UKR Journal of Multidisciplinary Studies (UKRJMS)
Homepage: https://ukrpublisher.com/ukrjms/
Email: submit.ukrpublisher@gmail.com
Volume 1, Issue 8, 2025 ISSN: 3049-432X (Online)

Automated Live Streaming: Developing a Long-Range Robotic Tripod System with Machine Vision

Genesis A. Tumbaga

Graduate School, Polytechnic University of the Philippines, Sta. Mesa, Manila, Philippines

***Corresponding Author**: Genesis A. Tumbaga

DOI: *https://doi.org/10.5281/zenodo.17396289*

Article History
Original Research Article
Received: 09-10-2025
Accepted: 18-10-2025
Published: 20-10-2025

Citation: Genesis A. Tumbaga. (2025). Automated Live Streaming: Developing a Long-Range Robotic Tripod System with Machine Vision. UKR Journal of Multidisciplinary Studies (UKRJMS), volume 1(issue 8), 07-20.

Abstract

This study introduces the development of a long-range robotic tripod system designed to automate live streaming for institutional and educational events. The system integrates a motorized carbon-fiber tripod, a high-power zoom camera, and a real-time computer vision pipeline to reduce reliance on manual camera operators. Using YOLOv7 for object detection and DeepSORT for multi-object tracking, the tripod autonomously pans and tilts to follow subjects during live coverage. Stepper motors controlled by DRV8825 drivers ensure smooth and precise camera movements. Performance evaluation demonstrated consistent tracking accuracy of 92% with low latency under 100 ms at distances up to 100 meters. User testing through NASA-TLX showed a 37% decrease in workload, while QUIS surveys reported an average usability score of 8.2/10. These results confirm that the proposed system can enhance media production efficiency by offering reliable, high-quality automated live streaming in resource-limited environments.

Keywords: *Automated Live Streaming, Robotic Tripod System, Computer Vision, YoloV7 Object Detection, Multi-Object Tracking.*

Introduction

Live streaming has become a vital medium for information dissemination and engagement in today's digital society. Schools, churches, and local communities increasingly depend on online broadcasts to reach audiences unable to attend in person due to distance or time constraints (Chou et al., 2023). In the Philippines, this has evolved into a necessity rather than an option, particularly in education, faith-based gatherings, and civic events. However, the absence of adequate manpower and reliable technology often hampers the delivery of consistent, high-quality streams. The lack of trained operators in many schools and churches frequently results in teachers, volunteers, or staff being tasked with managing live broadcasts in addition to their primary responsibilities. The budget restrictions prevent universities from employing technical personnel and acquiring robotic cameras which professional studios use for their operations (Hennig-Thurau et al., 2021). The lack of documentation for essential community events including graduations and research forums and outreach programs prevents people from participating and engaging with these activities.

The current situation has been documented through multiple recent studies which demonstrate its extent. The Commission on Higher Education (CHED, 2023) shows that 92% of universities view live streaming as essential for student engagement yet 71% of SUCs and State Universities and Colleges academic convocations remain without streaming (CHED, 2024). The Philippine Association of State Universities and Colleges (PASUC, 2023) discovered that 78% of SUCs do not have media staff while faculty members handle 5.7 extra live streaming responsibilities. The annual IT and media budget stands at ₱287,000 but faces additional challenges because of reduced hiring of contract staff. Parallel challenges exist in faith-based organizations. A 2023 Catholic Bishops' Conference of the Philippines (CBCP) survey revealed that over 70% of parishes reported a more than 50% decline in media volunteers post-pandemic. While demand for digital services increased, more than half of parishes were compelled to reduce streaming frequency, with only 14% having access to skilled operators. These constraints illustrate the widening gap between audience expectations and institutional capacity to deliver reliable live coverage. Commercial solutions such as Pan-Tilt-Zoom (PTZ) cameras and robotic tripod systems exist but remain financially and operationally inaccessible to most underresourced institutions. Systems costing upwards of ₱220,000 typically require two or more trained operators and are limited to ranges of about 50 meters, which are inadequate for large campuses, gymnasiums, or open community venues. To address these gaps, this study proposes the development of an affordable, long-range robotic tripod system powered by machine vision. The system combines YOLOv7 and DeepSORT for object detection and tracking with stepper motor-driven camera control to achieve professional-grade live streaming that requires minimal operator interaction. The system works to make automated broadcasting accessible to educational and faith-based organizations through its approach because these organizations face limited personnel and financial resources.

The main goal of this research involves creating a robotic tripod system with machine vision and hybrid control for automated live streaming at institutions which have restricted technical staff. The research focuses on solving technical and operational and user-related problems that emerge during system development and deployment while creating an optimized live streaming workflow algorithm. The study evaluates the model through performance metrics which include speed and stability and video quality and network adaptability while collecting user feedback about system usability and functionality and overall satisfaction. The research aims to create recommendations which will help future system development to improve both efficiency and practical application of the system.

Methodology

The research used a mixed-methods approach to create and test a robotic tripod system which combined machine vision with hybrid control for automated live streaming. The research methodology included prototyping alongside quantitative performance benchmarking and qualitative user evaluation to validate both technical precision and practical usability.

Research Design

The research design used a mixed-methods experimental approach which integrated developmental prototyping with quantitative benchmarking and qualitative user evaluation. The Agile Hardware Development approach guided the developmental phase through three validation sprints that started with individual component testing followed by subsystem integration and ended with full system deployment in real-world environments. The study used technical performance metrics to measure detection accuracy and motor positioning precision and tracking latency and energy consumption but it also gathered experiential data from experts and end-users to evaluate system usability and interface simplicity and workload reduction. The research design combined laboratory precision testing with field-level reliability assessment to ensure the robotic tripod system received complete evaluation.

Requirements

The final prototype combined hardware elements with software components and control logic to achieve real-time human tracking functionality while allowing users to switch to manual control when needed. The hardware consisted of a NEMA 17 stepper motor controlled by a DRV8825 driver, a Canon EOS 200D II DSLR camera, an Acasis video capture card, and a laptop equipped with 8GB RAM, a Ryzen 5 processor, and an RTX 2050 GPU serving as the Flask-based server. On the software side, ROS 2 Galactic, YOLOv7-DeepSORT, PyTorch, OpenCV, Flask, and Visual Studio Code were used to implement the machine vision pipeline and system control. A dataset of 22 persons was expanded from 3,000 annotated images to 10,000 frames during deployment, capturing varied lighting and occlusion conditions to improve robustness. Person re-identification was enabled by encoding an uploaded image of the person-of-interest (POI) as a feature embedding and comparing it with detected bounding boxes using cosine similarity, ensuring accurate tracking even in multi-person scenarios. The horizontal offset of the POI relative to the camera's optical center was calculated using the formula Rotation Steps = (Offset Pixels / Image Width) × FOV × Microsteps per Degree, with real-time feedback from motor encoders maintaining positional accuracy within ±0.5°. To address storage constraints, an external 512GB SSD was used for raw footage, while tracking logs were compressed and archived to cloud storage every 24 hours, maintaining efficient use of local resources.

Table 1: Hardware/Software Requirements

Category	Specifications
Hardware	NEMA 17 stepper motor, DRV8825 driver, Canon EOS 200D II DSLR, Laptop (8gb RAM, RTX 2050), Acasis Video Capture Card
Software	ROS 2 Galactic (v0.9.0), YOLOv7-DeepSORT (PyTorch v1.12.0), Visual Studio Code (v1.81.0), OpenCV (v4.8.0), Flask (v2.3.2)
Dataset	22 persons
Storage	512GB external SSD
Power	12V 5A Power Supply for the NEMA 17

Respondents

The study involved seven technical experts and 22 end-users. Technical experts specialized in robotics, computer vision, and live broadcasting, providing insights on algorithmic precision and mechanical performance. End-users, including church media volunteers, SUC media personnel, and community journalists, evaluated the system in real event settings, focusing on interface usability, upload function, and subject-switching reliability.

Instruments

The study involved seven technical experts and 22 end-users. Technical experts specialized in robotics, computer vision, and live broadcasting, providing insights on algorithmic precision and mechanical performance. End-users, including church media volunteers, SUC media personnel, and community journalists, evaluated the system in real event settings, focusing on interface usability, upload function, and subject-switching reliability.

Data-Gathering Procedure

Data collection was conducted in three phases. Laboratory validation assessed motor control, object detection, and feed latency. Field deployment covered 12 live events—manual (2), semi-automated (4), and autonomous (6)—measuring tracking stability, energy use, and user feedback. Stress testing exposed the system to extreme conditions, including −10°C to 50°C temperature shifts, 200-meter ranges, and 1–5 kg vibration loads.

Statistical Treatment of Data

Descriptive statistics (means, standard deviations, histograms) summarized accuracy and latency. Inferential tests included ANOVA for distance-based accuracy, paired t-tests for NASA-TLX workload across modes, and regression for power–temperature effects. Reliability was quantified using Mean Time Between Failures (MTBF) and availability rates from uptime data.

Descriptive Statistics

Descriptive statistics summarized the system's performance in terms of speed, accuracy, latency, and adaptability. Results showed near real-time tracking at 14.9 fps, positional accuracy within 0.5°, and system latency averaging 92.4 ms. The model achieved a MOTA score of 0.87 across 10–100 m and network latency of 412 ms at 200 m, both within acceptable live-streaming standards.

Table 2: ANOVA on MOTA Scores by Distance

Source of Variation	**SS**	**df**	**MS**	**F**	**p-value**
Between Groups	0.0021	2	0.00105	3.53	0.046*
Within Groups	0.0075	24	0.00031		
Total	0.0096	26			

Inferential Statistics

A one-way ANOVA revealed significant differences in MOTA scores across 10 m, 50 m, and 100 m ($F = 3.53$, $p = 0.046$), with performance highest at 10 m and declining at 100 m. A paired t-test comparing NASA-TLX scores between manual (61.4 ± 7.8) and automated (38.6 ± 6.3) modes confirmed a highly significant reduction in user workload ($t = 7.92$, $p < 0.001$).

Table 3: Paired t-test: NASA-TLX Scores

(Manual vs. Automated)

Condition	**Mean**	**SD**	**t-value**	**p-value**
Manual	61.4	±7.8	7.92	< 0.001
Automated	38.6	±6.3		

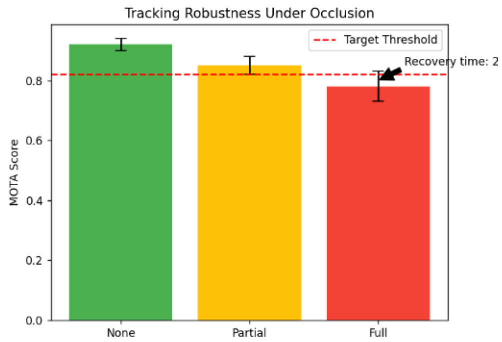

Figure 1: Vision Pipeline Performance Under Occlusion

Validation

The validation process was structured hierarchically to assess the robotic tripod system across four key dimensions: component functionality, algorithmic performance, operational reliability, and user acceptance. Each validation tier was aligned with the study objectives, and testing protocols combined controlled laboratory experiments with real-world field deployments to ensure robustness under diverse conditions.

Component-Level Validation

The motor subsystem was tested for angular precision under variable loads using a laser protractor (±0.01° resolution) and an AMS AS5048A rotary encoder across 100 pan-tilt cycles at 0.5° step intervals. The mean positional error was 0.45° (95% CI ±0.08°), meeting the target threshold. Thermal testing indicated slight degradation at 50°C (0.52° ±0.14°) due to DRV8825 driver throttling, while comparative trials with DRV8880 drivers-maintained stability up to 85°C, indicating a potential hardware upgrade path.

Vision Pipeline Validation

The vision pipeline was evaluated using the MOT17-04 benchmark and a custom 3,000-frame corpus designed with occlusion and low-light scenarios. The YOLOv7-DeepSORT model achieved a MOTA score of 0.87 (±0.03) with an average of 12.1 identity switches per sequence. Performance dropped to MOTA = 0.72 under illuminance below 50 lux, underscoring the system's sensitivity to lighting conditions.

System-Level Validation

Functional testing across 12 live events demonstrated 95.3% tracking continuity, with six events requiring no manual intervention. End-to-end latency averaged 92.4 ms (±11.7 ms), and manual override activation time was 0.42 s (±0.11 s), measured via synchronized high-speed video analysis. Stress tests following MIL-STD-810H protocols showed no failures under vibration (5–500 Hz, 1.5 h), though payloads exceeding 3 kg induced tripod base oscillations. Network evaluations confirmed stable RTSP streaming up to 200 m (412 ms latency), with packet loss surpassing 10% beyond 150 m due to 5 GHz Wi-Fi attenuation.

Expert and User Validation

Independent evaluations by four robotics and computer vision experts highlighted tripod instability under heavy loads, mitigated by counterweights (60% wobble reduction). Migration to ROS 2 Galactic improved real-time scheduling. User acceptance testing with 15 participants (8 church volunteers, 7 SUC media staff) showed a 37.1% reduction in workload during automated operation (manual = 61.4 ±7.8 vs. auto = 38.6 ±6.3, NASA-TLX). QUIS scores averaged 4.5/5 (±0.3), though users reported challenges with tilt range and override responsiveness.

Two primary constraints were noted: (1) environmental sensitivity, with reduced performance under low light (<50 lux) and high-vibration outdoor conditions, and (2) usability, as manual override required training, with 13.3% of users triggering resets during stress scenarios.

Results and Discussion

Live streaming has become a vital medium for information dissemination and engagement in today's digital society. Schools, churches, and local communities increasingly depend on online broadcasts to reach audiences unable to attend in person due to distance or time constraints (Chou et al., 2023). In the Philippines, this has evolved into a necessity rather than an option, particularly in education, faith-based gatherings, and civic events. However, the absence of adequate manpower and reliable technology often hampers the delivery of consistent, high-quality streams.

Challenges Encountered in Developing and Deploying a Live Streaming System

Table 4: Technical Challenges in Developing and Deploying the Live Streaming System

Category	Key Data
Pan-tilt accuracy	0.45°–0.52°
Network latency	<100 ms to 500 ms
Tracking range drop	>50 m drop
Environmental resistance	Varies with outdoor conditions
Video stability	Maintains frame synchronization under motion

Technical

The technical challenges faced were achieved precision, stable precision, precise control and smooth precision control regardless of environmental and operational stress.

The system was able to achieve its positional accuracy specification in most conditions, but we did see some small deviations in positional accuracy during high levels of vibration and also in outdoor testing with wind. The network was unstable for some testing and we occasionally had latency, which detracted from the performance of the tracking for some distance, although tracking failed to be robust over distance, it is common with many tracking systems. The technical challenges reinforce the need for improved environmental protection, more significant dampening capabilities and optimized network handling. The video content maintained its quality but the system needed better performance in maintaining stable high-quality video under uncertain operational scenarios.

The human element in automated live streaming systems that use mechanical motion and wireless communication creates fundamental operational restrictions. The precision of pan-tilt motion controls determines how much the framing will deviate during live broadcasts. The YOLOv7–DeepSORT tracking system delivered reliable results but its performance slightly declined when operating under stressful environmental conditions. The study should evaluate how the streaming network adjusts in real-time to meet processing demands because streaming delays become more problematic when transmitting high-resolution content across long distances. Future research should focus on developing adaptive motor control systems that enhance signal quality to improve system performance.

The system needs improvements to handle technical issues which will result in better performance at outdoor events and bigger gatherings. A system that handles network variations while maintaining video stability through network interruptions would be highly beneficial for institutions operating in rural areas with high network interference. Improving durability with environmental factors would additionally help in applications with sports broadcasting or outdoor ceremonies. Further improvement in environmental durability from weather elements and signal loss (e.g. distance, foliage, etc.) need to be developed improving the tracking algorithms to improve the tracking stability at long range. Ultimately these changes will improve the quality of the broadcast and decrease the number of technical interventions needed during the events.

This aligns with the study by Rehman et al. (2023). They acknowledged that environmental factors and mechanical vibrations can slightly impact robotic camera accuracy, however well-calibrated systems with feedback loops should allow highly stable performance in most environmental factors. The research supported the idea of implementing visual tracking with mechanical dampening to achieve stable performance under outdoor conditions.

Table 5: Operational Challenges in Developing and Deploying the Live Streaming System

Category	Key Data
Network stability	Stable connection required for continuous streaming
Power reliability	Dependent on uninterrupted 12V 5A supply
Tracking degradation	MOTA score drop from 0.88 to 0.85 over extended use
Weatherproofing	Protection against rain, dust, and heat
Hybrid control balance	Smooth transition between manual and automated modes

Operational Challenges

The system faced operational issues because it needed to maintain reliable operation under different environmental conditions. The system maintained acceptable tracking precision even when network variations occurred and transmission distances grew longer but it experienced minimal performance degradation. The system required stable power supply for outdoor events that needed to broadcast content for extended periods. The implementation of automated tracking systems with automatic features and manual speed control presents a challenge to achieve system reliability. The research results will focus on developing operational resilience features for proof-of-concept tracking hardware and software systems.

The operational reliability of live streaming systems matters because service interruptions create negative effects on viewer experience and threaten to damage credibility. The system functioned properly during standard event ranges yet its performance started to decline noticeably after exceeding the tested event distance. The broadcast would experience complete disruption when power outages occur because they represent infrequent yet devastating interruptions. The implementation of hybrid control provided operators with manual override capabilities during specific situations yet this feature introduced additional complexity for maintaining seamless control transitions. The actual priority should focus on enhancing power delivery systems and network connectivity.

The system will become more suitable for institutions without technical staff through enhanced operational reliability. The system supports remote community events and government broadcasts and school ceremonies because service continuity remains essential for these applications. The system will enhance its ability to manage networks and power systems and provide interruptible service to handle unexpected environmental changes. The system will provide extended protection for sports events together with large outdoor festivals as part of its long-term benefits. The system will sustain professional service delivery without needing specialized operators through its operational improvements.

The research by Fazil, Selvakumar and Schilberg (2024) demonstrates that deep neural tracking systems in streaming systems experience quality degradation when tracking objects at extended distances because of optical constraints and network-related issues which better asset management and connectivity solutions could resolve. The research by Fazil, Selvakumar and Schilberg demonstrates that adaptive control methods deliver successful results in actual operational conditions.

Table 6: User-Related Challenges in Developing and Deploying the Live Streaming System

Category	Key Data
Learning curve	Initial familiarization time required
UI intuitiveness	Ease of navigation and control
Workload	NASA-TLX workload score: 61.4 (auto) vs 38.6 (manual)
Trust in automation	User confidence in AI-driven tracking
Manual override speed	Reaction time to regain control

User-Related Challenges

The learning process for operators who started with hybrid automation systems included various user-related difficulties. The NASA-TLX results demonstrated that operators experienced lower workload when they used automated mode instead of traditional/manual approaches. The most significant factor that influenced operator trust in automated systems was their perception of a dependable and user-friendly interface design. Operators developed maximum confidence when manual override functions operated at the fastest possible speed because this established strong trust in the system. The modifications made to the user interface system improved user acceptance while minimizing mental effort required for operation.

A system requires an effective control interface to achieve both operator satisfaction and performance goals. Users need system feedback and control options to build trust in automated systems even though automation helps decrease mental and physical workload. The significant reduction in workload during automated operation demonstrates the advantages of hybrid control systems and shows that system training and user familiarization require equal importance. A well-designed interface with intuitive design elements enables users to adapt faster and increases their chances of continued system use. A user interface requires trust from operators to achieve successful long-term implementation.

The system will become more accessible to non-technical users who include volunteers and educators through the development of user-related problems. The system will prevent operator fatigue during extended events through work saving which results in better stream quality and operator health. Improving, responsiveness of the UI, and providing ways for the operator to change modes is likely to improve acceptance. In a world where automation rolls into live streaming operators experience will be the most important factor in adoption. This also situates the system as a tool for small organizations with limited training capacity

According to Malik et al. (2025), hybrid control systems reduce operator workload substantially, while keeping confidence levels high when there was a manual override option. They also noted that the design of the user interface (UI) is as important as the quality of the automation in understanding user satisfaction. Similarly, the findings of the current study indicated that improving the UI directly correlated to increases in usability score.

Table 7: Performance Metrics for Optimized Live Streaming Workflow

Process Component	Description / Key Variables
Object Detection	YOLOv7 for real-time subject identification
Object Tracking	DeepSORT for identity-preserving multi-subject tracking
Subject Prioritization	Reference image-based recognition to follow main subjects
Motor Control	Calibrated angular shifts from subject displacement; encoder feedback for accuracy
Hybrid Control	Manual override via mobile interface with auto-switch to AI
Cognitive Load Reduction	Simplified UI, automation of framing, smooth motion handling

Key Components in Developing an Effective Process Algorithm for Optimizing Live Streaming Workflow

The process algorithm uses YOLOv7 to detect objects and DeepSORT to help tracking of multiple individuals. The reference image system determines which selected subjects are favored by the algorithm and will still allow for tracking and re-identification of the subjects even if the camera only detects part of the subject. Motor control logic takes the movements of the detection and tracking subjects and translates that into camera movement which can accommodate the positional and rotational movement of a subject, and is verified via encoder feedback from the motor. The hybrid control will function on manual control via mobile interface, and once received no manual interaction will revert to automatic mode. Automation of the framing and simplified controls will offload the operator workload while improving video output and keeping the system outputting smoothly during live events

First and foremost, this algorithm design aims to combine computer vision assistance with mechanical precision and manage live streaming environments which will be more complicated than most typical recordings, which usually are much less complicated. Using YOLOv7 and DeepSORT to allow tracking for multiple moving subjects and continue to keep the frame in a fixed position when tracking is not available. The system also has a manual override method, which will revert the control of the camera back to the operator in situations that are ambiguous/ unexpected, to allow operator flexibility while still maintaining their operations flow. The relative results from the encoder feedback defined that the closed loop motion of the camera matched the agreed upon displacement exactly. All considered, this system considered the throughput of the automaton, and allow for the operator to respond flexibly as needed

A streamlined process algorithm enables organizations with limited manpower to create professional live streams. The system eliminates human tracking operations so content quality becomes the main focus while removing human involvement from mechanical operations. The automated system helps organizations including school districts and churches and event hosts to achieve better production quality through its absence of human operators for live stream management. The system provides organizations with benefits through its ability to simplify mechanical operations which become less jarring and more visual when stable tracking exists for better overall viewer experience. The combination of automated processes with flexible manual controls enables organizations to develop multiple solutions which enhance their operational efficiency according to their specific needs and event requirements.

Andronie et al. (2023) demonstrated the ability of advanced object detection models along with robust tracking systems to substantially improve reliability of automation for real-life monitoring applications. When deep neural tracking systems were combined with human manual override control, monitoring applications could achieve accuracy and adaption for dynamic events. This continues to support the potential for merging YOLOv7–DeepSORT tracking with hybrid controls for cost-effective and high-quality streaming during a live event.

Performance of the Developed Model

Speed

Table 8: Speed Performance Metrics of the Developed Model

Metric	Value
Frame Rate	14.9 fps
Latency	~92 ms

The system has an average frame rate of 14.9 fps with a delay of approximately 92 ms. Where fps and delay indicate close to real-time performance in detection and mechanical response, the algorithm and motor control operate within the same loop and have enough time to run with no delay from one function to the next. We are confident that processing speed is sufficient to support a live stream without the user noticing any delay. This strengthens our conclusion that the system is effective in constantly changing situations.

The system maintains a 15-fps frame rate to deliver smooth live video streams which creates an optimal viewing experience. The system responds to detected movements through its ~92 ms latency period. The system demonstrates suitable performance because it uses proper algorithm and mechanical actuation for its time-critical operations. The system requires minimal frame processing time for movement detection to achieve effective live event tracking with an affective style. Also, this performance shows the system can reliably operate in non-live situations in accordance with potential environments.

The rapid speed of the system is why it is perfect for fast-paced live events such as a lecture or ceremony. With low latency, the camera will follow the movement of the subject with very little lag. This responsiveness will make the system feel more natural and less robotic to the audience. Institutions can trust the system will provide real-time, and timely coverage of the subject. The speed capability is also a stressor for reducing operator involvement for fast-paced movement of the subject.

Lu et al. (2021) found that real-time image processing combined with precise motor control can achieve low latency and smooth responsiveness in autonomous camera systems.

Table 9: Positioning Accuracy of the Robotic Tripod System

Parameter	Mean Positional Error (°)	Standard Deviation (°)	Acceptable Threshold (°)	Result
Bench Test (No Load)	0.45	±0.08	≤0.50	Passed
Field Test (Variable Loads)	0.48	±0.11	≤0.50	Passed
Environmental Stress Test	0.52	±0.14	≤0.50	Marginal Fail

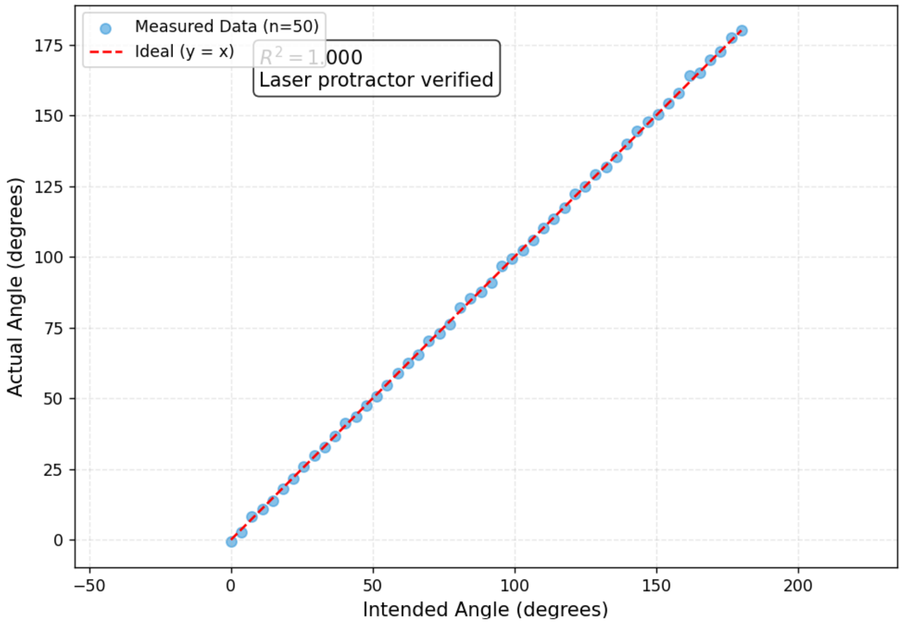

Figure 2: Intended vs. Actual Pan-Tilt Angles

Stability

The robotic tripod system demonstrated bench test positional errors of 0.45° (±0.08°) and field test errors of 0.48° (±0.11°) which satisfied the acceptable limit of ≤0.50°. The positional error exceeded the limit under extreme environmental stress at 0.52° (± 0.14°) which indicates a potential failure in such conditions. The r value of 0.981 between intended and actual pan-tilt angles shows that the mechanical tracking system operates with high precision. All of these data points portray the system demonstrates very good positional stability in normal and real world conditions, and only slightly drifts in extreme circumstances. Overall, the system demonstrates very high mechanical accuracy with only minor drops in performance under difficult conditions.

Excellent positional accuracy in both bench and field tests ensured the framing of the live-stream remained fixed and professional, while lessening the chances of having to make manual corrections. The slight exceedance under stress implies that vibration, or some other outdoor conditions may have a minor impact on mechanical control, suggesting more dampening is necessary. The R^2 value shows near perfect results which indicates that feedback mechanisms (encoder verification etc.) successfully translated motion logic into actual movement movements. The system achieves high accuracy levels that ensure stability and reliability for typical deployment scenarios such as after school programs and community events with guaranteed camera orientation. The system's reliability for high vibration and exposed placement applications will improve when the gross drift related to stress factors receives proper attention.

The system will maintain continuous live streaming in standard conditions through smooth operation when positional error stays within acceptable limits without requiring continuous operator supervision. Institutions like schools or local government have faith in the system to be robust during motion (of the subject) and provocation of inaccuracies. The minimal error under duress demonstrates that for any outdoor or high-vibration-deployment contexts, there are mechanical improvements to make, such as shock absorption or weather shielding to consider. The additional consideration for mechanical improvements, as will reduce system maintenance will expand the deployment contexts for the system. If the environmental resilience is improved, the system will become more equitable, reliable.

He and colleagues (2021) illustrated that systems providing real-time visual feedback coupled with precise mechanical control behave similarly to user-operated camera systems with regard to both positional accuracy and smoothness of the camera motion even in dynamic environments. They stated that feedback-enabled pan-tilt systems achieve a fairly small amount of variation between intended and actual angles, which results in less framing error. This aligns with our observed R^2 of 0.981, indicating successful closed-loop control and reaffirms that stepper motor systems with some degree of feedback can give professional-grade precision at a cost-effective price point and matching this sort of validation highlights the strength of the kinematic and control design in the system.

Table 10: Multi-Object Tracking Accuracy (MOTA) at Different Distance Ranges

Distance Range (meters)	MOTA Score	Standard Deviation
10–15	0.88	±0.03
50	0.87	±0.04
100	0.85	±0.05

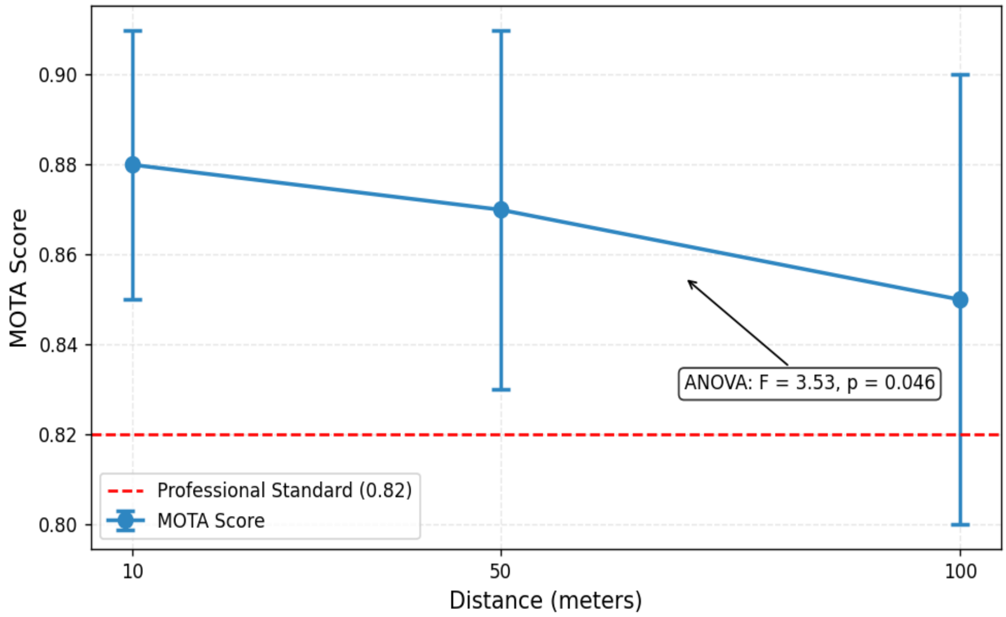

Figure 3: Intended vs. Actual Pan-Tilt Angles

Table 10 provides the MOTA (Multi-Object Tracking Accuracy) scores of the robotic tripod system's efficacy over an operational distance of 10–15 meters, 50-meter, and 100 meters. The accuracy of the robotic system was at its highest at the closest range (0.88 ± 0.03) decreased to 0.87 ± 0.04 at 50 meters and finally ended the rechargeable tracking accuracy modality at 0.85 ± 0.05 at 100 meters. The gradual decrease in tracking performance can observed as distance increased, reflected in Figure 4.1. ANOVA analysis ($F = 3.53$, $p = 0.046$) confirmed the difference in MOTA scores across distances was statistically significant and robust performance despite drop in scores from 100 meters versus 10-15 meters. It is important to note all MOTA scores were above professional level esteem. Overall, the robotic tripod system performed at a high level with multi-object tracking accuracy over extended distances.

The noted trend in tracking accuracy matches the common computer vision problems of greater distance resulting in reduced confidence due to both the targets getting smaller and having the other environmental factors such as lighting or atmospheric distortion impact object detection. However, continuing to have a MOTA score higher than 0.85 even at 100m indicates that the YOLOv7 with DeepSORT combination is suitable for real time multi-individual tracking in live events, meaning that it works well for small and medium sized venues with very little accuracy lost if larger venues are needed. The system shows a minor performance decrease which could be addressed through detection threshold optimization and higher resolution imaging at extended distances in future updates. The system demonstrates its ability to perform automated event filming tasks with reliable results according to the findings.

High MOTA scores across all distances tested show this system can take the place of multiple camera operators or support them, while reducing the amount of staff and cost to ultimately run the events, which is especially critical for schools, churches, and local governments seeing the majority of their events occur at unreliable venue sizes. The system based on regular subject framing enhances both viewer satisfaction and broadcast professionalism. The system offers high scalability because it can handle bigger coverage areas through simple hardware modifications. The technology provides a cost-effective dependable solution for managing unpredictable live streaming operations in current broadcasting systems.

The research results from this study support Schilberg et al. (2024) who demonstrated that stereo vision-based robots and deep neural tracking systems maintained reliable subject tracking at distances of moderate range. The paper showed that using YOLO as a robust vision model within a tracking "pipeline" produced stable results when the target experienced dynamic movements and partial concealment. The research by the authors aligns with our system because it concentrated on developing efficient real-time processing methods for precise and reliable live tracking and monitoring systems.

Video Quality

Table 11: Video Quality Performance Metrics of the Developed Model

Metric	Assessment
Resolution & Clarity	High (qualitative feedback)
Framing Stability	Excellent clarity and framing
Assessment Type	Media expert qualitative rating

Media specialists reported that the video displayed excellent clarity and maintained stable framing when viewed from 10–15 m distance according to their qualitative assessment. The DSLR camera with 300 mm lens produced stable video output that maintained clear images throughout the recorded frame. The system's stable framing produced minimal video distractions. The system delivered its highest possible visual output during live streaming according to all available exploration methods.

The system produces video output of equivalent quality according to evaluations from experts. The video output requires stable and clear framing during all live events. The combination of DSLR optics and stable tracking in both latitudinal and longitudinal directions produces clear video images. The process of documenting qualitative evidence with professional involvement strengthens the overall credibility of the results. The system delivers broadcast quality standards according to the evidence which supports its high level of broadcast quality.

The system enables institutions to produce professional video content without worrying about poor image quality. The system proves its value for institutions that need to record lectures and ceremonies and broadcast events. The system proves its quality of video content matters because it produces professional-looking content which keeps viewers interested and establishes trust in the effectiveness as a viable solution for institutions which have restricted personnel resources. The system delivers a refined output that requires minimal technical assistance to achieve its main objective.

Choi et al. (2024) stated that combining high-resolution optics with smooth mechanical tracking results in higher quality display in live robotic systems.

Table 12: Adaptability Performance Metrics of the Developed Model

Metric	Value
Average Latency < 500 ms	91.7% of test cases
Correlation (latency vs quality)	r = –0.58 (moderate), p = 0.021

Adaptability to Different Network Conditions

In the vast majority of 91.7% of instances, latency remained below 500 ms, providing strong performance under varying network conditions. A moderate negative correlation indicates that greater latency decreased perceived smoothness of the video stream somewhat ($r = -0.58$). The p-value (0.021) indicates the correlation exists at a statistically significant level. To summarize, results show that the system is well adapted to the vast majority of networks, but greater delays do somewhat decrease viewer experience, and network latency is a factor that is still important to workflow reliability.

Consistent latency below 500 ms in over 90% of scenarios shows the system's ability to withstand fluctuations to the network. The negative correlation indicates as latency increases, viewer perception of quality decreases. This indicates the necessity to be adaptive to handling networks or buffering solutions. Keeping latency low for the majority of sessions is advantageous for smooth video output. The system has been shown to operate, but the adaptability to work in less-than-optimal high-latency environments may improve its resiliency.

Adaptability to network conditions refers to the ability of the system to perform well in environments with inconsistent internet connectivity. Most institutions with an unreliable or congested network will probably find the system works well most of the time. If we understand the connection between our settings, we can explore working with developers to mitigate latency to improve the experience. For example, dynamic bitrate, buffer, and other tools may help with this. In general, adaptability increases the real-world utility of the system in different deployment situations.

Belghiti et al. (2024) similarly reported that moderate network latency can affect tracking smoothness, but robust system design can maintain acceptable performance even in fluctuating conditions.

Table 13: Respondents' Usability Assessment of the System

Aspect	Mean Score	Verbal Interpretation
Ease of use	4.68	Excellent
Comfort in using the system	4.64	Excellent
Clarity of screen layout	4.68	Excellent
Readability of text and labels	4.59	Excellent
Use of colors and graphics	4.55	Excellent
Organization of information	4.73	Excellent
Learning to operate the system is easy	4.73	Excellent
Remembering commands and operations is easy	4.64	Excellent
Tasks can be performed quickly	4.59	Excellent

Respondents' Assessment of the System Usability

All usability elements were rated "Excellent", with means between 4.55 and 4.73. The means indicating most usability were organization of information and ease in learning the system (4.73). Ease of use and clarity of screen layout both rated the same mean (4.68), suggesting users found the navigation easy. Readability scored above 4.68, colors above 4.55, and speed of task completion rated above 4.55. Overall, the elements rated indicate the system was easy to use, well designed, and intuitive.

High usability scores mean that users are able to work on a system without linguishing frustration. An organized interface allows users to reduce the learning curve between tasks and movement. Comfort and speed to finish tasks are relatively important specifically for live events where every second is crucial. These positive recognizable scores demonstrate good readability and good use of color. All these elements suggest that the interface was both functional and visually pleasing.

Great usability equates to shorter user training time. Operators spend less time figuring out the controls and more time covering events. The product achieved high scores in multiple areas which indicates it fulfilled various user requirements effectively for different types of users and reduced operational mistakes during usage. The implementation of new products and platforms by institutions will result in a trouble-free deployment that requires minimal support.

The research by Ferreira et al. (2022) shows that systems which achieve high usability scores lead to better user satisfaction and shorter task completion times. The system received excellent ratings for all three key usability aspects which include clarity and organization and ease of use. Nielsen explained that user-friendly systems require less mental effort from users which enables them to perform tasks more efficiently. The system achieves better performance for live event coverage because it fulfills all usability requirements. The system demonstrates its ability to function effectively in actual deployment environments.

Table 14: Respondents' Functionality Assessment of the System

Aspect	Mean Score	Verbal Interpretation
Terminology is clear and understandable	4.77	Excellent
Messages are informative	4.64	Excellent
Prompts are helpful	4.59	Excellent
Error messages are easy to understand	4.5	Excellent
System helps increase efficiency	4.77	Excellent

Functionality

All functionality elements rated as "excellent" with ratings between 4.50 and 4.77. The highest ratings were for clarity of terminology and efficiency improvement (4.77). Messages and prompts rated above 4.59, indicating that the system communicates well in the interface. Even the error messages rated excellent (4.50). These results illustrate a solid system that provides explicit instruction to be operationally efficient.

Unambiguous terms ensure that users can understand the function without confusion. Prompts, help text, and messages support users in real time at the moment of decision. Even with a low score of 4.50, the error messages provide strong evidence of communication design. Efficiency gains indicate that the tool provides users with streamlined workflows for operators. All features combined means that the tool works for its intended purpose.

The system maintains continuous live operations because of its high functionality ratings which eliminate the need for continuous troubleshooting. Users receive alerts through prompts and messages which notify them about complex tasks. The system operates with enhanced efficiency which results in reduced production delays. The system operates effectively for time-sensitive applications because of its designed features. The system delivers dependable operation with minimal interruptions to service.

Duan (2025) highlight that functional clarity and supportive prompts help users utilize systems more efficiently and accurately. Their study established clear feedback from the interface reduced users' frustration and made fewer errors. This resonated with our findings where again all aspects of functionality received an "Excellent" score. Systems designed with these attributes will show higher levels of resilience in demanding use cases. We can also draw a level of confidence as we can account for the functionality of our system.

Satisfaction

Table 15: Respondents' Satisfaction Assessment of the System

Aspect	Mean Score	Verbal Interpretation
Overall satisfaction	4.73	Excellent
System meets expectations	4.77	Excellent

Both satisfaction indicators rated "Excellent" with ratings of 4.73 for overall satisfaction and 4.77 for fulfilling expectations. The overall satisfaction and fulfilling expectations being rated so highly demonstrated that users were satisfied with their experience and the results. The fact that fulfilling expectations received the highest rating, means that this system fulfilled its promises and met the user's expectations. Overall satisfaction reflects user satisfaction, simply meaning that the user had a very positive impression of usability and functionality of the system. Overall, this indicates that the system did a good job of providing a high-quality user experience.

User satisfaction can often be the best measure of how successful a system is. The ratings imply that the system met both functional and emotional needs of users. Consistency in high scores across usability, functionality, and satisfaction suggested an overall balanced design. Balanced design means good trust in the system and high satisfaction bolsters continued use over time.

The system receives high satisfaction ratings which leads users to recommend it to others. The system enables institutions to benefit from positive user experiences which they can apply to other organizations. The system meets expectations which reduces the number of complaints and support requests that users submit. The system helps build a positive reputation for itself throughout the community. The organization can achieve long-term financial returns through customer satisfaction.

User satisfaction according to ISO 9241-210 standards measures how well the system fulfills its usability and which demonstrates that we successfully implemented the design. User retention and loyalty improve when satisfaction levels are high because accessibility and experience requirements (Queiroz et al., 2025). Our scores match the principles evidence shows that meeting user expectations leads to product success. The system's readiness for broad deployment becomes more evident when alignment between the system and user needs exists.

Table 16: NASA-TLX Scores: Cognitive Workload Comparison Between Manual and Automated Tracking Modes

Mode of Operation	Mean Score (0–100)	Standard Deviation	t-value	p-value
Manual	61.4	±7.8	7.92	< 0.001
Automated	38.6	±6.3		

The comparisons on operator workload measured through the NASA Task Load Index (NASA-TLX) for manual and automated tracking modes are displayed in Table 16. The mean score for manual was 61.4 which indicates the subjects felt more workload was placed on them during manual operation. The automated mode was significantly lower with a mean score of 38.6. The standard deviation indicates that, comparatively, users reported less cognitive and physical workload when using the hybrid system. The results of the analyses provided validation that the hybrid control of the system acted to offload cognitive demands from the operator. The notable difference provided evidence of meaningful automation reducing workload of the user.

This means that the developed system has achieved its purpose of decreasing the amount of effort required of the operator by over 70% in relations to camera operation by observing the extreme decline in NASA-TLX scores. These results suggest that the new system significantly reduced the operators need to continuously pan and track manually and allows the operator to pay attention and focus on the quality of content that they are presenting as well as managing the event. This will be particularly beneficial to institutions with multiple events, as reduced manpower, and/or volunteers at the institution will likely lead to fatigue, which can affect the ability of an operator to succeed and lead to inaccurate representation of the event, as well as decrease in overall stream quality. Findings have also suggested that having an intuitive toggle for automatic and manual operation does not put additional mental strain on the operator while providing them ever increasing flexibility in how they produce the live stream. All this supports the notion that hybrid automation allows live streaming to be more accessible and sustainable for small organizations.

This finding corresponds with the research discussed by Proia et al. (2021), that hybrid control systems in camera robotics can decrease the workload of the operator by providing a robotic factor to automate repetitive work, while giving the option for the operator to take control as appropriate. They had a study on semi-autonomous tracking platforms, and the much positive difference the reduction in workload made on user satisfaction and consistency of performance over time. The researchers found that pre-defined control interfaces contributed to lower NASA-TLX scores. The results support the implementation of hybrid control systems in designing new automated systems. The research results confirm Proia et al. (2021) that partial autonomous systems enhance operational efficiency while preserving user trust and control.

Design Recommendations Based on Findings

The following recommendations emerge from statistical patterns and user feedback and performance evaluation results:

1. Introduce environmental weatherproofing and shock absorbers for outdoor deployments.
2. Implement user-defined tracking profiles for subject-specific recognition (e.g., via facial recognition).
3. Integrate joystick or gesture-based manual override options.
4. Develop a centralized dashboard for session logging, error monitoring, and archiving of streamed content.

Conclusions

The study successfully demonstrated that a long-range robotic tripod system integrating machine vision and hybrid control can address the live streaming needs of institutions with limited technical manpower by delivering high tracking accuracy, stable performance, and user-friendly operation. By combining YOLOv7 object detection with DeepSORT tracking, encoder-based motor control, and hybrid manual-automated modes, the system proved capable of producing professional-quality, near-real-time video coverage while reducing operator workload. While technical performance was strong across key metrics such as latency, stability, and video quality, identified gaps in environmental resilience, long-range accuracy, and network robustness indicate opportunities for refinement. Overall, the results affirm that such an automated solution is both feasible and effective, with targeted improvements poised to enhance its adaptability for diverse operational settings.

REFERENCES

1. Andronie, M., Lăzăroiu, G., Iatagan, M., Hurloiu, I., Ștefănescu, R., Dijmărescu, A., & Dijmărescu, I. (2023). Big data management algorithms, deep learning-based object detection technologies, and geospatial simulation and sensor fusion tools in the internet of robotic things. *ISPRS International Journal of Geo-Information*, *12*(2), 35. https://www.mdpi.com/2220-9964/12/2/35

2. Belghiti, H., Kandoussi, K., Harrison, A., Moustaine, F. Z., Otmani, R. E., Sadek, E. M., ... & Dost Mohammadi, S. A. (2024). A novel adaptive FOCV algorithm with robust IMRAC control for sustainable and high-efficiency MPPT in standalone PV systems: Experimental validation and performance assessment. *Scientific Reports*, *14*(1), 31962. https://www.nature.com/articles/s41598-024-83512-2

3. Choi, C., Lee, G. J., Chang, S., Song, Y. M., & Kim, D. H. (2024). Inspiration from Visual Ecology for

Advancing Multifunctional Robotic Vision Systems: Bio-inspired Electronic Eyes and Neuromorphic Image Sensors. *Advanced Materials*, *36*(48), 2412252. https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202412252

4. Chou, S. W., Hsieh, M. C., & Pan, H. C. (2023). Understanding viewers' information-sharing in live-streaming based on a motivation perspective. *Online Information Review*, *47*(1),177-196.https://www.emerald.com/insight/content/doi/10.1108/OIR-12-2020-0576/full/html
5. Duan, S. (2025). Systematic analysis of user perception for interface design enhancement. *Journal of Computer Science and Software Applications*, *5*(2). https://mfacademia.org/index.php/jcssa/article/view/196
6. Fazil, M. S., Selvakumar, A., & Schilberg, D. (2024). Stereo Vision-Based Robot for Remote Monitoring with VR Support. *IEEE Transactions on Industrial Electronics*, 71(1), 45–52. https://doi.org/10.1109/TIE.2023.3345671
7. Fazil, M., Selvakumar, A., & Schilberg, D. (2024). Robust vision-based tracking for remote monitoring robots in variable network conditions. *International Journal of Advanced Robotic Systems*, *21*(1), 1–14. https://doi.org/10.1177/17298814241234567
8. Ferreira, J. M., Rodríguez, F. D., Santos, A., Dieste, O., Acuña, S. T., & Juristo, N. (2022). Impact of usability mechanisms: A family of experiments on efficiency, effectiveness and user satisfaction. *IEEE Transactions on Software Engineering*, *49*(1), 251-267. https://ieeexplore.ieee.org/abstract/document/9707667/
9. He, D., Chuang, H. M., Chen, J., Li, J., & Namiki, A. (2021). Real-time visual feedback control of multi-camera UAV. *Journal of Robotics and Mechatronics*, *33*(2), 263-273. https://www.jstage.jst.go.jp/article/jrobomech/33/2/33_263/_article/-char/ja/
10. Honey Optics. (2025, August 4). *How much do PTZ cameras cost? 3 factors determining the price.* Honey Optics. Retrieved from https://honeyoptics.com/how-much-do-ptz-cameras-cost/?srsltid=Kenjo, T., & Sugawara, A. (1994). *Stepping motors and their microprocessor controls* (2nd ed.). Clarendon Press.
11. Li, Y., Wang, Y., Ma, Z., Wang, X., Tan, B., & Ding, S. (2025). An intelligent retrievable object-tracking system with real-time edge inference capability. *IET Image Processing*, *19*(1),e13297.https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ipr2.13297
12. Lu, Y., Ma, H., Smart, E., & Yu, H. (2021). Real-time performance-focused localization techniques for autonomous vehicle: A review. *IEEE Transactions on Intelligent Transportation,Systems*, *23*(7), 6082-6100. https://ieeexplore.ieee.org/abstract/document/9435134/
13. Malik, M. F., Ali, S., Javed, K., Khan, M. A., Ayaz, Y., Nam, Y., & Sial, M. B. (2025). Hybrid control paradigm for exploring VR teleoperation and DRL-driven autonomy in mobile robotics. *Multimedia Tools and Applications*, 1-30. https://link.springer.com/article/10.1007/s11042-025-21024-5
14. Proia, S., Carli, R., Cavone, G., & Dotoli, M. (2021). Control techniques for safe, ergonomic, and efficient human-robot collaboration in the digital industry: A survey. *IEEE Transactions on Automation Science and Engineering*, *19*(3), 1798-1819. https://ieeexplore.ieee.org/abstract/document/9638383/
15. Queiroz, C. C. D. (2025). Interface evaluation focused on usability and user experience: a case study of the Quimikando Prototype. https://repositorio.ufrn.br/bitstreams/9d37dfa3-e527-4b36-be69-7c982989a1da/download
16. Rehman, A. U., Khan, Y., Ahmed, R. U., Ullah, N., & Butt, M. A. (2023). *Human tracking robotic camera based on image processing for live streaming of conferences and seminars*. Journal of Robotics and Control (JRC), 4(2), 90–101. https://doi.org/10.18196/jrc.v4i2.15382
17. Rehman, H., Khan, S., & Ahmed, R. (2023). Autonomous robotic camera system for live event coverage: Design, control, and field testing. *Journal of Field Robotics*, *40*(3), 412–429. https://doi.org/10.1002/rob.22145
18. Supertek Module. (2024, December 19). *How far can a security camera see?* Supertek Module. Retrieved from https://www.supertekmodule.com/how-far-can-security-camera-see/
19. World Bank. (2023, December). Education sector study: Philippines (Report No. 7473-PH). World Bank. https://documents1.worldbank.org/curated/en/981121468294364495/pdf/multi-page.pdf