Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: www.tandfonline.com/journals/tbit20

Gaze-aware support for security surveillance: a user-centered field study

Alexandre Marois , Daniel Lafond & Sébastien Tremblay

To cite this article: Alexandre Marois , Daniel Lafond & Sébastien Tremblay (07 Apr 2026): Gaze-aware support for security surveillance: a user-centered field study, Behaviour & Information Technology, DOI: 10.1080/0144929X.2026.2651124

To link to this article: https://doi.org/10.1080/0144929X.2026.2651124

Published online: 07 Apr 2026.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2026.2651124

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Gaze-aware support for security surveillance: a user-centered field study

Alexandre Marois[a,b], Daniel Lafond[c] and Sébastien Tremblay[a]

[a]École de psychologie, Université Laval, Québec, Canada; [b]School of Psychology and Humanities, University of Lancashire, Preston, UK; [c]Thales, CortAIx Labs, Québec, Canada

ABSTRACT
Security surveillance is characterised by substantial cognitive challenges to operators. Scantracker is a mixed-reality gaze-aware support tool that alerts surveillance operators to neglected cameras, attentional tunnelling, and vigilance decrements. Initial research efforts were conducted in simulated environments to examine the effects of Scantracker on surveillance performance; however, this tool has yet to be deployed and tested in a real-world operational environment. In the current study, we tested Scantracker in an airport operations centre to assess the feasibility of its integration and to collect expert feedback regarding its operational relevance. Operators used Scantracker voluntarily during their work shift, while gaze data enabled system notifications. They provided ratings on perceived utility, workload and ergonomic quality along with qualitative feedback on their experience. The pattern of results highlights the potential of Scantracker to support surveillance operators and demonstrates the value of user-centred field testing for developing intelligent monitoring assistants.

ARTICLE HISTORY
Received 27 June 2025
Accepted 8 March 2026

KEYWORDS
Security surveillance; eye tracking; mixed reality; user-centered research; field study

1. Introduction

Security surveillance is essential for protecting infrastructure and public places. Recent advances in surveillance equipment (e.g. computer vision and pattern recognition; Ahmed and Echi 2021) have led to the development of semi- and fully-automated surveillance systems. Despite the high performance these systems can achieve (e.g. Khan et al. 2024), surveillance still heavily relies on – and will likely continue to include – human operators. This is due to their critical role in overseeing, evaluating, and prioritising the various actions performed by surveillance technologies (Dadashi, Stedmon, and Pridmore 2013; Nicosia and Kristensson 2024), the need for clear accountability when automated systems make security-critical decisions (Santoni de Sio and Van den Hoven 2018), and the high value of human-automation joint operations (Suss et al. 2015).

Nonetheless, human operators face many cognitive challenges, including information overload, multitasking, background sound distraction, task interruptions, and fatigue. These factors can hinder the timely detection of incidents or threats (Hodgetts et al. 2017; Suss et al. 2015). For example, interruptions from colleagues or distractions caused by conversations or (low-priority) alerts may prevent optimal processing of the scene to monitor. Moreover, events requiring active intervention from operators are typically rare. Mucchielli (2016) reported, for instance, that a surveillance operator in a local police station might encounter one critical incident every four days on average. The abundance of information to process further exacerbates the difficulty of detecting these rare critical events. In addition to these challenges, reports of reduced motivation and situational uncertainty are also common (for a review, see Hodgetts et al. 2017).

Given the cognitive challenges faced by surveillance operators, strategies must be developed to support their work and reduce errors. The current paper adopts a user-centered approach to investigate the effects and perceptions of a mixed-reality gaze-based tool designed to assist surveillance operators. We first present the state-of-the-art on the main strategies used to support security surveillance and describe how the user-centered approach and methods can be leveraged to inform the development of support systems. Second, we present a mixed-method experiment conducted in a real operational environment involving actual surveillance operators. Finally, the discussion and conclusion sections

CONTACT Alexandre Marois alexandre.marois@psy.ulaval.ca École de psychologie, Université Laval, 2325, allée des Bibliothèques, Québec (QC), G1V 0A6, Canada

integrate the findings and outline implications for security surveillance operation centres, as well as directions for future research.

1.1. State-of-the-art

1.1.1. Methods for improving surveillance operations

Research has proposed several avenues to support surveillance operators, including, for instance, recruiting personnel with appropriate skill sets (Aitken, Champion, and Stainer 2019; Blechko, Darker, and Gale 2008). Although different approaches can be used to identify operators who possess surveillance-relevant abilities (e.g. cognitive tasks, personality assessments and psychometric tests; see, e.g. Edwards et al. 2018; Marois et al. 2021a), it can be difficult for organisations to implement these practices, especially given the high rate of attrition and turnover observed in surveillance centres (Gill et al. 2005; Piza and Moton 2023). Cognitive systems engineering and user experience (UX) design approaches can also be useful for analyzing operators' cognitive processes and then redesigning surveillance environments and camera visualisations accordingly (Pelletier et al. 2015). This ensures that the properties of the surveillance system are in line with how camera feeds are monitored and processed at the cognitive level.

Decision-support systems represent another promising avenue. These systems can support decision making by using, for instance, automated alerts or recommendations for further action (e.g. Parasuraman and Manzey 2010), harnessing the potential of human-autonomy teaming where humans can work and even cooperate with autonomous systems (see McNeese et al. 2017). Some of these systems have been developed for domains where visual monitoring represents the primary task, similar to security surveillance, mostly relying on gaze data analysis to index visual attention. For example, Rupert et al. (2016) developed a solution that triggered real-time warnings when loss of situation awareness was detected among aircraft pilots by using spatial disorientation and gaze measures. Similarly, Lounis, Peysakhovich, and Causse (2020) developed a tool that analyzes real-time visual monitoring behaviour in the cockpit, compares it to nominal visual circuits, and triggers an alarm in case of visual scan deviation. Studies on car drivers have also reported different algorithms capable of identifying driver distraction using oculometric data (for reviews, see Dong et al. 2011; Ghandour et al. 2020).

Closer to security surveillance applications, Taylor et al. (2015a) developed the EyeFrame, a gaze-aware tool to support the monitoring of semi-autonomous agents performing search and rescue tasks. Their system collects real-time gaze information from users controlling multiple agents (between 4 and 10) that were displayed on separate map-panels on a single screen. Based on gaze history, the system highlighted the most neglected map-panels with red frame cues, serving as a short-term memory aid. Performance on the rescue task was significantly improved with EyeFrame. Gaze-contingent notifications led to a greater number of targets rescued compared to two control conditions, especially in trials with a larger number of agents to monitor (>6). Taylor and his colleagues also reported faster reaction times and lower cognitive load as indexed by pupil size measures (see also Taylor et al. 2015b).

A tool developed specifically to support security surveillance was tested by Tremblay et al. (2018) within an urban security surveillance simulation (see Suss et al. 2015). The prototype system, called Scantracker, collects real-time gaze data from participants performing security surveillance tasks. It analyzes gaze data and triggers visual alerts to improve incident detection during surveillance. Scantracker produces three kinds of notifications based on specific eye movement criteria (Marois et al. 2020): (a) one focused on negligence, highlighting ignored cameras; (b) one aimed at reducing attention tunnelling, alerting participants when they over-focus on certain cameras; and (c) one that sends a notification when a decrement in vigilance is detected based on gaze velocity and fixation measures.

Tremblay et al. (2018) investigated the impact of Scantracker on an urban security surveillance simulation comprising eight camera feeds, only six of which could be displayed on monitors. The simulation scenario included realistic preprogrammed incidents to detect (e.g. a fire, a fight and other public disturbance events). Tremblay et al. reported no improvement in incident detection, although a significant decrease in visual search time (scan patterns) was found for participants who received feedback from Scantracker, compared to a control group without any aid. Similar benefits for visual monitoring behaviour and gaze-based measures of efficient surveillance were also reported in Marois et al. (2020), where Scantracker was integrated within a mobile eye tracking system, and in Williot et al. (2024) using a mixed reality (MR) device equipped with eye-tracking capabilities.

The MR implementation (cf. Marois et al. 2022) enables Scantracker to be universally deployed in practically any surveillance room, regardless of its infrastructure and user interfaces. MR supports the integration of a virtual environment (in Scantracker's case, the notifications) into a physical environment (in this case, the

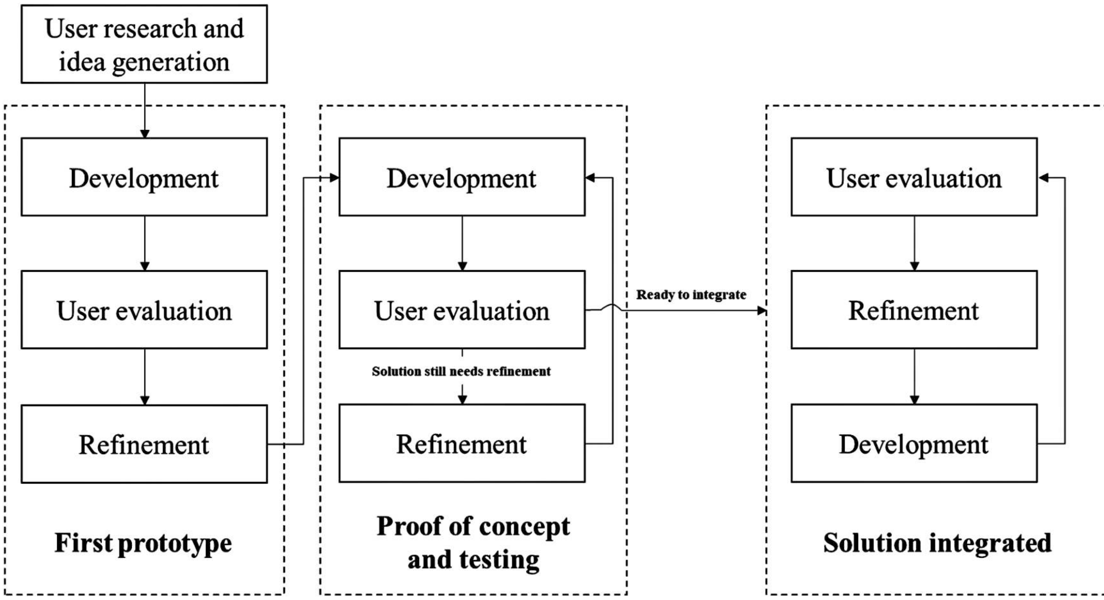

Figure 1. User-centered design iterative process.

camera feeds to monitor; Rokhsaritalemi, Sadeghi-Niaraki, and Choi 2020). Indeed, gaze-contingent notifications can be displayed directly in MR, within the user's field of view. The addition of virtual cues in other domains related to visual monitoring was previously shown to be useful for operators, for instance, to support air traffic control (Bagassi et al. 2020; Barbotin et al. 2022). Yet, while the studies discussed above outline how gaze-aware decision aids may impact monitoring behaviour and workload (cf. Marois et al. 2020; Tremblay et al. 2018; Williot et al. 2024), user-centered evaluation in real-life operational settings is critical before Scantracker can be deployed in the field.

MR and augmented reality (AR) technologies have great potential for supporting operators across different use cases. For example, Pillajo et al. (2025) showed that situation awareness could be improved via the integration of both MR and AR interfaces designed to support field managers in overseeing indoor work in high-rise construction projects. Similarly, Wu et al. (2022) developed a real-time mixed reality-based visual warning system to help construction workers identify accident risks and proactively avoid hazards on construction sites. Other studies have also outlined the potential of AR and MR systems to reduce cognitive workload and, in some cases, reduce decision time in a wide range of operational settings (Blattgerste et al. 2018; Funk et al. 2016; Tang et al. 2003). For these systems to be properly integrated and adapted to operational constraints, careful user experience evaluation and testing must be achieved (Bock, Bohné, and Tadeja 2025; Loizeau et al. 2021).

1.1.2. User-centered approach to system development and field evaluation

User-centered design (UCD) is an approach tailored to identify user needs and inform design decisions that is particularly useful to iteratively develop and ultimately implement support systems in real-life settings (e.g. Anderson et al. 2021). Systems engineers tend to design solutions solely based on technical requirements, which sometimes fail to consider how human operators might interact with them (Steane et al. 2023). UCD, however, provides a set of methods for identifying and designing systems for users' needs while considering the particular context of use (Norman and Draper 1986; Still and Crane 2017). UCD typically follows an iterative process where user evaluation, development, and refinement activities are sequentially carried out across the steps of prototyping, proof of concept and testing, and solution integration within the operational environment (see Figure 1).

Field studies are particularly relevant for the user evaluation step. This type of study takes place directly in the operational environment, where future end-users try to integrate and interact with the potential solution to support their usual tasks (Nielsen Norman Group 2016). Field studies are considered among the most important UCD methods (Eshet and Bouwman 2015; Hussain, Slany, and Holzinger 2009; Mao et al. 2009). They provide opportunities for gaining meaningful context and information on the operational environment and characteristics of the interaction with newly developed and implemented tools (Haines et al. 2022). As per Figure 1, evaluation for the 'Proof-of-concept and testing' step is required in a real-life environment. This represents a necessary step for further increasing the technological maturity of the tool, but also to assess the perception of surveillance operators towards the relevance of the Scantracker prototype.

1.2. Study objectives

The main goal of the current study was to evaluate the impacts of Scantracker in an operational environment among surveillance operators. Through this goal, we aimed at answering the following research question: *How do surveillance operators perceive the Scantracker tool, and what effects can this tool have on workload and surveillance behaviour?* Additionally, the study served a subgoal, which is to describe the UCD process used to collect end-user's insights, and the method to transition from laboratory testing to an operational environment for field tests. To reach these goals, we performed a field study in a real surveillance operations centre, namely the Operations Control Centre of the Quebec City Airport (YQB) in Canada. We collected feedback from surveillance operators either interacting with the tool or performing their typical surveillance tasks, without any support from Scantracker. Security considerations prevented us from intervening while operators performed actual surveillance activities. Unlike a laboratory experiment, the goal was not to provide an exhaustive portrait of the effects of Scantracker on surveillance performance. Therefore, this field study took the form of a quasi-experimental study without randomisation, or experimental control. We focused on collecting end-users workload measures, as well as ergonomic and usability feedback following their interaction with the tool.

2. Materials and methods

2.1. Participants

Ten volunteer surveillance operators (six women, four men) from the YQB team of security personnel with a mean age of 29.1 years ($SD = 6.77$) and an average of 2.29 years of experience ($SD = 0.96$) took part in the study. Their main task is to monitor surveillance cameras while responding to field calls, filling up security reports and coordinating with field security personnel to manage passengers' arrival. There were six participants assigned to the control condition and four to the Scantracker condition. Due to operational constraints from the operation centre, both conditions occurred at different moments and participants (operators) were assigned to the conditions depending on their work shifts. They reported having no experience with MR and other similar technologies. This study was approved by the Université Laval Research Ethics Committee (approval No. 2018-280 A-2 / 17-12-2019).

2.2. Apparatus and material

2.2.1. Field study testing environment

The study was conducted at the Operations Control Centre of YQB (see Figure 2). Participants were asked to continue with their usual tasks and could freely decide to participate in a data collection period during their work shift. Scantracker provided notifications to support screen negligence, attention tunnelling and vigilance decrement. Operators in the Scantracker condition could, however, freely deselect the notifications of their choice if they considered them irrelevant or disruptive, using the MR notification control panel (see Figure 3). Before using the tool, a vigilance baseline

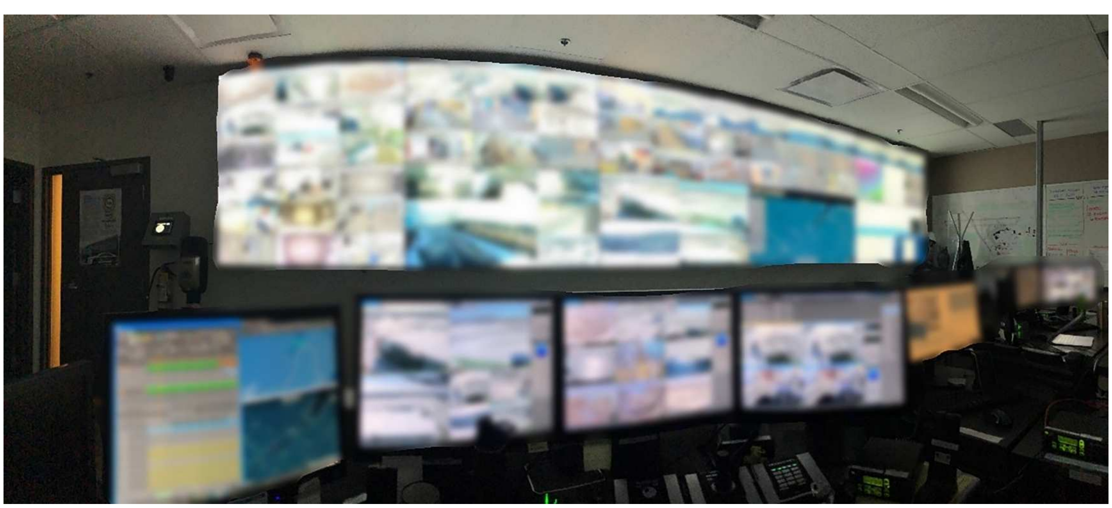

Figure 2. View from an operator workstation at the YQB airport operations centre. Content of the monitor is blurred for security considerations.

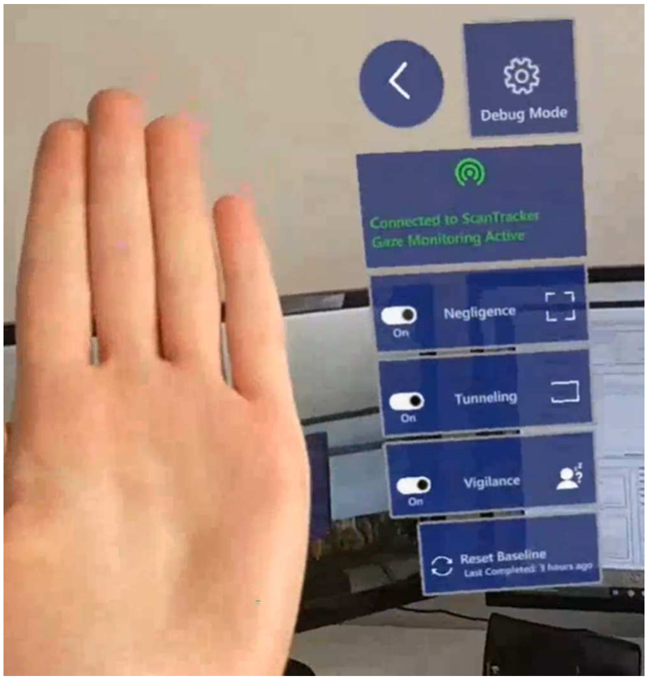

Figure 3. Depiction of the notification control panel appearing on the right of the hand of the user.

was collected at the beginning of each data collection period. Participants were asked, if possible, to keep the Scantracker MR headset on for at least 1 h during data collection.

The environment of the operators (workstation and the video wall) was modelled by the research team prior to the study. This included the subsections of the video wall (i.e. the 60 cameras), which enables Scantracker to present notifications at the camera level. Camera labels were also included to allow negligence notifications outside the field of view to provide information about the specific camera that was neglected. The four monitors located on the workstation were also included in the model and were associated with a parallel task feature. This feature meant that when an operator was looking at the monitors rather than at the video wall, they were considered to be performing a different task than monitoring the cameras. In this situation, a grace period of 20 s for the vigilance model was added to prevent the vigilance model from triggering a notification (e.g. should the gaze velocity be low because the operator was reading an email). Figure 4 displays the surveillance environment of the operators, including the video wall and the operator workstation.

2.2.2. Surveillance solution

The version of Scantracker reported in the present study was developed to be integrated within the Microsoft HoloLens 2 system, which collected eye-tracking data at a 30-Hz rate with a precision of ± 1.5°. The Microsoft HoloLens 2 system was selected for several reasons. At the time of the development of the solution (∼2021), this device was among the only technological solutions equipped with MR and eye-tracking capacities. Besides, the standalone system includes inertial sensors and a combination of sensing cameras that can produce a 3D scan map of the scene and thus enable the HoloLens to navigate in its referenced world. This integration eliminated the need for visual markers, unlike previously tested mobile eye-tracking systems that must rely on markers and OpenCV-based computer vision to detect surroundings and infer gaze position (Marois et al. 2021b). The relative ease of access to the Microsoft HoloLens 2 interface and app settings also made it more relevant for developing a solution tailorable and applicable to different use cases and work environments.

Scantracker is designed to elicit three types of notifications to counter: (i) negligence of areas of interest (in the use case of surveillance, camera feeds); (ii) attentional tunnelling; and (iii) vigilance decrement over time (see Figure 5). Within the MR version of Scantracker, notifications are programmed to be displayed on the visor of the Microsoft HoloLens 2 system. Negligence notifications took the form of four yellow chevrons (angular brackets) flashing around a neglected camera or in the top right section of the field of view for cameras out of sight. Attentional tunnelling notifications made participants aware that they were overfocusing on a particular camera. An orange bar

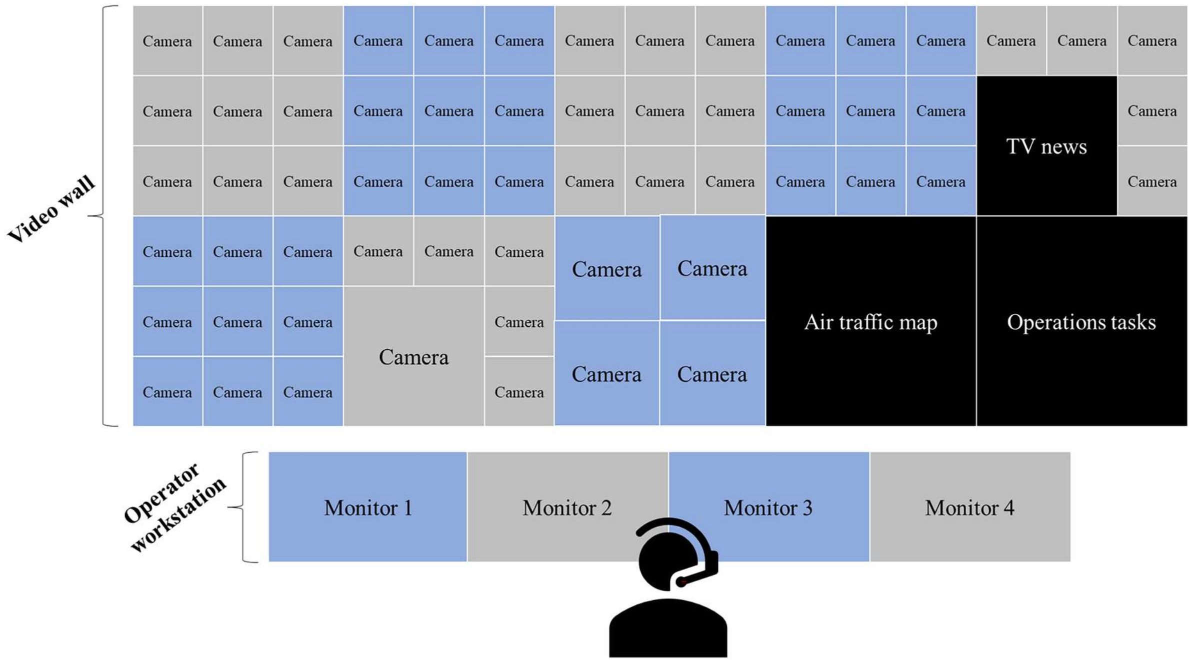

Figure 4. Depiction of the YQB operations centre surveillance environment model integrated into Scantracker, comprising a set of cameras on the video wall and the operator workstation. Note that the videowall content is adjustable and changes depending on operational needs. For the field study, however, the content remained as presented for all the surveillance periods.

progressed around that camera, starting from the upper left corner, and disappeared after 3 s or if the participant looked away. Vigilance notifications were represented by a small box depicting a clipboard and a cup of coffee in the lower right corner of the visual field. They encouraged participants to take a break by closing their eyes for 10 s, cued by a sound. Based on the gaze data, the vigilance model detected parallel tasks and added a 20-s delay to the vigilance model. An operator spending time on the monitors from their workstation performing clerical tasks would thus benefit from at least one delay before receiving a vigilance notification. Appendix A describes the algorithms and properties for each notification, screenshots of the solution, and the communication and integration protocol of the different systems involved in the Scantracker MR integration for data collection.

Scantracker is integrated with the Sensor Hub component to monitor gaze events and eye metrics for different visual behaviour models. Sensor Hub is a multipurpose software platform for bio-behavioural data fusion and real-time processing to provide assessments of cognitive and affective states (Gagnon et al., 2014). SensorHub allows easy integration of various sensors through ‘Drivers’ and implements the algorithms required to compute eye behaviours. The Sensor Hub ecosystem uses micro-services and communicates with other components via the MQTT protocol, the OASIS standard for IoT messaging. Sensor Hub publishes the information it collects over MQTT, which allows Scantracker to receive eye feature data and feed its behavioural models. Figure 6 depicts the communication protocol between Scantracker and Sensor Hub using the vigilance model as an example.

For Scantracker to analyse content from the environment and provide context-relevant notifications to the surveillance operators, we then integrated with the HoloLens 2. As shown in Figure 7 below, a Scantracker HoloLens app allowed for two main modes of operation, that is the Preparation Mode and the Standard Mode. The former was used for defining environmental modelling within the HoloLens Scantracker app (namely setting environmental model plane and adding scenes/ content to display notifications onto), and the latter served as the standard mode of operation which analyzed gaze coordinates, scenes and provided alerts per the Scantracker solution properties thanks to its communication with the Sensor Hub.

2.3. Procedure

Before taking part in the study, participants were asked to give informed consent. They were then informed by the research team that they could use Scantracker (as long and as often as they wished) or participate in the control data collection during their work shifts. For the Scantracker condition, at the beginning of each measurement session, they had to turn on the HoloLens

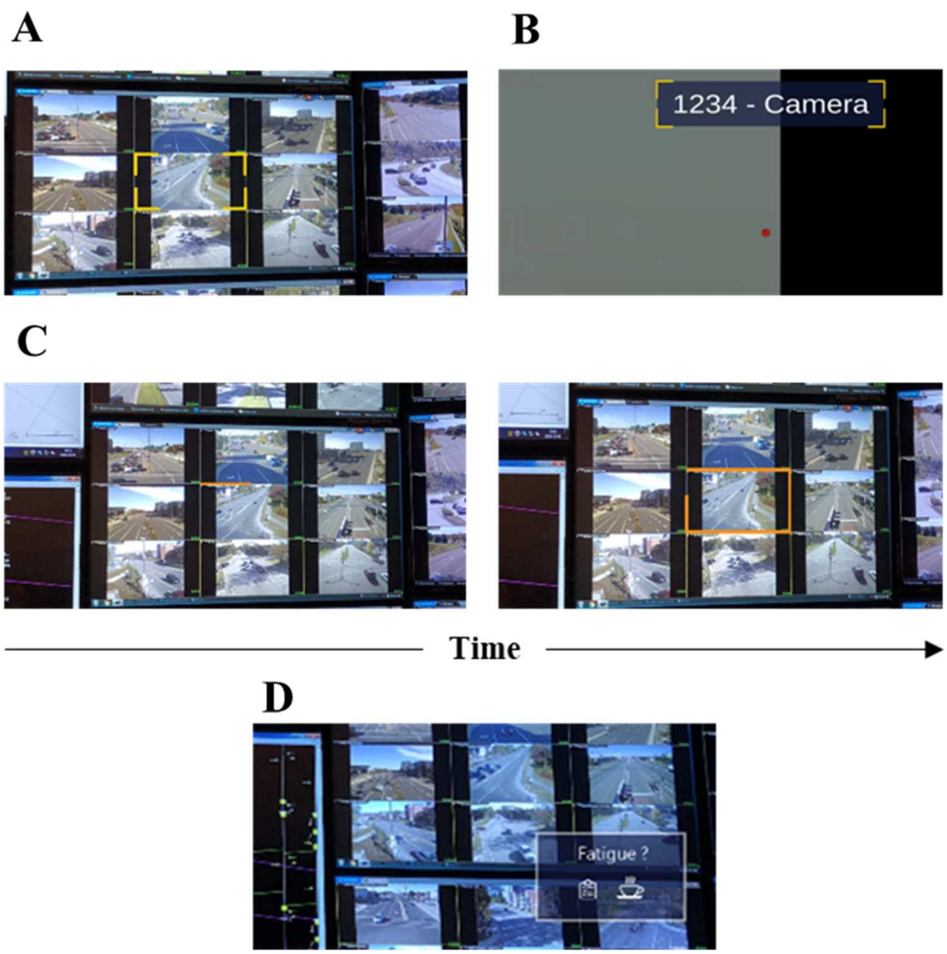

Figure 5. A. Example of a negligence notification around a neglected camera. B. Simulation of the negligence notification for a camera located outside the field of view of the user with the camera ID information. C. Example of the attention tunnelling notification with an orange progressive bar circling the over-focused camera. D. Example of the vigilance notification.

MR and the Scantracker app using a laptop that communicated with the MR system via Wi-Fi. The vigilance baseline was collected during the first 10 min of each session. For the control condition, participants only needed to note the exact start time of the session.

When a measurement session was over (i.e. at a break or when the work shift ended), participants completed questionnaires regarding that period. For the Scantracker condition, these questionnaires included: (a) the NASA-TLX workload questionnaire (Hart & Staveland, 1988) to measure subjective workload on a 10-point Likert scale from 1 (low) to 10 (high); (b) a questionnaire regarding the utility of each notification on a 10-point Likert scale from 1 (low) to 10 (high; Marois et al. 2021b); (c) a questionnaire concerning the ergonomic characteristics of the tool on a 5-point Likert scale from 1 (low) to 5 (high; inspired by an AR sickness questionnaire; Hussain, Park, and Kim 2023); and (d) sociodemographic questions. Participants in the Scantracker condition were also encouraged to provide qualitative feedback via an open question asking them to provide any comments on the tool and their experience. For the control condition, only the NASA-TLX and sociodemographic questions were presented.

2.4. Analyses

All analyses were conducted using a Bayesian framework with a Hamiltonian Monte Carlo technique implemented in Stan (Stan Development Team, n.d.), with four chains of 4,000 iterations and an *adapt-delta* value of 0.99 to minimise divergent transitions and ensure stable convergence. To do so, we used the brms package (Bürkner 2017), along with bridgesampling (Gronau, Singmann, and Wagenmakers 2020) and bayestestR (Makowski, Ben-Shachar, and Lüdecke 2019) in R. To evaluate differences in the NASA-TLX subscales across conditions, we first used a Bayesian hierarchical two-sample test, implemented as a multilevel regression model including a fixed effect of Condition and a random intercept for participants to account for repeated participants within each condition.

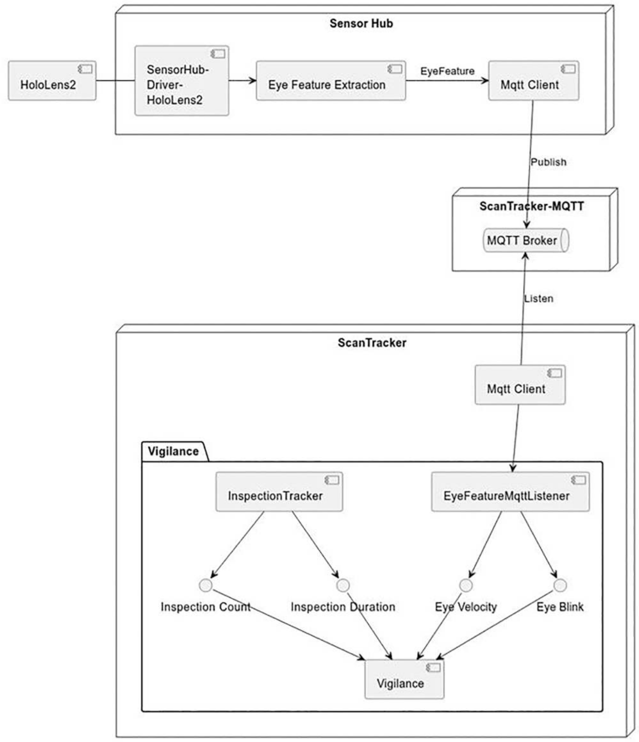

Figure 6. Depiction of the Scantracker communication protocol with the Sensor Hub system for the vigilance model.

The full model was compared with a null model containing only the intercept and the random effect of the participants to evaluate the impact of the Condition. For each NASA-TLX subscale model, we report the *SD* of the intercept (and 95% CI) to indicate the variability across participants, the β estimate for the impact of Scantracker over the control condition (with 95% CI), and the Bayes factor for the effect of the Condition under the alternative hypothesis (BF_{10}).

Subjective reports uniquely collected among the Scantracker condition (i.e. notifications utility and ergonomics properties) were also compared to their respective mid-scale values using a Bayesian hierarchical one-sample test including a random intercept for participants. The model with only the intercept (null model) was compared to the full model that also included the random effect of the participants. We report the variability of the intercepts ($SD_{\text{intercept}}$), the β estimate for the impact of the condition with respect to the reference value, and the BF_{10} for the difference with the reference value under H_1.

For all models, weakly informative priors were chosen to reflect reasonable expectations given the Likert scales. Intercepts and fixed effects were assigned normal (0, 1.5) priors given expectations that population-level means are likely to fall within the central portion of the scale while allowing uncertainty (McElreath 2016). Random-effect *SD*s and the residual error terms were assigned Student(3, 0, 1) priors to constrain variance parameters away from unrealistically large values while allowing potentially tailed distributions (a pattern typically observed in hierarchical models, see Bürkner 2017). BFs_{10} were computed using bridgesampling.

3. Results

3.1. Quantitative data

Twenty-five surveillance periods were completed by the participants over two one-week data collection periods, including 12 in the control condition and 13 in the Scantracker condition. In the control condition, two participants performed one session, two performed two sessions, and two performed three sessions (n = 12, M = 2.00, SD = 0.89). In the Scantracker condition, one participant performed one session, another performed three, a third performed four and the last one performed five sessions (n = 13, M = 3.25, SD = 1.71). The mean duration of the sessions was 62.82 min (SD = 16.29). Gaze validity of the eye-tracking data collected by the HoloLens in the Scantracker condition reached an average of 96.10% (SD = 6.08).

Table 1 depicts the subjective workload measures across sessions and conditions and the comparisons of the NASA-TLX subscales across conditions, as well as the results of the Bayesian hierarchical two-sample test. As shown in the table, reports on the different

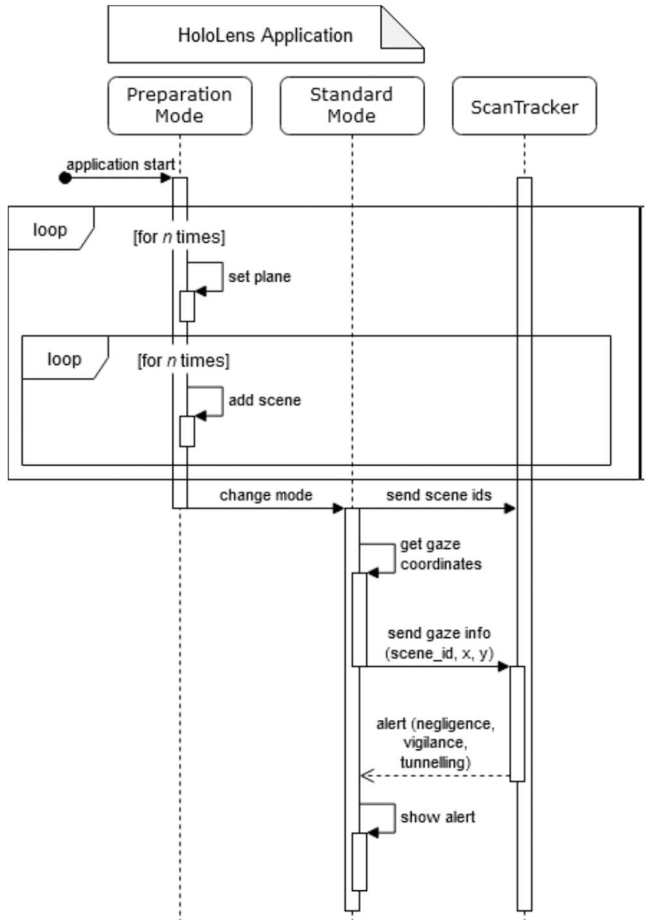

Figure 7. Sequence diagram of the HoloLens Scantracker application communicating with Scantracker.

NASA-TLX subscales were generally similar across conditions, except for the performance subscale, where higher performance values were reported for the control condition compared with the Scantracker condition. The Bayesian models for all subscales, except the Performance subscale, supported an absence of difference across conditions, with $BFs_{10} < 0.79$, suggesting anecdotal evidence for H_0 compared to H_1, also supported by the 95% CI. As for the Performance subscale, we observed a Bayes value of $BF_{10} = 1.25$, which suggests anecdotal evidence for H_1. However, this evidence was not supported by the 95% CI of the β estimate.

Table 2 presents the results of the ergonomic and notification utility measures for the Scantracker

Table 1. Measures of workload (means and SD, in parentheses) on the NASA-TLX subscales as a function of the condition and results of the Bayesian hierarchical two-sample tests.

	Mean values (*SD*)		Bayesian models[a]		
NASA-TLX subscale	Control ($n = 12$)	Scantracker ($n = 13$)	$SD_{\text{intercept}}$	β	BF_{10}
Mental	2.58	2.77	0.72	0.07	0.42
	(1.38)	(1.30)	[0.05, 1.64]	[−1.31, 1.31]	
Physical	1.58	1.62	0.44	0.03	0.36
	(1.73)	(0.65)	[0.02, 1.19]	[−1.14, 1.19]	
Temporal	2.42	2.39	0.52	−0.04	0.44
	(1.83)	(1.61)	[0.02, 1.46]	[−1.45, 1.35]	
Effort	2.42	2.08	1.47	−0.59	0.78
	(1.98)	(0.76)	[0.41, 2.70]	[−2.33, 1.06]	
Performance	9.42	6.23	3.45	−1.13	1.25
	(2.02)	(1.79)	[1.70, 6.65]	[−3.46, 1.35]	
Frustration	2.00	2.46	0.41	0.31	0.48
	(1.41)	(1.56)	[0.01, 1.17]	[−1.00, 1.55]	

[a]The results refer to the Bayesian hierarchical two-sample models comprised of the Condition as a fixed variable and the participant as a random variable, with the variability component across participants (i.e. *SD* of the random intercept, with 95% CI) and estimate of the effect of the Scantracker condition over the control condition (with 95% CI). The BF_{10} represents the comparison between both conditions for H_1.

Table 2. Measures of ergonomic properties and notification utility (means and SD, in parentheses) in the Scantracker condition and results of the Bayesian hierarchical one-sample tests.

		Bayesian models[a]		
Measure	Scantracker condition ($n = 13$)	$SD_{intercept}$	β	BF_{10}
Ergonomic properties (out of 5) – Test value = 3.0				
Comfort	2.85 (0.99)	1.01 [0.33, 2.20]	−0.15 [−1.23, 0.92]	0.33
Necessary breaks[a]	0.31 (0.48)	0.39 [0.02, 1.30]	−2.56 [−3.05, − 1.81]	291.87
Headache level	1.23 (0.44)	0.23 [0.01, 0.89]	−1.74 [−2.16, − 1.28]	277.80
Eye irritation	1.77 (1.24)	1.35 [0.66, 2.75]	−1.12 [−2.33, 0.28]	2.24
Nausea[b]	1.00 (0.00)	-	-	-
Visual quality	2.50 (1.24)	1.36 [0.60, 2.96]	−0.29 [−1.72, 1.19]	0.49
Visual fatigue	2.23 (1.24)	1.25 [0.59, 2.55]	−0.61 [−1.75, 0.70]	0.76
Global ease	3.58 (0.90)	0.86 [0.12, 2.25]	0.57 [−0.66, 1.67]	0.78
Notification utility (out of 10) – Test value = 5.5				
Negligence	4.00 (3.22)	2.60 [1.39, 4.86]	−0.83 [−2.72, 1.28]	1.05
Tunnelling	2.33 (3.21)	2.25 [0.17, 5.78]	−1.07 [−3.41, 1.53]	1.26
Vigilance	3.57 (2.23)	1.75 [0.23, 4.20]	−1.03 [−2.83, 1.19]	1.24

[a]The results refer to the Bayesian hierarchical one-sample models comparing the data to a reference value while adding the participant as a random variable, with the variability component across participants (i.e. *SD* of the random intercept, with 95% CI) and estimate of the deviation from the reference value (with 95% CI). The BF_{10} represents the comparison with the reference value for H_1.

condition, with the results of the Bayesian hierarchical one-sample tests. For all measures, *SD* of the intercept supported the presence of a participant random effect. Regarding the ergonomic properties of Scantracker, the number of necessary breaks and headache level were more likely to be rated under the midscale level. Data for breaks and headache levels were estimated to be 2.56 and 1.74 points under the reference value. This was supported by extreme evidence for H_1, with $BF_{10} = 291.87$ and $BF_{10} = 277.80$, for necessary breaks and headache level, respectively. The Bayesian model for the eye irritation measure showed anecdotal evidence in support of H_1 ($BF_{10} = 2.24$), but the 95% CI of the β estimate did not support this pattern, suggesting that it failed to differ from the midscale level. Comfort, visual quality and global ease were similar to the midscale level (with $BFs_{10} < 0.79$, providing anecdotal evidence for H_0). Measures of notification utility all suggested poor evidence favouring H_1 and important variability across participants, as shown by the $SD_{intercept}$. Analysis of the β estimate and the 95% CI however suggests that these differences with the reference value were not reliable.

3.2. Qualitative feedback from users

Following each surveillance session in the Scantracker condition, participants were encouraged to provide qualitative feedback about their experience with the tool. They were asked to consider the full integrated solution, that is the Scantracker system as well as the HoloLens 2 device. First, participants seemed to consider Scantracker useful to help them remember to look at the cameras while performing other clerical/secondary tasks. Participant A stated: '*I performed as I usually do, but I could see that the [Scantracker] helped me to be more conscious of monitoring the cameras more often*'. In another session, Participant A indicated: '*The [Scantracker] does not change my work habits (I can continue working on my other non-related surveillance tasks), this is one of the reasons why I think the notifications are relevant, because they allow me to stop neglecting the cameras*'.

In a later session, Participant A explained why they received many negligence notifications: '*I had a lot of other tasks to perform on my computer, so I tended to look less at the cameras and, consequently, I received many negligence notifications*'. On the opposite, Participant B noted: '*The focus on my [clerical] work was inferior, because I was constantly waiting for other notifications*'. Participant C reported: '*This can be relevant for employees who do not have the reflex of looking at the cameras*'.

An issue raised by the participants is the fact that the hand menu sometimes appeared automatically when the hands of the user were in their visual field. As stated by

Participant A: '*The menus were always appearing as soon as I was lowering my eyes*'. This even entailed a usability issue. As indicated by Participant A in another session: '*[…] once or twice, the negligence notifications were deactivated [through the hand menu] because I was typing on the keyboard and looking at it, so the notifications menu appeared without me wanting it*'. One participant reported that they felt a bit overwhelmed by the notifications. As they put it: '*[…] while we are trying to look at certain cameras, we receive other notifications to look at others. This is disruptive*'. Finally, some feedback was also provided by the users in terms of ergonomic properties of the HoloLens headset. Participant B specified that the tool was a bit heavy while Participant C indicated: '*The light tint of the visor can disrupt the other tasks, for instance when I'm looking at the phone display which is already dark, [I cannot see the digits]*'.

Despite these aspects, it seems that using Scantracker becomes easier with time. Participant D stated during their second session: '*Easier, no significant problem to report*'. Moreover, a participant raised that starting Scantracker, navigating through the app and using it became easier across sessions. Participants also quickly became proficient at troubleshooting and at reacting appropriately to some of the difficulties that arose with the tool. In some cases, the gaze of the user was not detected by Scantracker because the HoloLens inner software failed at recognising the room. When this occurred, participants were required to reset the HoloLens device and the app. They could also cross-check with a Debug mode whether the different scenes preprogrammed through the Preparation Mode in the HoloLens app were indeed detected. User feedback indicated that participants became increasingly proficient and comfortable with performing these troubleshooting tasks over time.

4. Discussion

The goal of the current study was to evaluate the effects and perceptions of the Scantracker tool on the work of surveillance operators in a real-life setting. As a subgoal, we wished to provide a description of the UCD process that we applied to improve the technological readiness of Scantracker, moving from the lab to the field. To serve our main goal, surveillance operators from the YQB airport operation centre underwent their security surveillance routine while either using the MR implementation of Scantracker or without using any tool. They were questioned on their perceived workload during the surveillance periods. Operators using Scantracker were also asked to answer a series of questions regarding the ergonomic properties as well as usability and utility of the tool, and to provide qualitative feedback.

Perceived workload was similar across conditions for all but one of the NASA-TLX subscales. Participants reported higher self-evaluated performance in the control condition, though the β estimate of the Bayesian model had wide credible intervals. This suggests that using Scantracker did not substantially impact perceived workload. Compared qualitatively, perceived workload in the current field study was lower than in previous laboratory studies. In Marois et al. (2020), the average workload was 5.19, in Williot et al. (2024), which tested the MR implementation of Scantracker, it was 5.14. In the current study, field experts reported lower workload overall ($M = 2.68$ with Scantracker and $M = 2.10$ in the control condition).

The higher performance ratings in the control condition may reflect that some participants found it difficult to keep track of which cameras to monitor, whereas Scantracker increased their awareness of monitoring behaviours. As some participants noted in qualitative feedback, especially for the negligence notifications, Scantracker helped them become more conscious of their monitoring strategies and performance. In the absence of notifications, operators may be less aware of, or less likely to actively reflect on, their monitoring behaviours. The Scantracker solution may therefore help operators remember to monitor cameras even when managing concurrent tasks (e.g. clerical tasks or discussions with colleagues and supervisors). However, this interpretation should be taken with caution given that the Bayes factor only provided anecdotal evidence for the alternative hypothesis.

Measures of ergonomic properties indicated that the MR integration of Scantracker was moderately acceptable, with low levels of headache and ocular fatigue. However, comfort and visual quality ratings were below the midpoint of the ergonomics questionnaire scale, though not statistically significant from it. This suggests that the MR-integrated solution could be improved in terms of notification visual rendering and comfort with the HoloLens. Other MR hardware technologies could be explored to determine how device characteristics impact these measures.

Measures of notification utility show that surveillance operators viewed the negligence notification as the most useful, followed by the vigilance notification and the tunnelling notification. Overall, utility ratings were near midscale, which was lower than those reported in a previous laboratory study (with means of 7.0, 5.5 and 5.0 out of 10 for the negligence, tunnelling and vigilance notifications, respectively; Marois et al. 2021b). One might interpret this difference as indicating that experts perceived Scantracker as less useful. However, this may suggest that experienced surveillance operators are less reliant on the tool for routine operations due to their expertise in performing the task daily. Perceived usefulness of decision aids varies with the task experience of users and task demands (Cohen, Parasuraman, and Freeman 1998; Mosier and Skitka 1996). Scantracker may offer particular value for: (1) assisting experienced operators during high-workload situations; and (2) supporting novice operators during initial training. Scantracker could indeed support training and the development of optimal strategies (e.g. by encouraging a parity monitoring strategy, which can increase the likelihood of detecting incidents; Hodgetts et al. 2018). Eye tracking is often used as a tool to evaluate an individual's ability to follow specific protocols during training (e.g. Guo et al. 2022; Tien et al. 2014), or as a comparison tool between experts and novices (Borg et al. 2018; Lounis, Peysakhovich, and Causse 2021; see also Rosch & Vogel-Walcutt, 2013, for a review). Further work is, however, needed before promoting such uses for Scantracker, given that its operational effectiveness in the field has yet to be more thoroughly investigated.

The collected qualitative feedback suggests that, once familiarised with the tool, Scantracker can help surveillance operators turn their attention toward the cameras to monitor and remember to look at them every now and then. It also supports that when operators are less focused on the cameras, Scantracker successfully triggers negligence notifications. However, this also means that Scantracker can sometimes disrupt clerical tasks. These observations are informative on the way the tool should be implemented into the workflow. For some users, the notifications should not be presented while they are already actively looking at the cameras. The conceptual principle of Scantracker rests on the idea that surveillance operators monitor the cameras almost continuously, and it is mainly driven by bottom-up principles whereby notifications depend on gaze data. An adaptive functionality could be developed to reduce the disruptive impact of the tool and to pinpoint neglected cameras only when the operator is performing a secondary task (i.e. engaging in a parallel task) or to preprogramme systematic notifications for higher-priority cameras. Such a hybrid bottom-up and top-down approach might prove useful, as both processes can be complementary (Katsuki and Constantinidis 2014) and can differently influence visual search strategies (e.g. Foulsham et al. 2014).

Qualitative feedback from users also revealed some hardware-related concerns, including the unexpected appearance of the hand menu when hands were detected by the HoloLens, the weight of the headset, and the light tint of the visor. This feedback highlights the benefits of testing Scantracker in a real-world environment. Although these issues are not related to Scantracker per se, they must still be considered, as the current integration relies on the HoloLens 2 device. Such considerations may even be relevant for other display technologies (e.g. head-up displays).

Despite advances in surveillance technologies, security centres still strongly rely on human operators to detect critical events (Nicosia and Kristensson 2024). Cognitive human factors such as workload and limited attentional capacity have been shown to affect the ability of operators to detect critical incidents (see Hodgetts et al. 2017, for a review). Developing human-in-the-loop support systems, such as gaze-aware assistive technologies, to mitigate cognitive limitations may significantly enhance the effectiveness of surveillance work. A tool such as Scantracker could potentially alleviate the cognitive burden experienced by surveillance operators. At a metacognitive level, our study suggests that using assistive technologies can help surveillance operators to be more aware of the cameras they need to monitor. The support and perceived utility, however, may depend on both user characteristics and context, highlighting the necessity of providing flexible solutions that users can tailor to their needs. This has also been previously observed in other contexts involving assistive technologies (e.g. in assembly lines; Battini et al. 2011; Peron, Sgarbossa, and Strandhagen 2022).

4.1. Field studies in UCD: challenges and value

Our study adopted the UCD approach, which provides techniques for identifying user needs and iteratively designing, testing and refining solutions (Norman and Draper 1986; Still and Crane 2017). Specifically, we conducted a field study, a key UCD method proven valid for collecting user feedback in real-world contexts (den Dekker 2020; Eshet and Bouwman 2015; Hussain, Slany, and Holzinger 2009; Mao et al. 2009; Uebernickel and Brenner 2016; van Reine 2017). This step represents a central component of the UCD iterative process, promoting user evaluation to identify potential refinements and orient future developments (cf. Figure 1). Although user-centered studies involving interviews and workshops are common, field studies remain relatively scarce. Despite their perceived value, they are seldom conducted (Eshet, 2012; Eshet and Bouwman 2015; Hussain, Slany, and Holzinger 2009; Mao et al. 2009) largely due to significant time and resource demands (Bergvall-Kåreborn and Howcroft 2011; Vredenburg et al. 2002).

Field studies in operational environments present a range of methodological challenges that researchers must navigate (Stanton et al., 2013). Access to subject matter experts (SME) is often limited, as specialised professionals have constrained availability and restricted work schedules. Technical and security constraints inherent to operational environments (e.g. safety protocols, continuous operations requirements, and confidentiality concerns) can restrict data collection methods and experimental procedures. Resulting small sample sizes constrain statistical power and limit the precision of quantitative inferences. Additionally, field studies involve trade-offs between controlled experimentation and ecological validity (see McNeese, Bautsch, and Narayanan 1999). These challenges necessitate complementary analytical approaches, including qualitative methods and mixed-method designs that can accommodate smaller samples.

Despite these challenges, our study demonstrates the value of field studies for gaze-aware assistive technology development. Field studies with end-users in an operational environment yield insightful and ecologically valid data across multiple dimensions: objective performance metrics (e.g. gaze-based indicators of visual monitoring behaviour), user experience measures (e.g. workload, ergonomics, ease of use), and qualitative feedback. This comprehensive, naturalistic evaluation reveals aspects of real-world use that laboratory studies cannot, including how technology integrates with existing workflows, how users adapt tools to meet operational demands, and what unforeseen technical or ergonomic issues emerge in actual practice. Such findings are essential for determining whether emerging technologies are both effective and acceptable in operational settings. However, recent analyses show that only approximately 10% of human factors studies are conducted in real-world settings (Chen et al. 2025), which highlights the need for field studies to be conducted more frequently within the design, cognitive systems engineering, and human factors communities.

4.2. Limitations and future work

While the present field study provides empirical evidence that Scantracker can offer useful assistance to surveillance operators, some results must be taken with caution. The quasi-experimental approach adopted for the field study may reduce the validity of the feedback provided about Scantracker, because of the absence of experimental control over the moment the technology was used. In fact, participants from the field study were experienced surveillance operators performing their daily tasks as usual. In that regard, it remains difficult to provide strong inference over other surveillance environments and operators regarding how an MR-integrated tool such as Scantracker may affect workload.

As is often the case with field studies that involve SMEs, the size of the sample of participants can also be seen as a limitation. This constraint must, however, be interpreted in light of the methodological choices made. Conducting research in real-life surveillance operation centres imposes substantial challenges on recruitment given the necessity to study operators without disrupting critical operational work. As a result, many studies rely on old footage or simulations to be able to present enough incidents to assess detection performance (e.g. Dadashi, Stedmon, and Pridmore 2013; Gelertner 2013; Hodgetts et al. 2018; Marois et al. 2020; Rankin et al. 2012; van Voorthuijsen et al. 2005; Williot et al. 2024). In contrast, and in line with recent methodological recommendations supporting field studies in ergonomics and human factors research (cf. Chen et al. 2025), our study was conducted during real surveillance periods, therefore increasing ecological validity and providing a more rigorous test of Scantracker's effectiveness in supporting surveillance activities. This approach may limit the achievable sample size. Still, the mixed-method design that we relied on (i.e. mix of quantitative and qualitative feedback) helped overcome some of the limitations caused by the sample size. Specifically, the qualitative feedback collected from the participants in the Scantracker condition provided rich, complementary information that extended the quantitative findings.

Drawing lessons learned from the present study, future work could focus on extending this experiment with a larger set of participants and a longitudinal design to assess long-term effect of MR decision-support tools on surveillance activities, including workload measures, performance metrics, and other subjective evaluations of the solution using standardised usability assessment tools. Efforts could also be deployed to adapt the solution to different MR technologies. The potential of adapting the notifications to the context and to the users, as well as the possibility of using Scantracker in other visual monitoring contexts (e.g. aviation, industrial monitoring) could also be explored.

5. Conclusions

Our work proposed a method to test and analyze the perception of end-users toward a prototype gaze-aware support tool developed to enhance security surveillance. Ultimately, this could help keep surveillance personnel engaged, accountable and in charge given the known relationship between automation, motivation decrement and loss of situation awareness. Per the users' feedback, Scantracker seemed to provide appropriate support for aiding camera monitoring, although some components related to ergonomics and to context of use may need to be refined. Future developments could include top-down prioritisation of certain zones to monitor. Future work will aim at further adjusting the tool and conducting empirical work – with a larger sample size and objective performance measures – for testing the potential of Scantracker to alleviate the cognitive burden experienced by surveillance operators while sustaining effective monitoring behaviours.

Acknowledgements

We are grateful to Vincent Poiré, Jonathan Richard, Jean-Sébastien Thivierge, Alexandre Williot, and all the students involved in the data collection. Special thanks are also due to Eric R. Harvey, Bruno Martin, Jonathan Roy-Noël and to all the team from the *Centre en imagerie numérique et médias interactifs* (CIMMI) for their key role in contributing to develop the mixed reality components of the system and to Louis-Philippe Bellerose and Laura Salvan for their implication in the user experience interviews and workshop. We thank the surveillance operations centre at the *Centre intégré de gestion de la circulation* from the *Ministère des Transports et de la Mobilité durable du Québec* (MTQ) and the *Aéroport international Jean-Lesage de Québec* (YQB) and their operators for their collaboration throughout the data collection.

Author contributions

CRediT: **Alexandre Marois:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft; **Daniel Lafond:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – review & editing; **Sébastien Tremblay:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Innovation for Defence Excellence and Security (IDEaS) program and by Mitacs Canada.

References

Ahmed, A. A., and M. Echi. 2021. "Hawk-eye: An AI-Powered Threat Detector for Intelligent Surveillance Cameras." *IEEE Access* 9:63283–63293. https://doi.org/10.1109/ACCESS.2021.3074319.

Aitken, B. M., J. C. Champion, and M. J. Stainer. 2019. "Anxious Individuals Predict the Onset of Aggression Earlier in a CCTV Surveillance Task." *Journal of Experimental Psychology: Applied* 25:343–353. https://doi.org/10.1037/xap0000199.

Anderson, T., K. Fogarty, H. Kenkel, J. Raisigel, S. Zhou, L. M. Reggia, S. G. Manizade, D. R. Lesniak, and G. J. Gerling. 2021. "User Experience Design for Human-Machine Teaming in Commanding a Distributed Constellation of Unmanned Assets in Search and Rescue." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 65 (1): 62–66. https://doi.org/10.1177/1071181321651130.

Bagassi, S., F. De Crescenzio, S. Piastra, C. A. Persiani, M. Ellejmi, A. R. Groskreutz, and J. Higuera. 2020. "Human-in-the-loop Evaluation of an Augmented Reality Based Interface for the Airport Control Tower." *Computers in Industry* 123:103291. https://doi.org/10.1016/j.compind.2020.103291.

Barbotin, N., J. Baumeister, A. Cunningham, T. Duval, O. Grisvard, and B. H. Thomas. 2022. "Evaluating Visual Cues for Future Airborne Surveillance Using Simulated Augmented Reality Displays." In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 213–221. https://doi.org/10.1109/VR51125.2022.00040.

Battini, D., M. Faccio, A. Persona, and F. Sgarbossa. 2011. "New Methodological Framework to Improve Productivity and Ergonomics in Assembly System Design." *International Journal of Industrial Ergonomics* 41:30–42. https://doi.org/10.1016/J.ERGON.2010.12.001.

Bergvall-Kåreborn, B., and D. Howcroft. 2011. "Mobile Applications Development on Apple and Google Platforms." *Communications of the Association for Information Systems* 29:30. https://doi.org/10.17705/1CAIS.02930.

Blattgerste, J., P. Renner, B. Strenge, and T. Pfeiffer. 2018. "In-situ Instructions Exceed Side-by-Side Instructions in Augmented Reality Assisted Assembly." In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 133–140. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3197768.3197778.

Blechko, A., I. Darker, and A. Gale. 2008. "Skills in Detecting gun Carrying from CCTV." In *2008 42nd Annual IEEE International Carnahan Conference on Security Technology*, edited by M. Klima. IEEE Press. https://doi.org/10.1109/CCST.2008.4751312.

Bock, L., T. Bohné, and S. K. Tadeja. 2025. "Decision Support for Augmented Reality-Based Assistance Systems Deployment in Industrial Settings." *Multimedia Tools and Applications* 84:23617–23641. https://doi.org/10.1007/s11042-024-19861-x.

Borg, L. K., T. K. Harrison, A. Kou, E. R. Mariano, A. D. Udani, T. E. Kim, C. Shum, and H. K. Howard. 2018. "Preliminary Experience Using eye-Tracking Technology to Differentiate Novice and Expert Image Interpretation for Ultrasound-Guided Regional Anesthesia." *Journal of Ultrasound in Medicine* 37:329–336. https://doi.org/10.1002/jum.14334.

Bürkner, P.-C. 2017. "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80:1–28. https://doi.org/10.18637/jss.v080.i01.

Chen, Y., X. Chen, S. Chen, J. Xu, and P. Liu. 2025. "Investigating Human Factors and Ergonomics Research: A 4s Framework." *Ergonomics*, 1–17. https://doi.org/10.1080/00140139.2025.2537778.

Cohen, M., R. Parasuraman, and J. Freeman. 1998. "Trust in Decision Aids: A Model and Its Training Implications." In *Proceedings of the 1998 Command and Control Research and Technology Symposium*, 1–37.

Dadashi, N., A. W. Stedmon, and T. P. Pridmore. 2013. "Semi-automated CCTV Surveillance: The Effects of System Confidence, System Accuracy and Task Complexity on Operator Vigilance, Reliance and Workload." *Applied Ergonomics* 44:730–738. https://doi.org/10.1016/j.apergo.2012.04.012.

den Dekker, T. 2020. *Desing Thinking*. Utrecht: Noordhoff Uitgevers.

Dong, Y., Z. Hu, K. Uchimura, and N. Murayama. 2011. "Driver Inattention Monitoring System for Intelligent Vehicles: A Review." *IEEE Transactions on Intelligent Transportation Systems* 12:596–614. https://doi.org/10.1109/TITS.2010.2092770.

Edwards, J. D., B. A. Fausto, A. M. Tetlow, R. T. Corona, and E. G. Valdés. 2018. "Systematic Review and Meta-analyses of Useful Field of View Cognitive Training." *Neuroscience & Biobehavioral Reviews* 84:72–91. https://doi.org/10.1016/j.neubiorev.2017.11.004.

Eshet, E. 2012. "Human-Centered Design in Mobile Application Development." *International Journal of Mobile Human Computer Interaction* 4 (4): 1–21. https://doi.org/10.4018/jmhci.2012100101.

Eshet, E., and H. Bouwman. 2015. "Addressing the Context of use in Mobile Computing: A Survey on the State of the Practice." *Interacting with Computers* 27:392–412. https://doi.org/10.1093/iwc/iwu002.

Foulsham, T., C. Chapman, E. Nasiopoulos, and A. Kingstone. 2014. "Top-down and Bottom-up Aspects of Active Search in a Real-World Environment." *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale* 68:8–19. https://doi.org/10.1037/cep0000004.

Funk, M., J. Heusler, E. Akcay, K. Weiland, and S. Schmidt. 2016. "Haptic, Auditory, or Visual? Towards Optimal Error Feedback at Manual Assembly Workplaces." In *Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 1. https://doi.org/10.1145/2910674.2910683.

Gagnon, J.-F., S. Tremblay, D. Lafond, M. Rivest, and F. Couderc. 2014. "Sensor-Hub: A Real-time Data Integration and Processing Nexus for Adaptive C2 Systems." In *The Sixth International Conference on Adaptive and Self-Adaptive Systems and Applications*, 63–67. Valencia, Spain.

Gelertner, J. 2013. "Effective Threat Detection for Surveillance." In *IEEE International Conference on Technologies for Homeland Security*, 290–296. https://doi.org/10.1109/THS.2013.6699016.

Ghandour, R., B. Neji, A. M. El-Rifaie, and Z. Al Barakeh. 2020. "Driver Distraction and Stress Detection Systems: A Review." *International Journal of Engineering and Applied Sciences* 7:39–46. https://doi.org/10.31873/IJEAS.7.04.10.

Gill, M., R. Little, A. Spriggs, J. Allen, J. Argomaniz, and S. Waples. 2005. *Assessing the Impact of CCTV: The Hawkeye Case Study. Home Office Online Report*. London: Home Office.

Gronau, Q. F., H. Singmann, and E.-J. Wagenmakers. 2020. "bridgesampling: An R Package for Estimating Normalizing Constants." *Journal of Statistical Software* 92:1–29. https://doi.org/10.18637/jss.v092.i10.

Guo, Y., D. Freer, F. Deligianni, and G.-Z. Yang. 2022. "Eye-tracking for Performance Evaluation and Workload Estimation in Space Telerobotic Training." *IEEE Transactions on Human-Machine Systems* 52:1–11. https://doi.org/10.1109/THMS.2021.3107519.

Haines, E. R., M. A. Kirk, L. Lux, A. B. Smitherman, B. J. Powell, A. Dopp, A. M. Stover, and S. A. Birken. 2022. "Ethnography and User-Centered Design to Inform Context-Driven Implementation." *Translational Behavioral Medicine* 12:1–14. https://doi.org/10.1093/tbm/ibab077.

Hart, S. G., and L. E. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." *Advances in Psychology* 52: 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hodgetts, H. M., C. Chamberland, J.-D. Latulippe-Thériault, F. Vachon, and S. Tremblay. 2018. "Priority or Parity? Scanning Strategies and Detection Performance of Novice Operators in Urban Surveillance." In *Proceedings of the Human Factors and Ergonomics Society*, Vol. 62, 1541–1117. https://doi.org/10.1177/1541931218621255.

Hodgetts, H. M., F. Vachon, C. Chamberland, and S. Tremblay. 2017. "See No Evil: Cognitive Challenges of Security Surveillance and Monitoring." *Journal of Applied Research in Memory and Cognition* 6:230–243. https://doi.org/10.1016/j.jarmac.2017.05.001.

Hussain, M., J. Park, and H. K. Kim. 2023. "Augmented Reality Sickness Questionnaire (ARSQ): A Refined Questionnaire for Augmented Reality Environment." *International Journal of Industrial Ergonomics* 97:103495. https://doi.org/10.1016/j.ergon.2023.103495.

Hussain, Z., W. Slany, and A. Holzinger. 2009. "Current State of Agile User-Centered Design: A Survey." In *HCI and Usability for e-Inclusion. USAB 2009. Lecture Notes in Computer Science, vol 5889*, edited by A. Holzinger and K. Miesenberger, 416–427. Berlin: Springer. https://doi.org/10.1007/978-3-642-10308-7_30.

Katsuki, F., and C. Constantinidis. 2014. "Bottom-up and top-Down Attention: Different Processes and Overlapping Neural Systems." *The Neuroscientist* 20:509–521. https://doi.org/10.1177/1073858413514136.

Khan, M., A. E. Saddik, W. Gueaieb, G. De Masi, and F. Karray. 2024. "Vd-net: An Edge Vision-Based Surveillance System for Violence Detection." *IEEE Access* 12:43796–43808. https://doi.org/10.1109/access.2024.3380192.

Loizeau, Q., F. Danglade, F. Ababsa, and F. Merienne. 2021. "Methodology for the Field Evaluation of the Impact of Augmented Reality Tools for Maintenance Workers in the Aeronautic Industry." *Frontiers in Virtual Reality* 1:603189. https://doi.org/10.3389/frvir.2020.603189.

Lounis, C., V. Peysakhovich, and M. Causse. 2020. “Flight Eye Tracking Assistant (FETA): Proof of Concept.” In *Advances in Human Factors of Transportation. AHFE 2019. Advances in Intelligent Systems and Computing, vol. 964*, edited by N. Stanton, 739–751. Cham: Springer. https://doi.org/10.1007/978-3-030-20503-4_66.

Lounis, C., V. Peysakhovich, and M. Causse. 2021. “Visual Scanning Strategies in the Cockpit Are Modulated by Pilots’ Expertise: A Flight Simulator Study.” *PLoS ONE* 16:e0247061. https://doi.org/10.1371/journal.pone.0247061.

Makowski, D., M. Ben-Shachar, and D. Lüdecke. 2019. “Bayestestr: Describing Effects and Their Uncertainty, Existence and Significance within the Bayesian Framework.” *Journal of Open Source Software* 4:1541. https://doi.org/10.21105/joss.01541.

Mao, J.-Y., K. Vreenburg, P. W. Smith, and T. Carey. 2009. “The State of User-centered Design Practice.” *Communications of the ACM* 48:105–109. https://doi.org/10.1145/1047671.1047677.

Marois, A., H. M. Hodgetts, C. Chamberland, A. Williot, and S. Tremblay. 2021a. “Who Can Best Find Waldo? Exploring Individual Differences That Bolster Performance in a Security Surveillance Microworld.” *Applied Cognitive Psychology* 35: 1044–1057. https://doi.org/10.1002/acp.3837.

Marois, A., D. Lafond, A. Williot, F. Vachon, and S. Tremblay. 2020. “Real-time Gaze-aware Cognitive Support System for Security Surveillance.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 64:1145–1149. https://doi.org/10.1177/1071181320641274.

Marois, A., J. Roy-Noël, D. Lafond, A. Williot, E. R. Harvey, B. Martin, and S. Tremblay. 2022. “Adaptation of a Gaze-Aware Security Surveillance Support Tool for Augmented Reality.” In *Human Interaction, Emerging Technologies and Future Systems V. IHIET 2021. Lecture Notes in Networks and Systems, vol. 319*, edited by T. Ahram and R. Taiar, 781–789. Cham: Springer. https://doi.org/10.1007/978-3-030-85540-6_99.

Marois, A., L. Salvan, D. Lafond, A. Williot, N. Lemaire, and S. Tremblay. 2021b. “Improving Usability of a Gaze-Based Surveillance Support Tool through User-centered Design.” In *Advances in Usability, User Experience, Wearable and Assistive Technology. AHFE 2021*. Lecture Notes in Networks and Systems, vol. 275, edited by T. Z. Ahram and C. S. Falcão, 732–740. Cham: Springer. https://doi.org/10.1007/978-3-030-80091-8_87.

McElreath, R. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. New York: Champman and Hall/CRC. https://doi.org/10.1201/9781315372495.

McNeese, M. D., H. S. Bautsch, and S. Narayanan. 1999. “A Framework for Cognitive Field Studies.” *International Journal of Cognitive Ergonomics* 3:307–331. https://doi.org/10.1207/s15327566ijce0304_3.

McNeese, N. J., M. Demir, N. J. Cooke, and C. Myers. 2017. “Teaming with a Synthetic Teammate: Insights into Human-Autonomy Teaming.” *Human Factors* 60:262–273. https://doi.org/10.1177/0018720817743223.

Mosier, K. L., and L. J. Skitka. 1996. “Human Decision Makers and Automated Decision Aids: Made for Each Other?” In *Automation and Human Performance: Theory and Applications*, edited by R. Parasuraman and M. Mouloua, 201–220. Mahwah, NJ: Routledge.

Mucchielli, L. 2016. “À quoi sert la vidéosurveillance de l’espace public ?” *Déviance et Société* 40:25–50. https://doi.org/10.3917/ds.401.0025.

Nicosia, M., and P. O. Kristensson. 2024. “Risk Management in Human-in-the-Loop AI-Assisted Attention Aware Systems.” In *Putting AI in the Critical Loop*, edited by P. Dasgupta, J. Llinas, T. Gillespie, S. Fouse, W. Lawless, R. Mittu, and D. Sofge, 81–92. Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-0-443-15988-6.00013-3.

Nielsen Norman Group. 2016. *Field Studies*. https://www.nngroup.com/articles/field-studies/.

Norman, D. A., and S. W. Draper. 1986. *User Centered System Design: New Perspectives on Human-Computer Interaction*. Boca Raton: Erlbaum Associates.

Parasuraman, R., and D. H. Manzey. 2010. “Complacency and Bias in Human use of Automation: An Attentional Integration.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 52:381–410. https://doi.org/10.1177/0018720810376055.

Pelletier, S., J. Suss, F. Vachon, and S. Tremblay. 2015. “Atypical Visual Display for Monitoring Multiple CCTV Feeds.” In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1145–1150. https://doi.org/10.1145/2702613.2732840.

Peron, M., F. Sgarbossa, and J. O. Strandhagen. 2022. “Decision Support Model for Implementing Assistive Technologies in Assembly Activities: A Case Study.” *International Journal of Production Research* 60:1341–1367. https://doi.org/10.1080/00207543.2020.1856441.

Pillajo, E., C. Mourgues, A. Neyem, and V. A. González. 2025. “An Interface Design Method Based on Situation Awareness and Immersive Analytics for Augmented and Mixed Reality Decision Support Systems in Construction.” *Applied Sciences* 15:7820. https://doi.org/10.3390/app15147820.

Piza, E. L., and L. N. Moton. 2023. “Proactive Monitoring and Operator Discretion: A Systematic Social Observation of CCTV Control Room Operations.” *Journal of Criminal Justice* 86:102071. https://doi.org/10.1016/j.jcrimjus.2023.102071.

Rankin, S., K. Cohen, K. MacLennan-Brown, and K. Sage. 2012. “CCTV Operator Performance Benchmarking.” In *IEEE International Carnahan Conference on Security Technology*, 325–330. https://doi.org/10.1109/CCST.2012.6393580.

Rokhsaritalemi, S., A. Sadeghi-Niaraki, and S.-M. Choi. 2020. “A Review on Mixed Reality: Current Trends, Challenges and Prospects.” *Applied Sciences* 10:636. https://doi.org/10.3390/app10020636.

Rosch, J. L., and J. J. Vogel-Walcutt. 2013. “A Review of Eye-Tracking Applications as Tools for Training.” *Cognition, Technology & Work* 15 (3): 313–327. https://doi.org/10.1007/s10111-012-0234-7.

Rupert, A. H., G. Woo, J. C. Brill, and B. Lawson. 2016. “Countermeasures for Loss of Situation Awareness: Spatial Orientation Modeling to Reduce Mishaps.” *2016 IEEE Aerospace Conference*. https://doi.org/10.1109/AERO.2016.7500725.

Santoni de Sio, F. S., and J. Van den Hoven. 2018. "Meaningful Human Control over Autonomous Systems: A Philosophical Account." *Frontiers in Robotics and AI* 5:15. https://doi.org/10.3389/frobt.2018.00015.

Stanton, N. A., P. M. Salmon, L. A. Rafferty, G. H. Walker, C. Baber, and D. P. Jenkins. 2013. *Human Factors Methods: A Practical Guide for Engineering and Design*. London: CRC Press. https://doi.org/10.1201/9781315587394.

Steane, V. A., S. Hart, J. Oakes, S. Palmer, and M. Chattington. 2023. "Using EAST to Inform Systems Architecture Design: Considerations Relating to the use of UAVs in Search and Rescue Missions." In *Human Factors in Robots, Drones and Unmanned Systems. AHFE (2023) International Conference, AHFE Open Access, vol. 93* edited by T. Ahram and W. Karwowski, 132–141. USA: AHFE International. https://doi.org/10.54941/ahfe1003755.

Still, B., and K. Crane. 2017. *Fundamentals of User-centered Design: A Practical Approach*. Boca Raton: CRC Press, Taylor & Francis Group.

Suss, J., F. Vachon, D. Lafond, and S. Tremblay. 2015. "Don't Overlook the Human! Applying Principles of Cognitive Systems Engineering to the Design of Intelligent Video Surveillance Systems." In *Proceedings of the 12th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 1–6. https://doi.org/10.1109/AVSS.2015.7301795.

Tang, A., C. Owen, F. Biocca, and W. Mou. 2003. "Comparative Effectiveness of Augmented Reality in Object Assembly." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 73–80. https://doi.org/10.1145/642611.642626.

Taylor, P., N. Bilgrien, Z. He, and H. T. Siegelmann. 2015a. "EyeFrame: Real-time Memory aid Improves Human Multitasking via Domain-General eye Tracking Procedures." *Frontiers in ICT* 2:17. https://doi.org/10.3389/fict.2015.00017.

Taylor, P., Z. He, N. Bilgrien, and H. T. Siegelmann. 2015b. "Human Strategies for Multitasking, Search, and Control Improved via Real-time Memory aid for Gaze Location." *Frontiers in ICT* 2:15. https://doi.org/10.3389/fict.2015.00015.

Tien, T., P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G. Z. Yang, and A. Darzi. 2014. "Eye Tracking for Skills Assessment and Training: A Systematic Review." *Journal of Surgical Research* 191:169–178. https://doi.org/10.1016/j.jss.2014.04.032.

Tremblay, S., D. Lafond, C. Chamberland, H. M. Hodgetts, and F. Vachon. 2018. "Gaze-aware Cognitive Assistant for Multiscreen Surveillance." In *Advances in Intelligent Systems and Computing*, vol. 722, edited by W. Karowski and T. Ahram, 230–236. Cham: Springer. https://doi.org/10.1007/978-3-319-73888-8_36.

Uebernickel, F., and W. Brenner. 2016. "Design Thinking." In *Business Innovation: Das St. Galler Modell*, edited by C. Hoffmann, S. Lennerts, C. Schmitz, W. Stölzle, and F. Uebernickel, 243–265. Wiesbaden: Business Innovation Universität St. Gallen. Springer Gabler. https://doi.org/10.1007/978-3-658-07167-7_15.

van Reine, P. P. 2017. "The Culture of Design Thinking for Innovation." *Journal of Innovation Management* 5:56–80. https://doi.org/10.24840/2183-0606_005.002_0006.

van Voorthuijsen, G., H. van Hoof, M. Klima, K. Roubik, M. Bernas, and P. Pata. 2005. "CCTV Effectiveness Study." In *International Carnahan Conference on Security Technology*, 105–108. https://doi.org/10.1109/CCST.2005.1594815.

Vredenburg, K., J.-L. Mao, P. W. Smith, and T. Carey. 2002. "A Survey of User-Centered Design Practice." In *CHI '02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 471–478. https://doi.org/10.1145/503376.503460.

Williot, A., D. Lafond, S. Tremblay, and A. Marois. 2024. "Surveillance-behavior Support by a Real-Time Gaze-Based Tool Integrated with Augmented Reality." In *2024 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 1–7. https://doi.org/10.1109/CogSIMA61085.2024.10553855.

Wu, S., L. Hou, G. Zhang, and H. Chen. 2022. "Real-time Mixed Reality-based Visual Warning for Construction Workforce Safety." *Automation in Construction* 139:104252. https://doi.org/10.1016/j.autcon.2022.104252.

Appendix A: Details on Scantracker notifications

Scantracker produced three kinds of notifications based on eye movement criteria.

Negligence

Negligence notifications took the form of four yellow chevrons flashing around a neglected camera within the field of view or in the top right section of the field of view for cameras out of sight. They were triggered when no inspection of the content of a particular camera was detected for at least 10 min. An inspection is a period of one second, updated every second, comprising at least 20 gaze data measured during this time and 80% of them being collected on a given camera (Figure A1). A minimal grace period of 2 min was added between two different negligence notifications. Negligence notifications disappeared when the participant moved their gaze upon the neglected camera or after a period of 40 s.

Attention tunnelling

Tunnel notifications made participants aware that they were over-focusing on a particular camera. A notification was triggered when, for a period of 2 min, at least 1,200 gaze points were recorded and 80% were related to a single camera feed. In this case, an orange bar would progress around that camera feed for 3 s, starting from the upper left corner. The notification disappeared after 3 s, or if the participant looked away from the over-focused camera. A minimal period of 28 s separated any two tunnel notifications. Figure A2 displays the sequence for a tunnelling notification to be triggered.

Vigilance

Vigilance notifications were represented by a small box depicting a clipboard and a cup of coffee, in the lower right corner of the visual field. They notified a decrement in vigilance and encouraged participants to take a break by closing their eyes for 10 s, cued by a sound. Vigilance notifications were triggered if: (a) the number, (b) the duration, or (c)

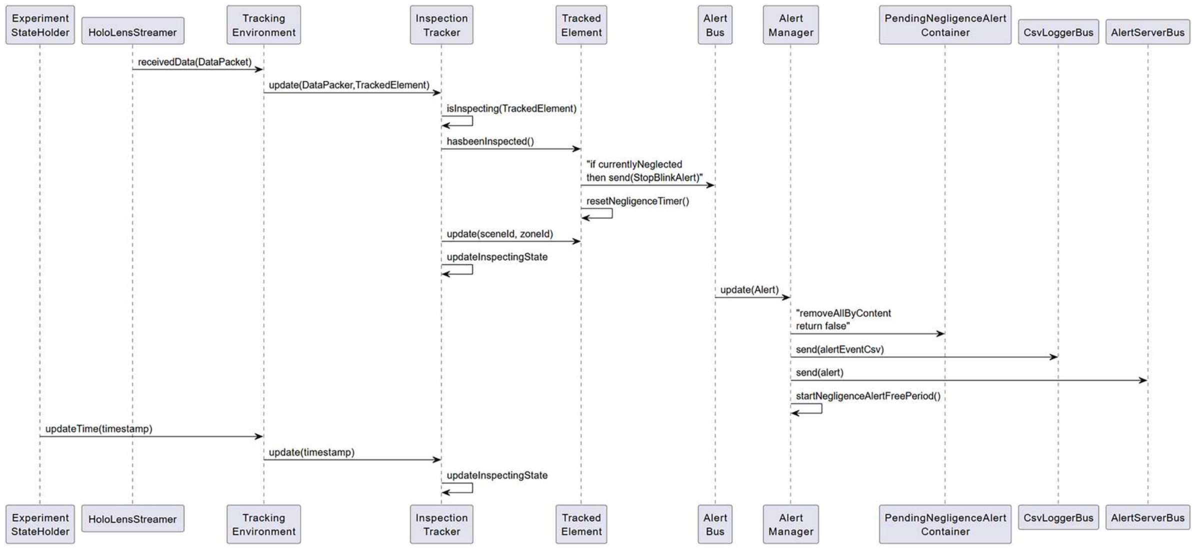

Figure A1. Sequence diagram for inspection detection and negligence notifications.

the velocity of an inspection – i.e. the observation of at least 80% of gaze data within each 1 s period within the same zone – was under 1.5 *SD* of a normative baseline determined during the first 10 min of each scenario (or baselines self-initiated by participants; see Figure A3). The vigilance notification relied on a 30-s rolling time window for the identification of the number of fixations and their duration. Saccadic velocity rather relied on 5-s rolling time windows. A minimal delay of 20 s was inserted between two vigilance notifications. Figure A4 summarises the sequence for a vigilance notification to be elicited.

Based on the gaze data, the vigilance notification model considered a parallel task detection feature that adds a 20 s delay to the three notification models when 80% gaze data (with a minimum count of 10 gaze points) was detected upon a specific zone of their work environment during a 4-s rolling time window. An operator spending time on the monitors from their workstation performing clerical tasks would thus benefit from at least one 20 s delay added to the vigilance notification. Figure A5 depicts the sequence for the detection of a parallel task.

Notifications Protocol

All the different notifications presented depended on a protocol involving a Message Queuing Telemetry Transport (MQTT) workflow. This allowed triggering the notifications according to the different properties presented above. Figure A6 depicts how the environment and the predefined zones are being tracked, while Figure A7 presents the full MQTT protocol workflow used to present the notifications.

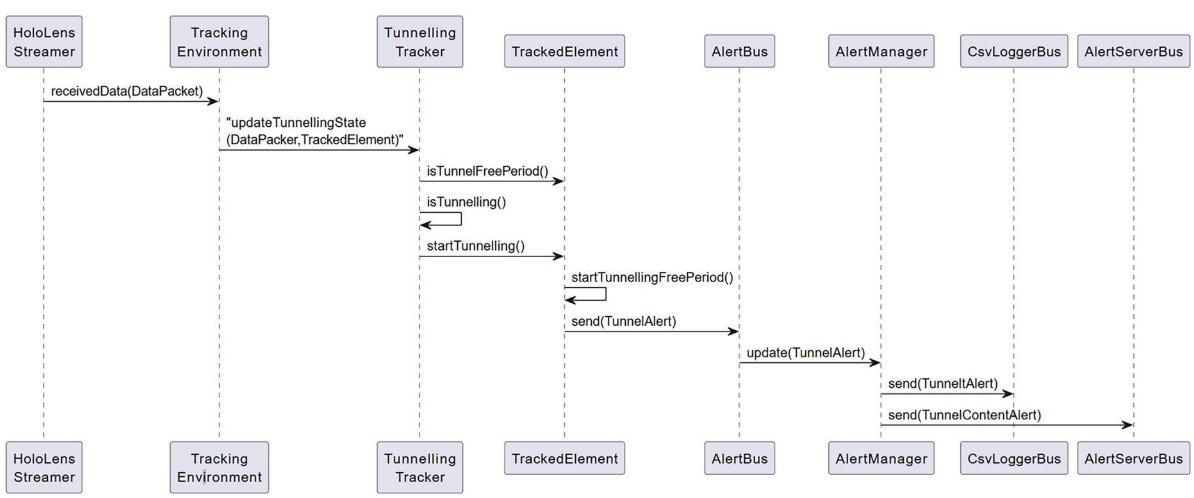

Figure A2. Sequence diagram for the tunnelling notification.

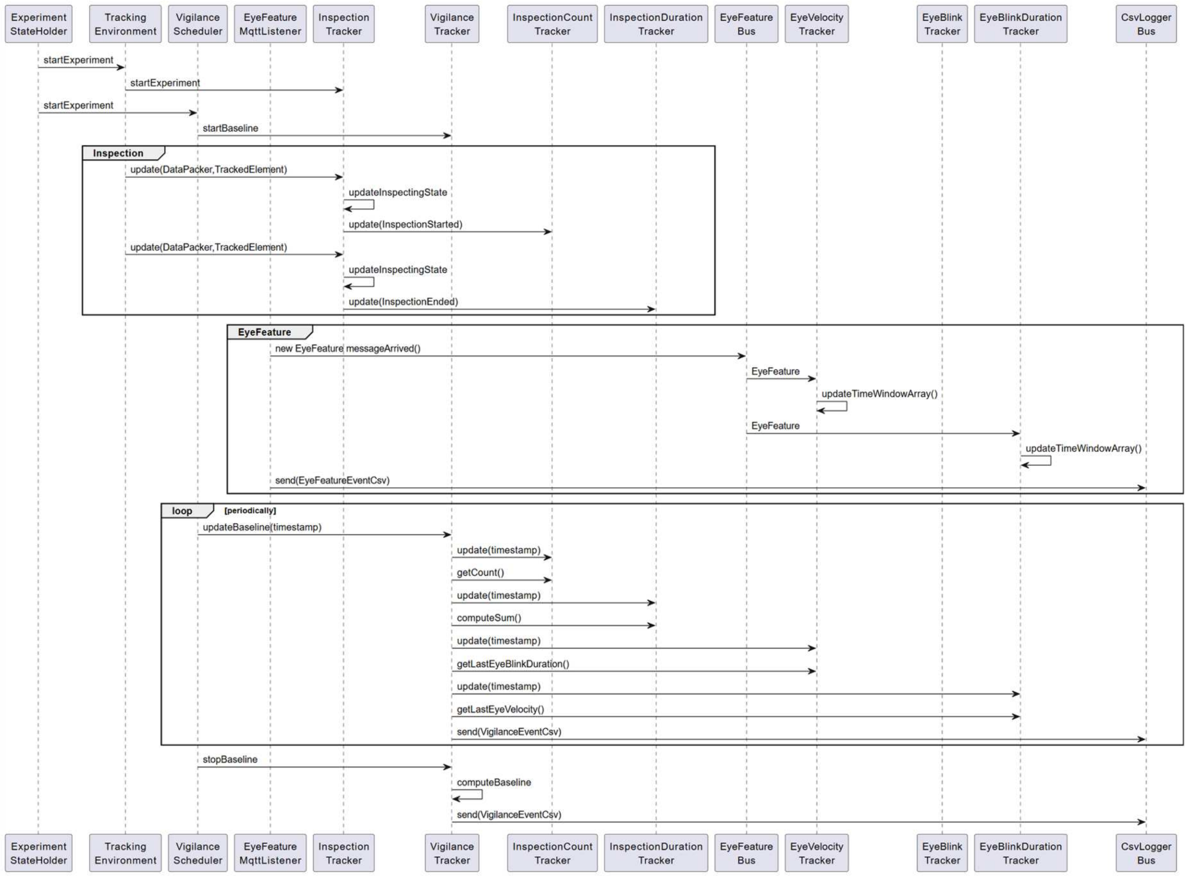

Figure A3. Sequence diagram for the vigilance baseline.

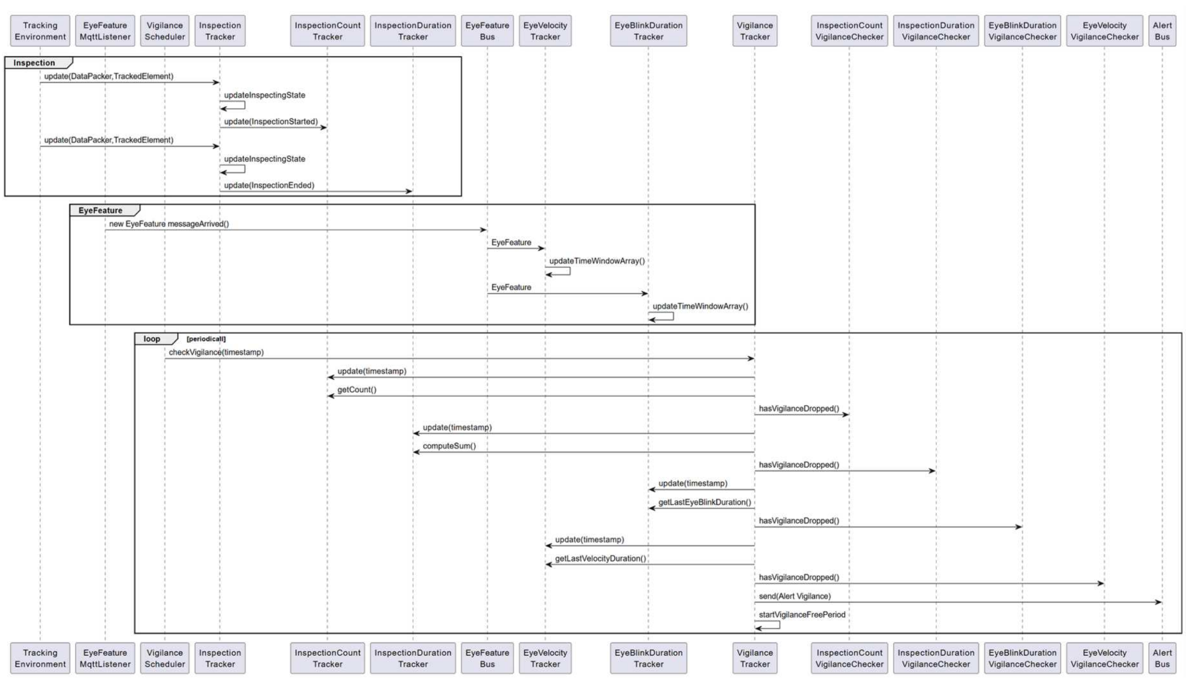

Figure A4. Sequence diagram for the vigilance notification.

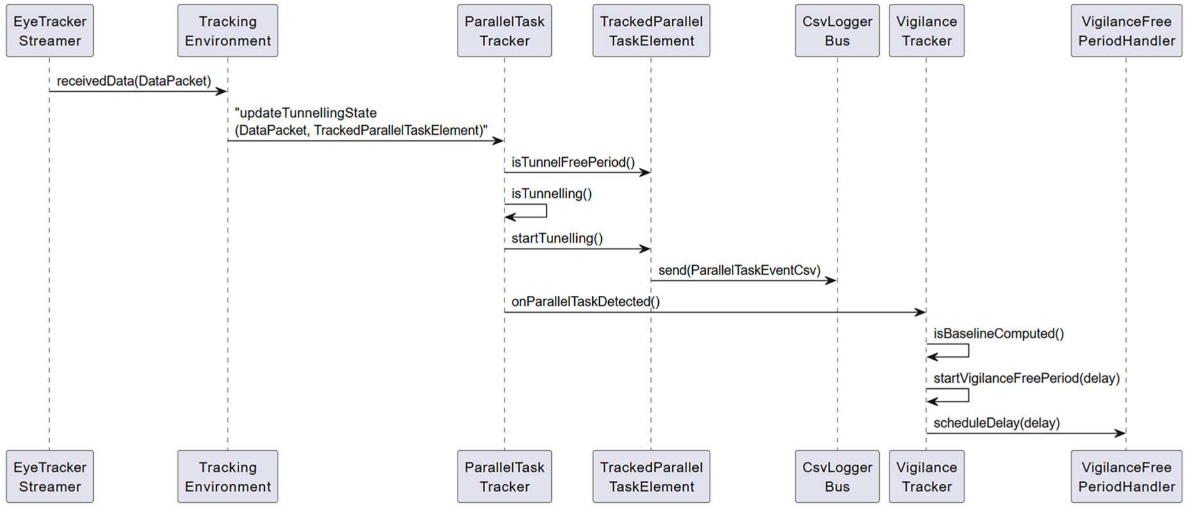

Figure A5. Sequence diagram for the detection of a parallel task.

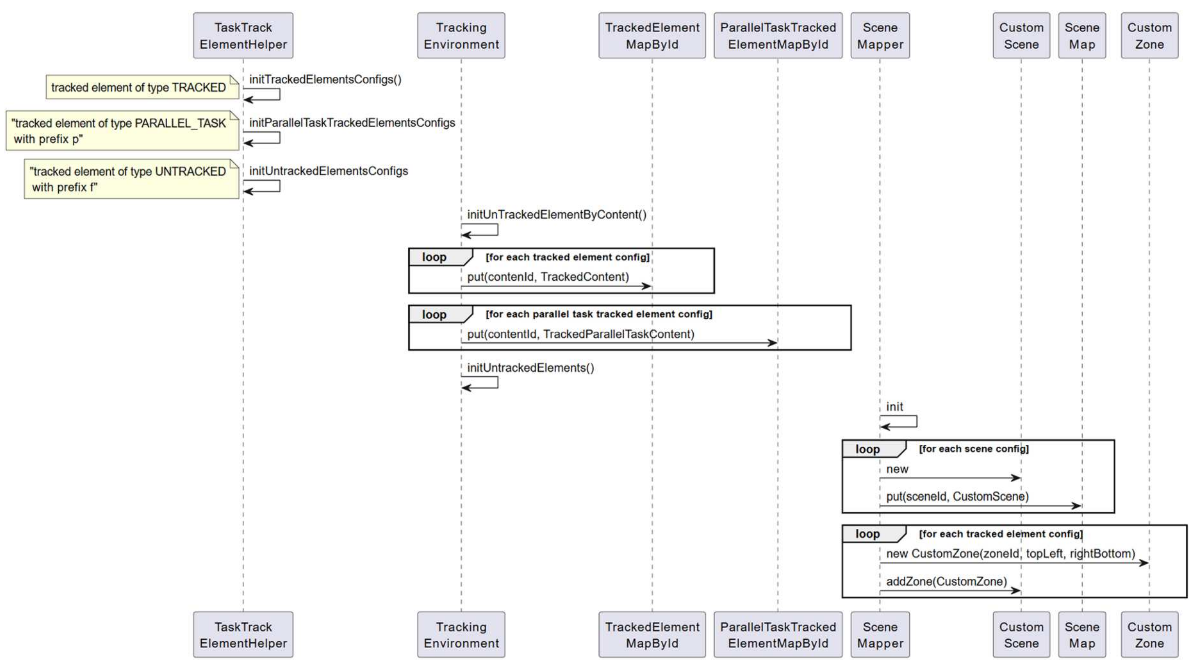

Figure A6. Sequence diagram for the tracking of the predefined elements of the environment.

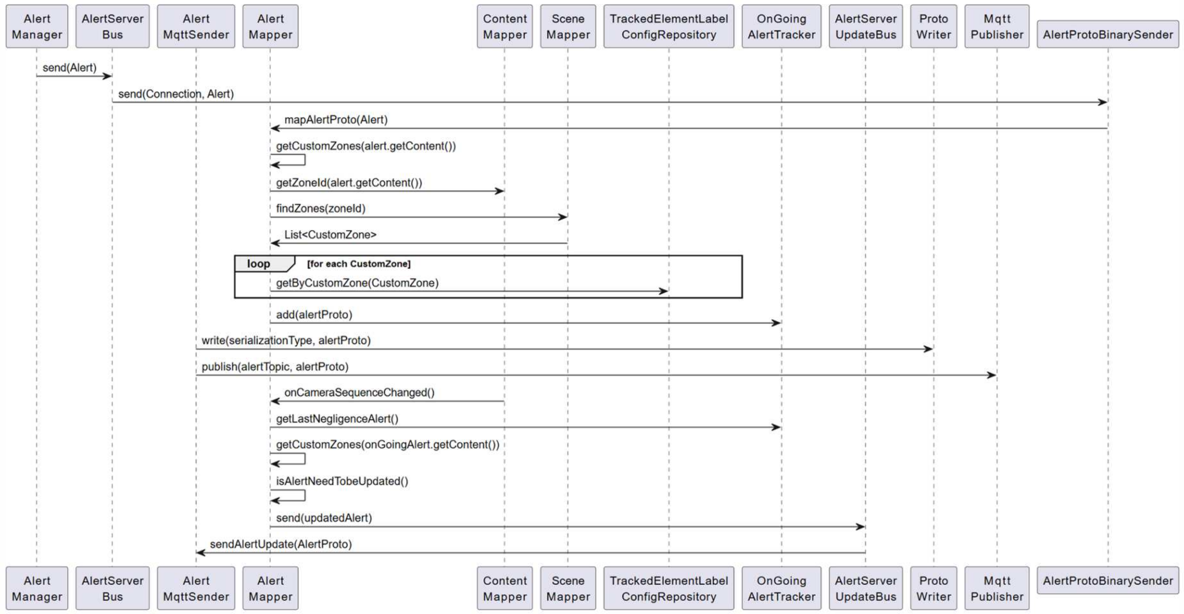

Figure A7. Sequence diagram for MQTT workflow of the notifications protocol.