorithm, for seamless detection in dynamic environments. The algorithm is trained on the target object before being used in the system and utilizes both real-world and implemented images for training. The

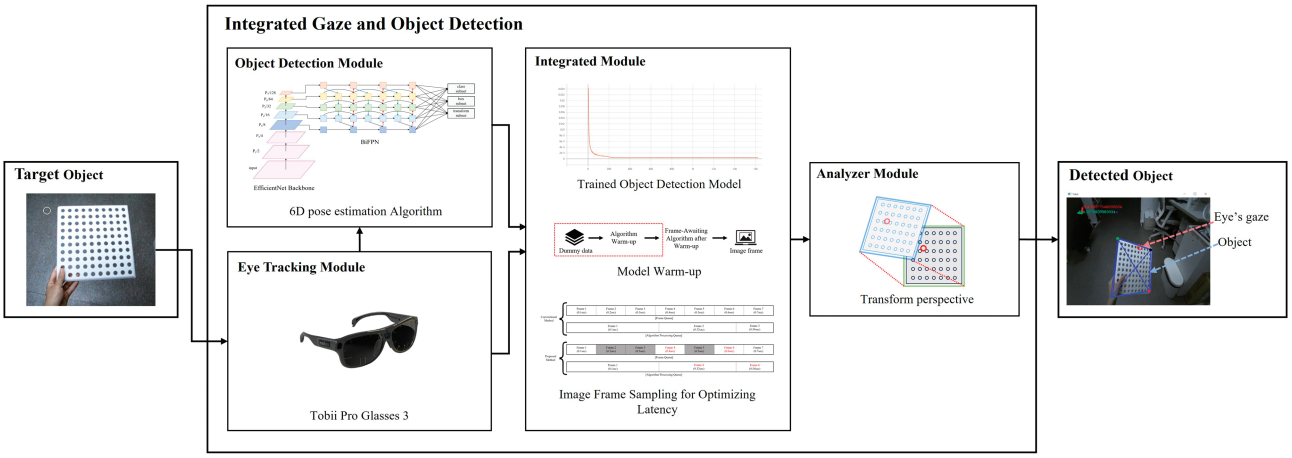


Figure 2: ISGOD structure.

module processes the received images for object recognition and subsequently sends the identified information to the integration module. Consequently, within the integration module, the results of object detection and gaze tracking obtained from these two systems were combined to finalize the development of the proposed system.

In addition, an analysis module is necessary when using the proposed system for analysis. As dynamic objects are detected, the coordinates, angles, rotation of the object, and the relative position of the gaze at that moment vary. To address this, an additional algorithm that standardizes the size and position of all objects was implemented, enabling more accurate gaze comparisons. In addition, owing to its modular architecture, the system offers ease for future modifications or redesigns (Kang *et al.*, 2021).

3.1. Object detection module

There is a general tendency to use depth data collected with an RGB-D camera to estimate the 6D pose (He *et al.*, 2020). However, in this study, a method of directly estimating the 6D pose using only RGB data was used for fast real-time recognition (Son & Ko, 2022; Yin *et al.*, 2021). In the proposed system, object detection is achieved by implementing an algorithm based on the EfficientPose framework. EfficientPose, a detailed deep learning architecture, is capable of determining the class of single or multiple objects within a single-shot RGB image while also estimating their 2D bounding boxes and rotational angles (roll, pitch, and yaw) across the three axes.

The algorithm operates as follows: initially, it acquires an input image from a camera or another imaging device to extract features from this image. This phase is critical for analyzing and interpreting the shape and structural attributes of an object to obtain essential information. Subsequently, using these extracted features, the algorithm estimates the 6D pose of an object in Fig. 3. The term "6D" pertains to both the position and rotation within a 3D space, thereby defining the spatial coordinates and the directional orientation of the object. After successfully determining the position and orientation of the object, the algorithm finalizes the object detection process by computing the bounding box of the object.

3.1.1. *6D pose estimation architecture*

This study used a modified EfficientPose model architecture to estimate the 6D pose by reflecting the structural characteristics of an object, as shown in Fig. 4 (J. Y. Kim *et al.*, 2022). This architecture includes an EfficientNet backbone, a bidirectional feature

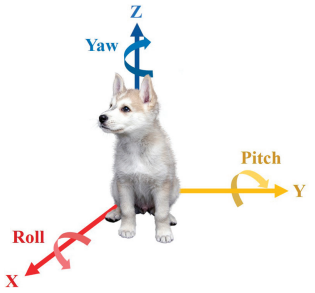


Figure 3: 6D pose: X, Y, Z, yaw, pitch, and roll.

pyramid network (BIFPN), and lower subnetworks. The backbone employs EfficientNet, a convolutional neural network architecture renowned for its superior accuracy and computational efficiency relative to existing ConvNet models (Z. Li *et al.*, 2021). In fact, EfficientNet-B7 recorded a top-1 accuracy of 84.4% and a top-5 accuracy of 97.1% on the ImageNet dataset, demonstrating its capability to realize a model that is 8.4 times smaller in size and 6.1 times faster than traditional ConvNet architectures (Tan & Le, 2019). For the neck, a BIFPN is utilized to enhance the detection accuracy of objects of various sizes. The head comprises a classifier for object type recognition, a bounding box for determining the position of the object, and a regressor for angle estimation.

The complete loss function utilized for training the 6D pose estimation architecture, which is specifically engineered for 6D pose recognition, is composed of three distinct components: L_{class} for classification loss, L_{bbox} for bounding box regression loss, and L_{TR} for transformation loss, as explained in equation (1). Furthermore, the variable influence of each constituent loss is regulated by the hyperparameter λ.

$$\text{Loss} = \lambda_{class} \cdot L_{class} + \lambda_{bbox} \cdot L_{bbox} + \lambda_{TR} \cdot L_{TR}. \quad (1)$$

The classification loss L_{class}, used in EfficientDet to classify the classes of objects is a modified cross-entropy loss function known as the Focal Loss function (Tan *et al.*, 2020). It was developed to address the class imbalance problem, which is a challenge in model training where the "negative" class significantly outnumbers the "positive" class. The estimated probability P_t corresponds to the likelihood that a given instance is classified as a foreground class by a deep learning model. The term a_t is a weight that balances the positive and negative classes. In addition, a modulating factor expressed as $(1 - P_t)^y$ is incorporated to mitigate the imbalance

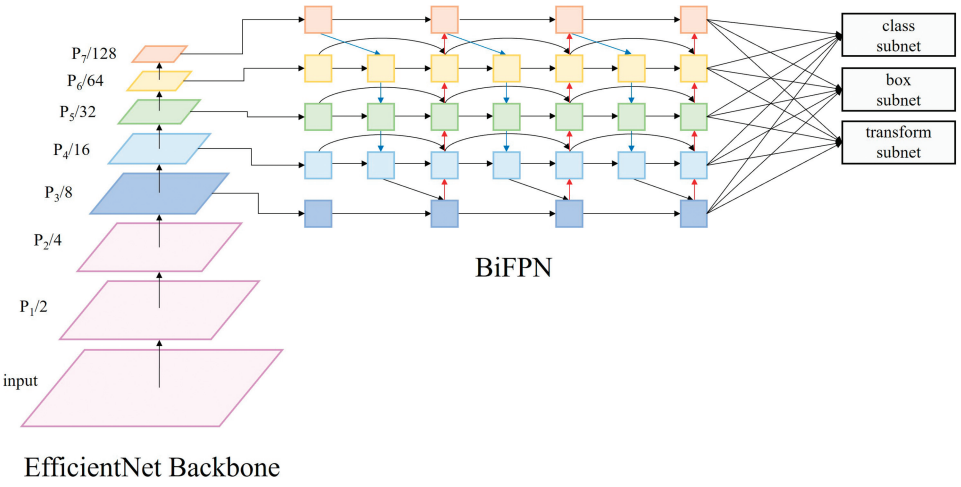


Figure 4: 6D pose estimation network architecture.

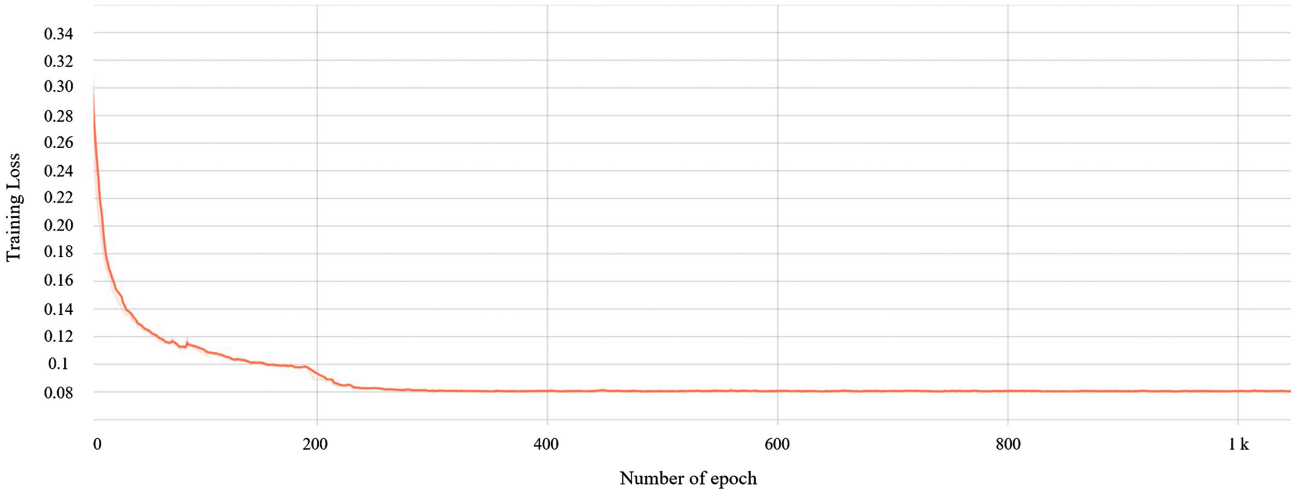


Figure 5: 6D pose estimation algorithm loss graph.

between straightforward and complex examples, thereby refining the focus of the model and improving its performance for more challenging instances within the training data.

$$L_{\text{class}} = FL(P_t) = -a_t(1-P_t)^{\tau}\log(P_t), \quad P_t = \begin{cases} p & if\ y = 1 \\ 1-p & otherwise \end{cases}. \tag{2}$$

The bounding box regression loss, represented by L_{bbox} employs a smooth $L1$ loss, which enhances the precision of object localization. In the context of bounding box regression for the target class u, let t represent the ground-truth values and v denote the corresponding predicted values. This relationship is represented by the following equations: $t^u = (t_x^u,\ t_y^u,\ t_w^u,\ t_h^u)$, $v = (u_x,\ u_y,\ u_w,\ u_h)$.

$$L_{\text{bbox}}(t^u, v) = \sum_{i\in x,y,w,h} smooth_{L1}(t_i^u - v_i),\ smooth_{L1}(x) \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases}. \tag{3}$$

Transformation loss, L_{TR} defined by an $L2$ loss framework, is utilized to recognize the pose of the object, which is expressed as relative positional coordinates and rotational angles in space with reference to the coordinate system of the camera. The translation vector T_f is a 3D vector from set R^{3x1} containing elements t_x, t_y, and t_z, which are aligned with the ground truth position of the object. Similarly, the rotation vector R_f, expressed in the compact Rodrigues form within R^{3X1}, encapsulates the ground truth orientation. The translation and rotation vectors, $\overline{T_f}$ and $\overline{R_f}$, respectively, denote the predicted pose parameters that are essential for the estimation process of the model. Furthermore, the set M_f consists of 3D model points, typically presented as point-cloud data, whereas m_f represents the count of these points, which are factored into the loss computation.

$$L_{\text{TR}} = \frac{1}{m_f}\sum_{M_1\in M_f}\min_{x_2\in M_f} |\left(R_f x_1 + T_f\right) - \left(\bar{R}_f x_2 + \bar{T}_f\right)|^2. \tag{4}$$

During training, using the given loss function, we saw the loss values drop over time, as Fig. 5 shows. This steady decrease in loss means the algorithm is getting better at making accurate predictions. Consequently, this suggests that the reliability of the proposed loss function improves as training advances.

3.1.2. *Learning environment*

In this study, a high-performance computational framework was assembled to facilitate the training and assessment of cutting-edge deep learning architectures. The foundational system infras-

Table 2: Specific learning environment.

Type	Product	Version
OS	Windows 10 Education	19 042.1826
GPU driver	NVIDIA GPU driver	30.0.14.7168
NVIDIA GPU computing toolkit	CUDA	11.2
Programing language	Python	3.8
Deep learning framework	Tensorflow	2.5.0

tructure operated on the Windows 10 Education platform and other information, such as language and driver versions, are listed in Table 2.

3.2. Eye-tracking module

In this study, the Tobii Pro Glass 3 was used for real-time gaze tracking and image acquisition. This wearable eye tracker is adept at integrating into a wide range of settings and unobtrusively captures the viewpoint of the wearer. The device is equipped with a built-in scene camera that provides live visual feedback from the user's perspective (T. H. Li *et al.*, 2020). This feature is instrumental in concurrently gathering both visual field data and images and is an essential component of this study for comprehensive analysis. The device has also been proven to be reliable in various studies (Jonas *et al.* (2021)).

However, using the program of Tobii Pro Glasses 3 was financially prohibitive and lacked the freedom to be used in this algorithm. Therefore, information from Tobii Pro Glasses 3 was utilized in the module, following the structure depicted in Fig. 6. We also needed to integrate the device into the system, which could cause problems with communication. For this purpose, we wirelessly connected the glasses to a PC and imported the data into Python using the Real-Time Streaming Protocol (RTSP). This protocol is pivotal for establishing a stable communication link, which is essential for effectively streaming live visual and gaze data into the proposed Python-based analysis system (Muhammad *et al.*, 2013). The primary role of the RTSP in this configuration was to ensure the seamless transfer of real-time data from the glasses, thereby facilitating the efficient and continuous acquisition of eye-tracking and visual data, which is crucial for the research objectives of this study.

3.3. Integrated module

For real-time analysis of a worker's gaze upon object detection, it is essential to integrate the two modules above to determine which part of the object is being observed by the worker. In other words, the integration module outputs a real-time display by combining the bounding box information procured from the object detection module with the current gaze coordinates provided by the eye-tracking module. However, because the system levels and characteristics differ, it is challenging to implement and integrate them into one environment (Ham *et al.*, 2018; B. S. Kim *et al.*, 2020; Tran *et al.*, 2014). This section discusses troubleshooting during the integration of these two modules.

3.3.1. *Integrated module structure*

The integration process of the eye-tracking module with the object detection module is shown in Fig. 7. Initially, when the target object was observed through Tobii Glasses, two distinct types of data were captured: image and gaze data. The image data are relayed to the object detection module, whereas the gaze data

Figure 6: Networking communication structures.

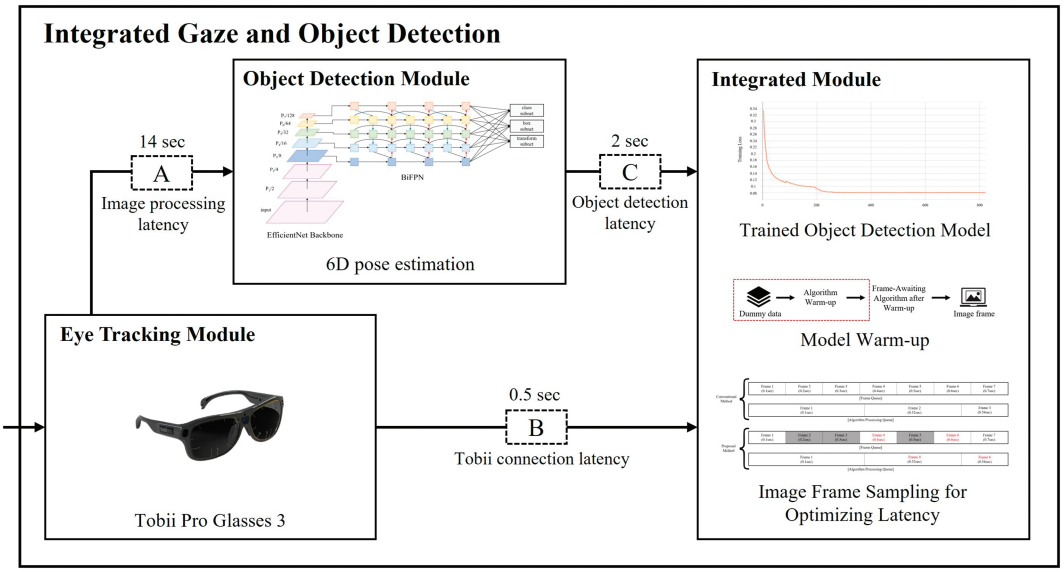


Figure 7: Three primary latency of ISGOD.

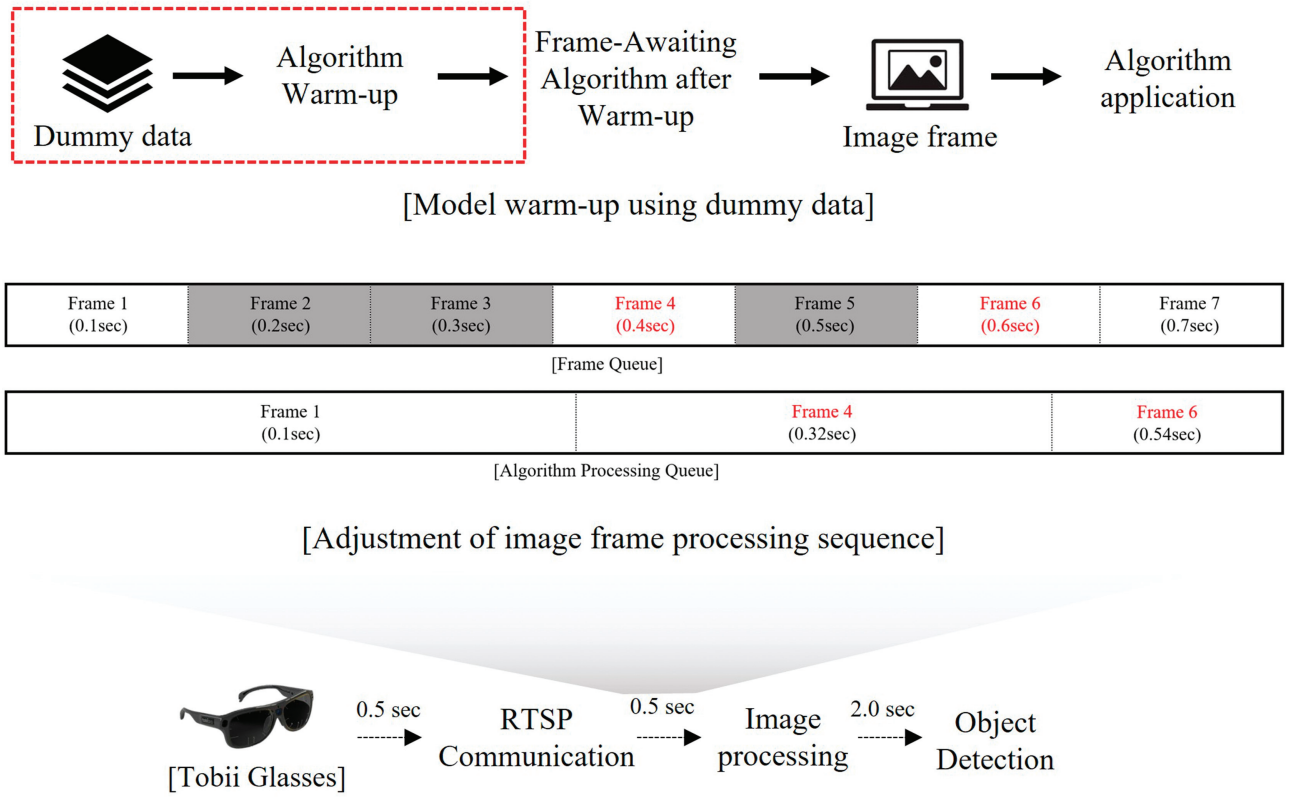


Figure 8: Two primary delaying causes and solutions.

are transmitted to the integration module. Subsequently, the object detection module, which employs a pre-trained algorithm on the image frames, generates an output encompassing the bounding box, rotation, and depth information of the object. Ultimately, the task of merging information from each module is essential, which consequently introduces a delay due to the integration of disparate data sources. Three primary sources of delay were identified: image processing latency (A), Tobii connection latency (B), and object detection latency (C). Latency B originates from the communication between the Tobii Glasses and the PC. In contrast, latency C is a consequence of the computational requirements of the 6D pose estimation algorithm, mainly owing to its capability to handle multi-object and dynamic object detection.

3.3.2. *Latency reduction method*

The delay known as latency A in this study comes from how long it takes for the algorithm to start working on the image. This latency is a result of the time lag between the transmission of the image frame from the gaze-tracking module and processing by the algorithm of the object detection module. To address this issue, as shown in Fig. 8, two primary causes were identified, and the corresponding solutions were implemented. The delay known as latency A in this study comes from how long it takes for the algorithm to start working on the image. This latency is a result of the time lag between the transmission of the image frame from the gaze-tracking module and processing by the algorithm of the object detection module. To address this issue, as shown in Fig. 8, two primary causes were identified, and the corresponding solutions were implemented.

The first issue pertains to the algorithm initialization. The algorithm remains inactive until it receives an image frame from the gaze-tracking module. Consequently, the initial operation of the object detection algorithm includes both preparation time and actual processing time. To alleviate this problem, a strategy of pre-activating the algorithm with dummy data was employed to eliminate preparation delays.

The second issue is time synchronization. Tobii Glass operates at an actual FPS of 25, sending 25 frames per second to the object detection module. As shown in Fig. 9, the frame-processing sequence begins with object detection in the first frame. Subsequent frames, such as Frames 2 and 3, must wait until the processing of Frame 1 is complete. This sequential processing leads to the accumulation of waiting times. To resolve this, a sampling method that selects the frame closest to the completion of the current frame from the waiting frames for algorithm application was adopted. This approach effectively addresses the continuous buildup of latency.

3.4. Analyzer module

As illustrated in the left segment of Fig. 10, the size and position of each recognized image vary according to the location of the operator. The implementation of an analyzer module is necessary to confirm and analyze these discrepancies accurately. This module standardizes the representation of objects across different frames, as shown on the right side of Fig. 10.

To analyze eye gaze data from the eye-tracking module and detect object data through the object detection module, a sequence of the perspective transfer algorithm of the analyzer module was undertaken, as depicted in Fig. 11. The sequence diagram illustrates the operation of the analyzer module. The eye-tracking module continuously captures images until the recording session on the Tobii Glasses is concluded and subsequently transmits the data to the object detection module. Once the recording was concluded, the analyzer module collected the data of the detected ob-

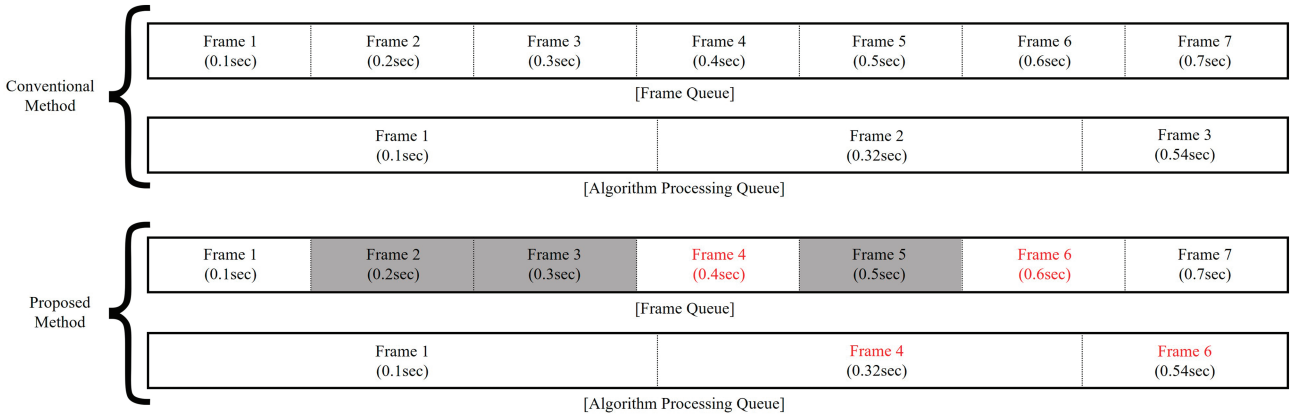


Figure 9: Frame sampling for latency improvement.

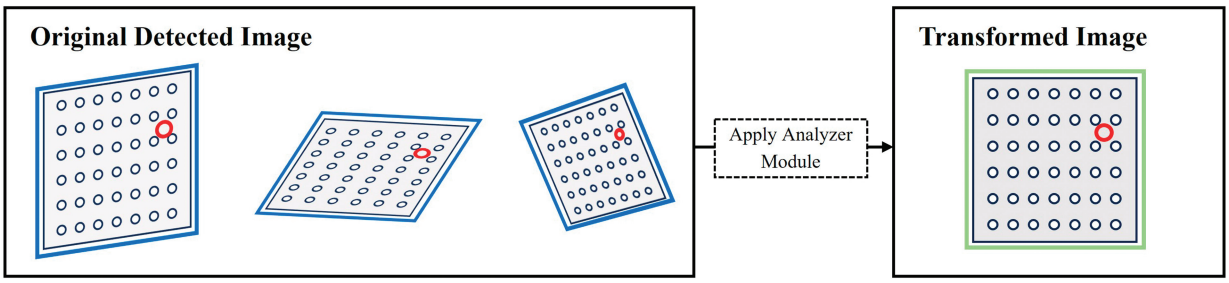


Figure 10: Applying analyzer module.

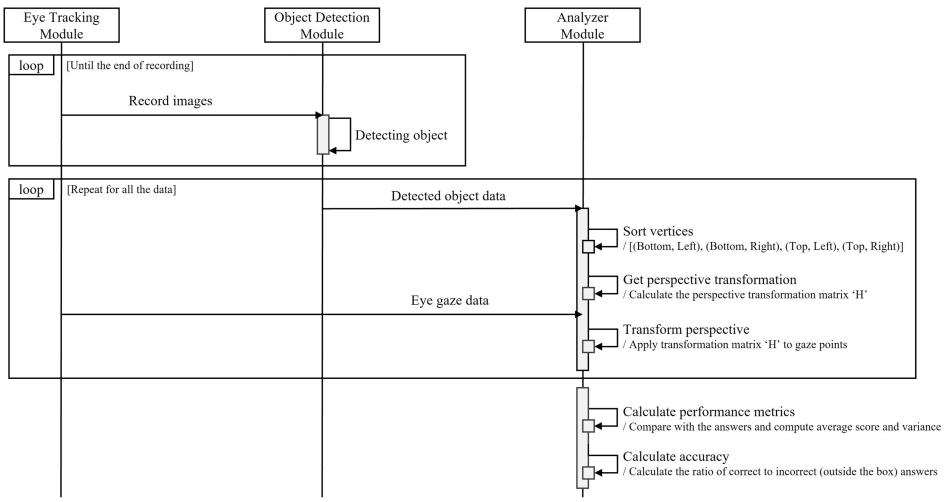


Figure 11: Analyzer module sequence diagram.

jects and arranged the coordinates of the vertices for each square in a specified sequence. This alignment is a prerequisite for performing perspective transformation. Following this, the module computes a transformation matrix, denoted as "H", which is used to convert the coordinates of the detected square to a standardized square form. Concurrently, the eye-tracking module procures the eye gaze data and applies the perspective transformation using matrix "H" on the gaze points. Once all the data were handled, the analyzer module worked out the average scores and how much they varied by looking at where people were supposed to look compared with where they actually looked. It also figured out how accurate the gaze was by checking how many gaze points matched up with the correct spots.

4. Experiments

Figure 12 shows the two principal experiments conducted in this study. Experiment 1 assessed the robustness of the proposed algorithm. This experiment involved scenarios in which both the observer and object were stationary (i), situations in which the observer was stationary while the object alone was in motion (ii), and cases in which both the observer and object were moving (iii). In a dynamic environment, objects can be rotated 360 degrees and flipped upside down or backwards. In contrast, in static environments, they must remain stationary. Humans can also use their arms to move objects in dynamic environments, and their neck and eyes to detect objects. However, in a static environment, only

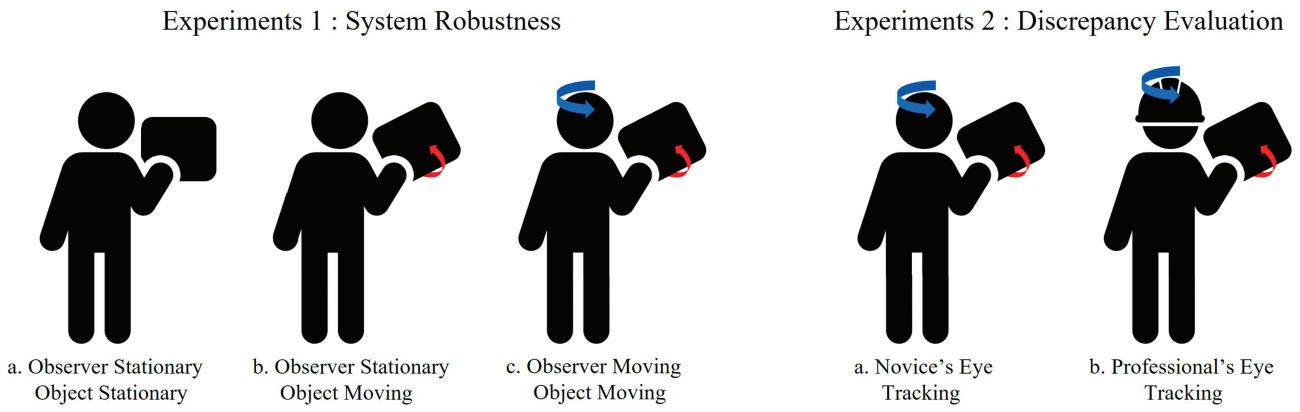


Figure 12: Two experiments scenarios: system robustness and discrepancy evaluation.

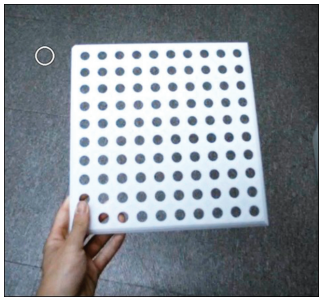


Figure 13: A square plate as a target object.

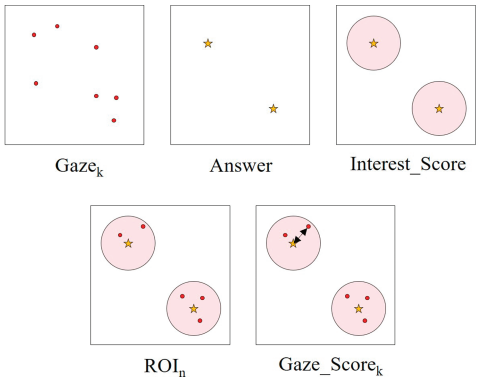


Figure 14: An evaluation matrix for system robustness.

the subject's eyes can be used to detect objects. Evaluation metrics were defined for the experiment, and the average, variance, and accuracy of these scores were compared to demonstrate the robustness of the system in a dynamic environment. Experiment 2 focused on tracking the eye movements of novice and professional workers in conditions in which both the operator and object were in motion. The analysis was conducted by visualizing the gaze positions on an object over time, providing insights into the differing observational strategies of novice and professional operators.

The experimental design was predicted using a square plate, as shown in Fig. 13. The dimensions of the object were configured as 26.5 cm in width, 26.5 cm in length, and 1 cm in height. This object was arbitrarily selected to facilitate the experimental procedure and could be adapted to accommodate different objects.

4.1. Experiment 1: evaluation of system robustness

In this, an experiment is delineated to assess the robustness of the system. First, an evaluation matrix was established, followed by a detailed description of the experimental scenarios. Subsequently, an experiment was conducted, yielding results that were subjected to analysis, culminating in a discussion.

4.1.1. Evaluation matrix

This study initially established an evaluation matrix, as depicted in Fig. 14, to assess the proximity of gaze points to the ROI of the object in addition to the ratio of points within to those outside the ROI.

Assuming the detection of a square object, the gaze coordinates projected onto the object, i.e., post-perspective transformation, are defined as $gaze_k$ in equation (5):

$$Gaze_k = (X_k, Y_k). \tag{5}$$

Furthermore, the set of coordinates representing the critical points on the object to be observed is expressed as *A nswer* to equation (6):

$$Answer = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}. \tag{6}$$

The vicinity of $Answer_n$ extends to *Interest_Score* as per equation (7) and is designated as the ROI_n, denoted as in equation (8):

$$Interest\_Score = C \quad (constant) \tag{7}$$

$$ROI_n = \left\{(x, y) \mid (x - x_n)^2 + (y - y_n)^2 \leq Interest\_Score^2\right\}. \tag{8}$$

Consequently, $Gaze\_Score_k$ as expressed in equation (9), quantifies the closeness of a gaze point to the point of interest, $Answer_n$. This score is designed to measure the precision of gaze alignment, where a score closer to 1 indicates a direct match with the point of interest, signifying high accuracy, while a score closer to 0 suggests a greater distance from the target point, indicating lower accuracy.

$$Gaze\_Score_k = \frac{Interest\_Score - \sqrt{(X_k - x_n)^2 + (Y_k - y_n)^2}}{Interest\_Score}. \tag{9}$$

The average of these $Gaze\_Score_k$ from 1 to K is defined as the *Accuracy_Score* in equation (10). This metric serves as an overall indicator of gaze accuracy across multiple points of interest, offering a comprehensive view of the participant's focus and precision

Table 3: Experiment 1 scenarios.

	Target point	Observer	Object
Scenario 1–1	1 dot	Stationary	Stationary
Scenario 1–2	1 dot	Stationary	Moving
Scenario 1–3	1 dot	Moving	Moving
Scenario 1–4	4 dots	Stationary	Stationary
Scenario 1–5	4 dots	Stationary	Moving
Scenario 1–6	4 dots	Moving	Moving

Table 4: Evaluation matrix of Scenarios 1–1, 1–2, and 1–3.

	Average	Variance	Concentration(%)
Scenario 1–1	0.84	0.01	99.38
Scenario 1–2	0.67	0.05	91.88
Scenario 1–3	0.65	0.07	88.12

in following the intended gaze path.

$$Accuracy\_Score = \frac{1}{K}\sum_{k=1}^{K} Gaze\_Score_k. \tag{10}$$

Finally, the proportion of $Gaze_k$ points falling inside the ROI relative to those outside the ROI is defined as the *Concentration_Index* in equation (11):

$$Concentration\_Index = \frac{N_{inside}}{N_{outside} + N_{inside}},\ N_{inside} \begin{cases} 1, & if\ in\ ROI \\ 0, & otherwise \end{cases}. \tag{11}$$

4.1.2. Scenario

The parameters used in Experiment 1 are listed in Table 3. The experimental parameters were categorized into three distinct classes: target point of the object, mobility of the observer, and mobility of the object. Initially, representing the object on a 50 × 50 coordinate plane, the target points on the object are demarcated. This included a central point at (25, 25), and four peripheral points located in the upper right (40, 40), lower right (40, 10), lower left (10, 10), and upper left (10, 40) quadrants. Following this initial setup, the experiment divides the same observer into various cases to compare differences. Despite the limitation of the number of observer, the detailed insights gained from each participant can still support robust research findings (Sharma *et al.*, 2020).

The experimental setup was categorized based on the state of motion of the observers and objects. This encompasses scenarios where both observers remain stationary and the object is stationary, scenarios where the observer is stationary while the object is moving, and scenarios where there is concurrent movement of both the observer and the object. In addition, there were a total of four subjects performing the scenarios, and the results, excluding heat maps, utilized the average of the eight subjects' evaluation metrics. Consequently, analyses were conducted according to the defined scenario assessment metrics for Scenarios 1–1, 1–2, and 1–3, followed by analogous comparisons for Scenarios 1–4, 1–5, and 1–6.

4.1.3. Experimental results

The experimental findings are illustrated using a heat map in Fig. 15. This visualization reveals that the dispersion of gaze points in Scenario 1–2 is slightly more pronounced than in Scenario 1–1, with Scenario 1–3 displaying the highest level of dispersion. Such a dispersion is an expected consequence of the dynamic nature of the experiment, in which objects and observers are in motion, leading to a broader gaze distribution.

To provide an evaluation matrix, the scores for each scenario, based on the designated evaluation index, are listed in Table 4. Scenario 1–1 registers an average score of 0.84 with a variance of 0.01 and a concentration index rate of 99.38%. In comparison, Scenario 1–2 yields an average score of 0.67, a variance of 0.05, and a concentration index rate of 91.88%. Finally, Scenario 1–3 attains an average score of 0.65, variance of 0.07, and concentration index rate of 88.12%.

The experimental results are depicted using a heat map, as illustrated in Fig. 16. The visual plot indicates a gradual increase in the variance of the gaze points, with Scenario 1–5 showing slightly more variance than Scenario 1–4 and Scenario 1–6 exhibiting the highest variance. This increase in variance, especially compared with Scenarios 1–1, 1–2, and 1–3, is attributed to the requirement for the plate to be moved and the four points to be observed alternately.

The evaluation matrix for each scenario, according to the defined metrics, is presented in Table 5. Scenario 1–4 achieves an average score of 0.59, a variance of 0.08, and a concentration index rate of 82.50%. Scenario 1–5 had a mean score of 0.55, a variance of 0.12, and a concentration index rate of 74.06%. Lastly, Scenario 1–6 records a mean score of 0.46, a variance of 0.13, and a concentration index rate of 63.75%.

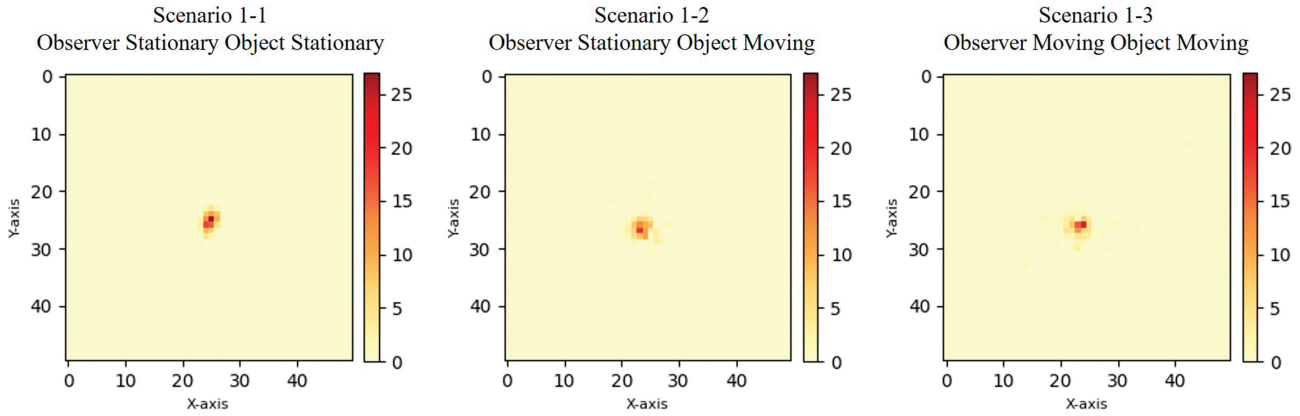


Figure 15: Heat map of Scenarios 1–1, 1–2, and 1–3.

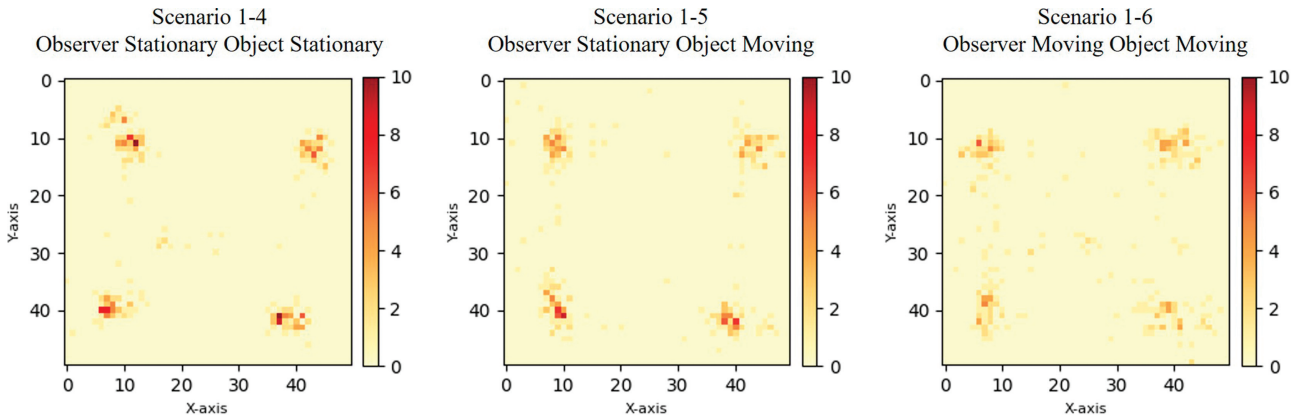


Figure 16: Heat map of Scenarios 1–4, 1–5, and 1–6.

Table 5: Evaluation matrix of Scenarios 1–4, 1–5, and 1–6.

	Average	Variance	Concentration(%)
Scenario 1–4	0.59	0.08	82.50
Scenario 1–5	0.55	0.12	74.06
Scenario 1–6	0.46	0.13	63.75

4.1.4. *Discussion*

Through the execution of experiments across six scenarios, it was ascertained that gaze data can be effectively gathered, even in dynamic contexts where both objects and observers are in motion. Nevertheless, as the scenarios became more dynamic, there was a slight decrease in the average accounting score, an increase in variance, and a reduction in the concentration index rate. This trend is likely due to the difficulty in maintaining a steady gaze in complex environments, especially when either the observer or the object is in motion, compared with a stationary setting. Consequently, the scores were influenced by gaze points that deviated from the correct target. Despite these variations in the evaluation matrix scores, the heat maps provided a precise analysis of which areas were observed, indicating that the dynamic nature of the scenarios did not significantly impede the ability to analyze the focus areas of observation.

4.2. Experiment 2: discrepancy evaluation

The preceding section focused on examining the robustness of the system, effectively demonstrating its ability to integrate gaze information collection with object detection. In this section, we introduce an experimental scenario designed to leverage the capabilities of the proposed system.

4.2.1. *Scenario*

We conducted experiments with eight participants: Novice A, Professional B, Novice C, Professional D, Novice E, Professional F, Novice G, and Professional H, delineating eight specific scenarios that involved either a novice or a professional observer. Subsequently, segmenting the observer group into distinct subsets allowed for a detailed examination of the variations among them (Atkins *et al.*, 2012, Khan *et al.*, 2012). Despite the small sample size, the detailed observations and analyses derived from the proposed eye-tracking system provided essential insights that reinforced the credibility of the research outcomes (Sharma *et al.*, 2020, Tien *et al.*, 2012). Following this, we analyzed and discussed

Table 6: Experiment 2 scenarios.

	Target point	Subject	Observer/Object
Scenario 2–1	4 dots	Novice A	Moving/Moving
Scenario 2–2	4 dots	Professional B	Moving/Moving
Scenario 2–3	8 dots	Novice C	Moving/Moving
Scenario 2–4	8 dots	Professional D	Moving/Moving
Scenario 2–5	4 dots	Novice E	Moving/Moving
Scenario 2–6	4 dots	Professional F	Moving/Moving
Scenario 2–7	8 dots	Novice G	Moving/Moving
Scenario 2–8	8 dots	Professional H	Moving/Moving

the observed gaze patterns across these scenarios, providing insights derived from the data. Through this analysis, we aimed to demonstrate the versatility of the system and focus on showcasing its potential applications in various contexts.

The parameters used in Experiment 2 are listed in Table 6. The experiment involved eight groups: A, B, C, D, E, F, G, and H. The scenarios were bifurcated based on the assumption that the target plate had either four or eight defects. The specific locations of these defects are illustrated in Fig. 17, with Fig. 17a depicting the scenario with four defect points and Fig. 17b showing eight defect points.

Additionally, gaze analysis for both the observer and object was conducted in scenarios where both were moving. To facilitate a straightforward analysis of the scenarios, the target object was segmented into 25 sections arranged from top left to bottom right, as shown in Fig. 18. Finally, before conducting the actual experiment, we performed a baseline experiment where we assigned each subject to look at the sections in a specified order and verified that they looked at the sections according to the scenario. The results showed that four of the subjects looked at each section in the specified order, confirming that the eye-tracking device correctly worked.

4.2.2. *Experimental results*

Figure 19a presents the gaze-point analysis of Novice A, while Fig. 19b showcases that of Professional B. The graphs illustrate the spatial focus of each observer's gaze over time. The x-axis represents time in seconds, and the y-axis denotes the sections of the object being observed. The red circles in the graphs indicate the moments when each observer detected a defect in the area. From Fig. 19a, it is evident that Novice A took approximately 37 s to

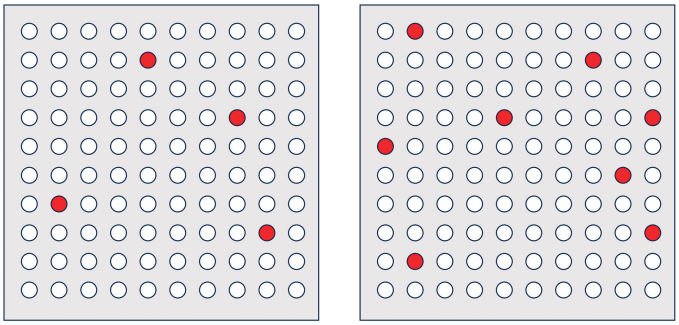


Figure 17: Target points on object.

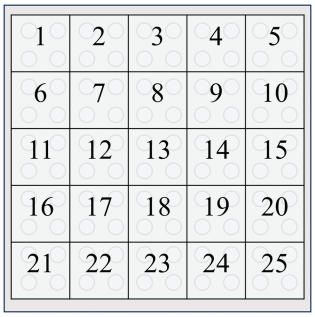


Figure 18: 25 Sections on object.

identify all four defects. Lacking prior knowledge about the defects, Novice A appeared to methodically inspect the entire area sequentially. Notably, Sections 3 and 20 were initially scrutinized for an extended duration; however, defects were not initially identified in these sections, necessitating subsequent reexamination. Conversely, Professional B completed the entire inspection in approximately 21 s, and all defects were identified within approximately 13 s. Rather than following a sequential pattern, Professional B's inspection strategy involved a more sporadic approach, crossing different areas to scan the entire section. This approach highlights a more efficient inspection methodology, possibly stemming from professional experience and a strategic understanding of defect detection.

In Fig. 20, Novice C as depicted in Fig. 20a, exhibits a pattern similar to that of Novice A in Fig. 19a. The total inspection duration for Novice C was 55 s, with the identification of all eight defects occurring at the 51 s mark. This pattern demonstrates that Novice C conducted a comprehensive check of various sections, such as Sections 1, 15, and 19, and revisited certain areas to confirm the presence of defects. Moreover, it is noticeable that Novice C spent a longer time in each area than its professional counterparts. Professional D finished the inspection in 26 s and found all the defects in 24 s, as you can see in Fig. 20b. Professional D quickly checked each section and found defects fast, spending less time looking at each one. However, it was also observed that Professional D had to revisit certain areas, such as part 4, for a second inspection, indicating that even experienced professionals may need to double-check certain sections to ensure thoroughness.

Based on the results of the previous experiments, we wanted to increase the reliability of the results by conducting additional experiments on a total of four people. The results for each novice and expert did not show the same trends as the previous experiments, but in most cases, the experts spent significantly less time to detect defects than the novices and were more efficient at selecting detection zones.

In Fig. 21, which shows the results of a four-defect detection experiment, the novice detected a defect in about 30 s and immediately terminated the experiment, while the expert detected a defect in about 15 s, but after further detection, the expert com-

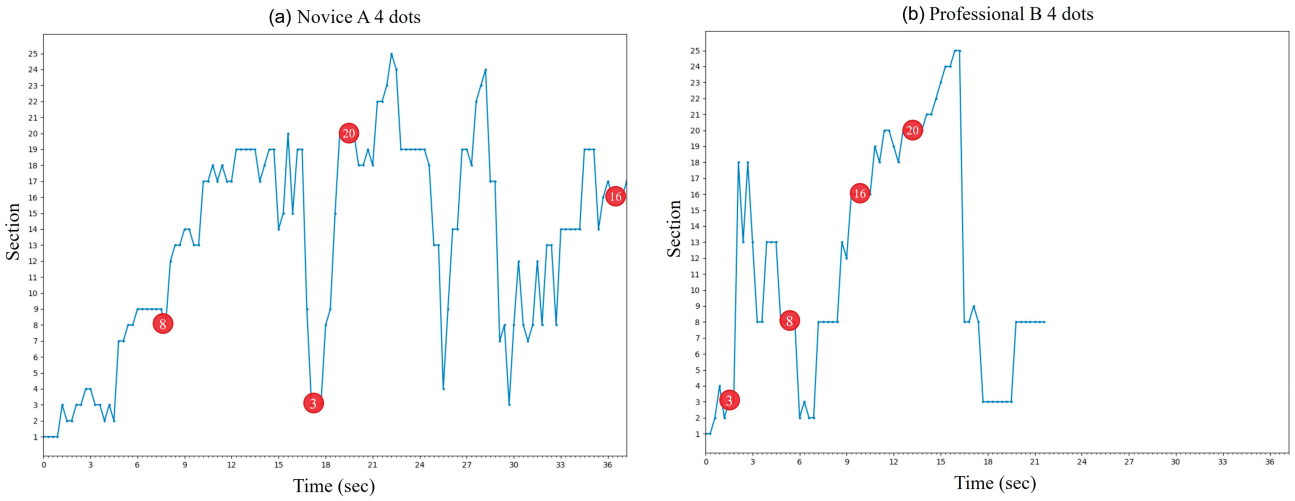


Figure 19: Time-based section graphs for Novice A and Professional B.

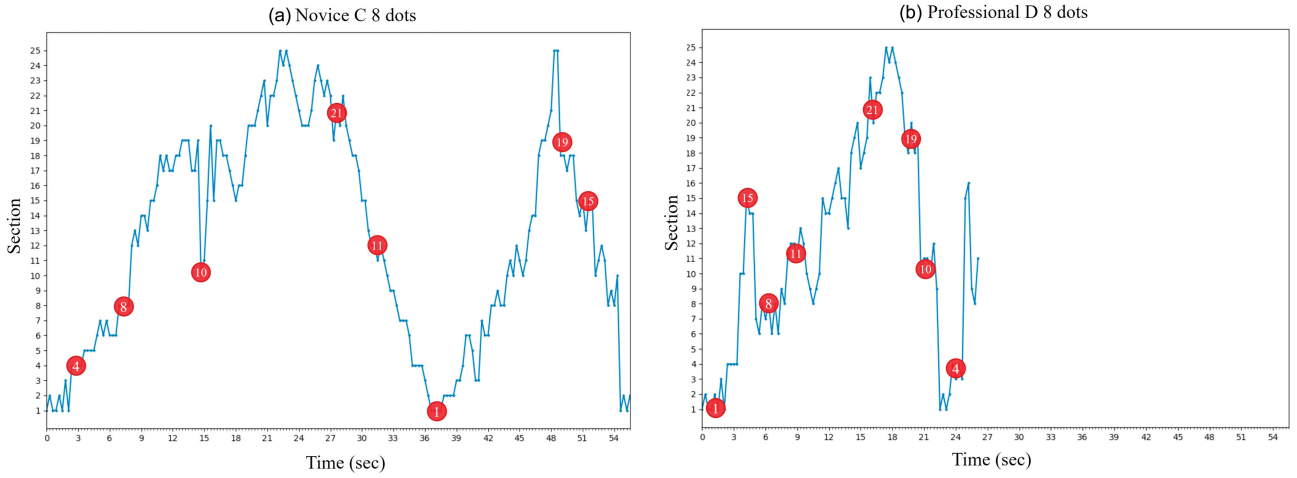


Figure 20: Time-based section graphs for Novice C and Professional D.

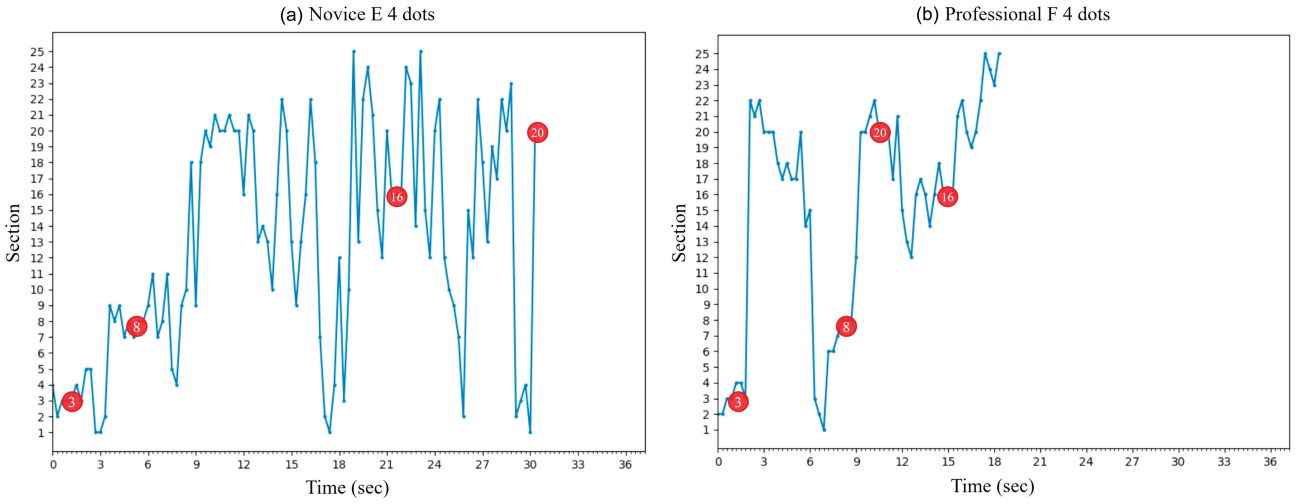


Figure 21: Time-based section graphs for Novice E and Professional F.

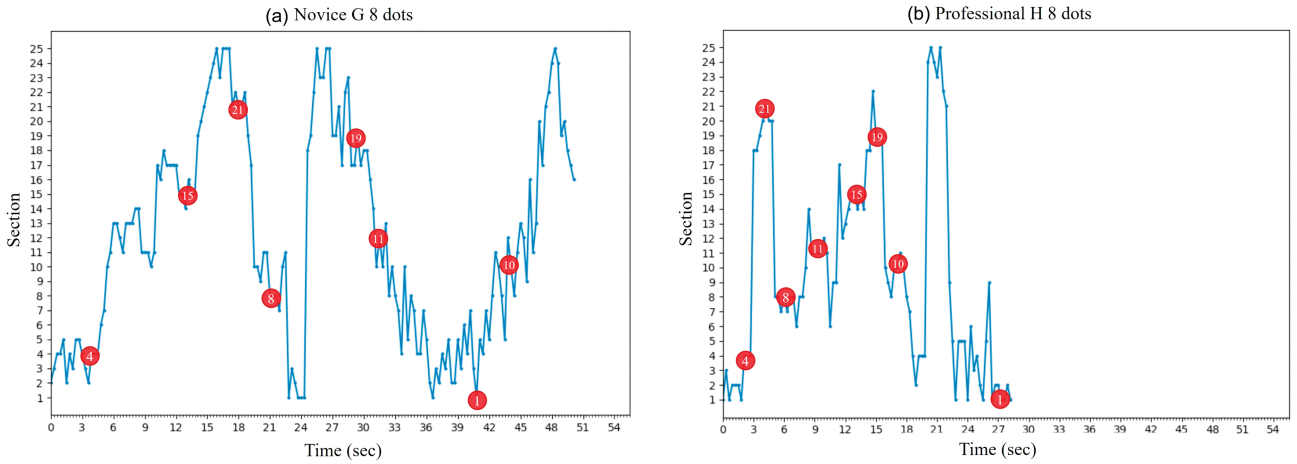


Figure 22: Time-based section graphs for Novice G and Professional H.

pleted the experiment in about 20 s. In Fig. 22, which shows the results of an experiment with eight defect detections, the novice completed the defect detection in about 45 s, but the experiment ended at 51 s after checked for any missed areas. In contrast, experts completed the defect detection and concluded the experiment in approximately 27 s. The differences in results from one experiment to another happened because each person checked for defects in their own way and order. Therefore, it was deter-

mined that additional experiments contributed to the system's reliability.

4.2.3. *Discussion*

The implemented system, as demonstrated through four distinct scenarios in Experiment 2, provides insights into the gaze patterns of expert observers. The system enabled analysis of the entire inspection duration and highlighted the differences in gaze characteristics between novice and professional observers. The novices tended to inspect each section more thoroughly, spending more time on average in each area. By contrast, professionals allocated a shorter duration per section, indicating a more efficient inspection strategy.

The gaze data difference in the presence of know-how is discussed. However, managing human error is difficult, especially in dynamic settings where both the observer and the object are in motion. Such conditions may lead to unintended gaze movements, thus complicating the accuracy of gaze tracking and analysis. This factor necessitates additional considerations in the system to interpret gaze data under fluctuating circumstances accurately. Furthermore, variations in inspection order and methodology can be observed, even among professionals. Each individual may adopt a unique approach to the task, underscoring the necessity of gathering and analyzing gaze data from a broad spectrum of gaze samples. This method helps us fully appreciate the range of inspection techniques and accurately assess how experts use their gaze.

Especially since our experiments were conducted with a limited number of participants, there is a potential issue in adequately capturing the variations across individuals and levels of expertise. Moreover, by utilizing our system to expand the number of various dynamic scenarios, there lies the potential for gaining many insights through different analytical techniques, highlighting the possibility for more extensive analysis and understanding of gaze behavior across different conditions and expertise levels.

5. Conclusions

This study presents a system designed for real-time object detection and eye-tracking in dynamic environments, with the objective of analyzing worker gaze. The system integrates two specialized modules: one for object detection and another for eye-tracking. These modules work concurrently and allow the simultaneous performance of both functions. The system's effectiveness in variable settings has been confirmed through tests assessing its stability and by comparing how experts and beginners differ in their eye-movement patterns.

In the proposed integrated system, object detection was conducted using a 6D pose estimation algorithm, and gaze tracking was performed using Tobii Glasses. The 6D pose estimation algorithm is adept at detecting the 6D aspects of an object, such as position, rotation, and angle. The structured loss function within this algorithm ensures optimal parameter selection. Tobii Glasses are employed for real-time gaze tracking and processing of both image and gaze data. Subsequently, to integrate these two modules, the delay that occurs during integration was solved. The main problems were algorithm initialization and time synchronization, and the delay time was reduced using dummy data and image frame sampling.

A robustness experiment of the integrated system for reliability was conducted for six scenarios by judging the dynamic presence/absence of workers and objects when there were one and four defect points. Although the accuracy and variance in complex environments were generally lower than those in static settings, the differences were not substantial, indicating the capability of the system to collect and analyze gaze data effectively. In addition, a second discrepancy evaluation experiment confirmed the difference in work proficiency between experts and novices. The differences in gaze data according to the presence of expertise were analyzed. These experiments should be followed up with additional experiments using other wearable devices other than Tobi glasses to derive generalized performance of the system. However, the purpose of this study is to demonstrate feasibility rather than performance of the system, which will be addressed in future research.

The ability of the system to analyze worker gaze information in a dynamic environment was demonstrated. This research lays the groundwork for future systems aimed at facilitating the transfer of expertise among workers. However, it did not delve into detailed gaze analyses, such as how expert gaze patterns change during accurate defect detection or which areas novices should specifically avoid. While this study did not provide a direct methodology for the transfer of expert gaze information to novices, it paves the way for future endeavors. Upcoming research will aim to enable novices to attain expert-level proficiency by employing technologies such as augmented reality or voice guidance for an immersive learning experience through in-depth analysis of gaze patterns. Moreover, there is a plan to develop algorithms specifically designed for analyzing expert gaze patterns, enhancing the understanding of efficient gaze analyzing strategies. Additionally, if the object detection algorithm is trained on a diverse array of objects, it will enable the tracking of gaze points across various items in more manufacturing environments.

Acknowledgments

This work was supported by the Industrial Technology Innovation Program (No. 20023014, Development of an Agricultural Robot Platform Capable of Continuously Harvesting more than 3 Fruits per minute and Controlling Multiple Transport Robots in an Outdoor Orchard Environment) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Conflict of interest statement

None declared.

References

Ahrens, M., & Nagel, L. (2023). All eyes on traceability: An interview study on industry practices and eye tracking potential. In *Proceedings of the 2023 IEEE 31st International Requirements Engineering Conference (RE)* (pp. 77–88). IEEE. https://doi.org/10.1109/RE57278.2023.00017.

Akhlaq, M., & Sheltami, T. R. (2013). RTSP: An accurate and energy-efficient protocol for clock synchronization in WSNs. *IEEE Transactions on Instrumentation and Measurement*, **62**, 578–589. https://doi.org/10.1109/TIM.2012.2232472.

Atkins, M., Tien, G., Khan, R., Meneghetti, A., & Zheng, B. (2012). What do surgeons see: Capturing and synchronizing eye gaze for surgery applications. *Surgical Innovation*, **20**, 3. https://doi.org/10.1177/1553350612449075.

Aust, J., Mitrovic, A., & Pons., D. (2021). Assessment of the effect of cleanliness on the visual inspection of aircraft engine blades: An eye tracking study. *Sensors*, **21**, 6135. https://doi.org/10.3390/s21186135.

Borgianni, Y., Rauch, E., Maccioni, L., & Mark, B. G. (2018). User experience analysis in Industry 4.0–The use of biometric devices in engineering design and manufacturing. In *Proceedings of the 2018 IEEE International Conference on Industrial EngiSEering and Engineering Management (IEEM)* (pp. 192–196). IEEE. https://doi.org/10.1109/IEEM.2018.8607367.

Bukschat, Y., & Vetter, M. (2020). EfficientPose: An efficient, accurate and scalable end-to-end 6D multi-object pose estimation approach. https://doi.org/10.48550/arXiv.2011.04307.

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, **42**, 692–700. https://doi.org/10.3758/BRM.42.3.692.

Ghanbari, L., Wang, C., & Jeon, H. W. (2021). Industrial energy assessment training effectiveness evaluation: An eye-tracking study. *Sensors*, **21**, 1584. https://doi.org/10.3390/s21051584.

Ham, S. H., Roh, M. I., & Zhao, L. (2018). Integrated method of analysis, visualization, and hardware for ship motion simulation. *Journal of Computational Design and Engineering*, **5**, 285–298. https://doi.org/10.1016/j.jcde.2017.12.005.

He, Y., Sun, W., Huang, H., Liu, J., Fan, H., & Sun, J. (2020). PVN3D: A deep point-wise 3D keypoints voting network for 6DoF pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11632–11641). IEEE. https://doi.org/10.48550/arXiv.1911.04231.

Hinterstoisser, S., Holzer, S., Cagniart, C., Ilic, S., Konolige, K., Navab, N., & Lepetit, V. (2011). Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In *Proceedings of the 2011 International Conference on Computer Vision (ICCV)* (pp. 858–865). IEEE. https://doi.org/10.1109/ICCV.2011.6126326.

Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, **108**, 67–76. https://doi.org/10.1016/j.visres.2015.01.013.

Kang, B. G., Park, H. M., Jang, M., & Seo, K. M. (2021). Hybrid model-based simulation analysis on the effects of social distancing policy of the COVID-19 epidemic. *International Journal of Environmental Research and Public Health*, **18**, 11264. https://doi.org/10.3390/ijerph182111264.

Khan, R. S., Tien, G., Atkins, M. S., Zheng, B., Panton, O. N., & Meneghetti, A. T. (2012). Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation?. *Surgical Endoscopy*, **26**, 3536–3540. https://doi.org/10.1007/s00464-012-2400-7.

Kim, B. S., Nam, S., Jin, Y., & Seo, K.-M. (2020). Simulation framework for cyber-physical production system: Applying concept of LVC interoperation. *Complexity*, **2020**, 4321873. https://doi.org/10.1155/2020/4321873.

Kim, J. Y., Pyo, H. R., Jang, I. H., Kang, J. H., Ju, B. K., & Ko, K. E. (2022). Tomato harvesting robotic system based on Deep-ToMaToS: Deep learning network using transformation loss for 6D pose estimation of maturity classified tomatoes with side-stem. *Computers and Electronics in Agriculture*, **201**, 107300. https://doi.org/10.1016/j.compag.2022.107300.

Kuma, P., Mittal, A., & Kumar, P. (2010). Addressing uncertainty in multi-modal fusion for improved object detection in dynamic environment. *Information Fusion*, **11**, 311–324. https://doi.org/10.1016/j.inffus.2009.10.002.

Lee, W.-J., Roh, M.-I., Lee, H.-W., Ha, J., Cho, Y.-M., Lee, S.-J., & Son, N.-S. (2021). Detection and tracking for the awareness of surroundings of a ship based on deep learning. *Journal of Computational Design and Engineering*, **8**, 1407–1430. https://doi.org/10.1093/jcde/qwab053.

Li, T. H., Suzuki, H., & Ohtake, Y. (2020). Visualization of user's attention on objects in 3D environment using only eye tracking glasses. *Journal of Computational Design and Engineering*, **7**, 228–237. https://doi.org/10.1093/jcde/qwaa019.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827.

Lušić, M., Fischer, C., Braz, K. S., Alam, M., Hornfeck, R., & Franke, J. (2016). Static versus dynamic provision of worker information in manual assembly: A comparative study using eye tracking to investigate the impact on productivity and added value based on industrial case examples. *Procedia CIRP*, **57**, 504–509. https://doi.org/10.1016/j.procir.2016.11.087.

Mark, B. G., Rauch, E., & Matt, D. T. (2021). Worker assistance systems in manufacturing: A review of the state of the art and future directions. *Journal of Manufacturing Systems*, **59**, 228–250. https://doi.org/10.1016/j.jmsy.2021.02.017.

Nakamura, J., & Nagayoshi, S. (2019). The pottery skills and tacit knowledge of a master: An analysis using eye tracking data. *Procedia Computer Science*, **159**, 1680–1687. https://doi.org/10.1016/j.procs.2019.09.338.

Niemann, J., Fussenecker, C., & Schlösser, M. (2019). Eye tracking for quality control in automotive manufacturing. In Walker A., O'Connor R., & Messnarz R. (Eds.), *Proceedings of the Systems, Software and Services Process Improvement: 26th European Conference, EuroSPI 2019* (Vol. **26**, pp. 289–298). Springer. https://doi.org/10.1007/978-3-030-28005-5_22.

Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., & Witlox, F. (2012). Interpreting maps through the eyes of expert and novice users. *International Journal of Geographical Information Science*, **26**, 1773–1788. https://doi.org/10.1080/13658816.2011.642801.

Ramachandra, C. K., & Joseph, A. (2021). IEyeGASE: An intelligent eye gaze-based assessment system for deeper insights into learner performance. *Sensors*, **21**, 6783. https://doi.org/10.3390/s21206783.

Ren, Z., Fang, F., Hou, G., Li, Z., & Niu, R. (2023). Appearance-based gaze estimation with feature fusion of multi-level information elements. *Journal of Computational Design and Engineering*, **10**, 1080–1109. https://doi.org/10.1093/jcde/qwad038.

Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 141–149). Association for Computing Machinery. https://doi.org/10.1145/1054972.1054993.

Sampaio, I. G. B., Machaca, L., Viterbo, J., & Guérin, J. (2021). A novel method for object detection using deep learning and CAD models. https://doi.org/10.48550/arXiv.2102.06729.

Sharma, K., Giannakos, M., & Dillenbourg, P. (2020). Eye-tracking and artificial intelligence to enhance motivation and learning. *Smart Learning Environments*, **7**, 1–19. https://doi.org/10.1186/s40561-020-00122-x.

Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A., & Blake, A. (2013). Efficient human pose estimation from single depth images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 2821–2840. https://doi.org/10.1109/TPAMI.2012.241.

Son, M., & Ko, K. (2022). Learning-based essential matrix estimation for visual localization. *Journal of Computational Design and Engineering*, **9**, 1097–1106. https://doi.org/10.1093/jcde/qwac046.

Takahashi, R., Suzuki, H., Chew, J. Y., Ohtake, Y., Nagai, Y., & Ohtomi, K. (2018). A system for three-dimensional gaze fixation analysis using eye tracking glasses. *Journal of Computational Design and Engineering*, **5**, 449–457. https://doi.org/10.1016/j.jcde.2017.12.007.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks.

Tan, M., Pang, R., & Le, Q. V. (2020). EfficientDet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10781–10790). IEEE. https://doi.org/10.48550/arXiv.1911.09070.

Tien, G., Atkins, M. S., & Zheng, B. (2012). Measuring gaze overlap on videos between multiple observers. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 309–312). Association for Computing Machinery. https://doi.org/10.1145/2168556.2168623.

Tran, T. A., & Park, J. Y. (2014). Development of integrated design methodology for various types of product—service systems. *Journal of Computational Design and Engineering*, **1**, 37–47. https://doi.org/10.7315/JCDE.2014.004.

Ulutas, B. H., FÖzkan, N., & Michalski, R. (2020). Application of hidden Markov models to eye tracking data analysis of visual quality inspection operations. *Central European Journal of Operations Research*, **28**, 761–777. https://doi.org/10.1007/s10100-019-00628-x.

Wang, F. S., Gianduzzo, C., Meboldt, M., & Lohmeyer, Q. (2022). An algorithmic approach to determine expertise development using object-related gaze pattern sequences. *Behavior Research Methods*, **54**, 493–507. https://doi.org/10.3758/s13428-021-01652-z.

Ye, L., Yang, S., Zhou, X., & Lin, Y. (2023). Supporting traditional handicrafts teaching through eye movement technology. *International Journal of Technology and Design Education*, **33**, 981–1005. https://doi.org/10.1007/s10798-022-09748-z.

Yin, P., Ye, J., Lin, G., & Wu, Q. (2021). Graph neural network for 6D object pose estimation. *Knowledge-Based Systems*, **218**, 106839. https://doi.org/10.1016/j.knosys.2021.106839.

Zheng, T., Glock, C. H., & Grosse, E. H. (2022). Opportunities for using eye tracking technology in manufacturing and logistics: Systematic literature review and research agenda. *Computers & Industrial Engineering*, **171**, 108444. https://doi.org/10.1016/j.cie.2022.108444.

Received: January 29, 2024. **Revised:** April 29, 2024. **Accepted:** April 29, 2024