

Revolutionizing Video Production: An AI-Powered Cameraman Robot for Quality Content [†]

Bara Fteiha , Rami Altai, Maha Yaghi  and Huma Zia * 

Electrical, Computer and Biomedical Engineering Department, College of Engineering, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates; 1071373@alumni.adu.ac.ae (B.F.); 1071991@alumni.adu.ac.ae (R.A.); maha.yaghi@adu.ac.ae (M.Y.)

* Correspondence: huma.zia@adu.ac.ae

[†] Presented at the 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES 2023), Plovdiv, Bulgaria, 23–25 November 2023.

Abstract: In today's world of growing user-generated content on social media, this study addresses the challenge of producing high-quality content, be it for social engagement or educational purposes. Conventionally, using a cameraman has been an effective yet expensive way to enhance video quality. In this context, our research introduces an innovative AI-driven camera robot that autonomously tracks the content creator, thereby improving video production quality. The robot uses an object detection model composed of YOLOv3 and Kalman filter algorithms to identify the content creators and create a bounding box around them within the frame. Using motion detection control, the robot adjusts its position to keep the bounding box centered in the frame, ensuring a continuous focus on the content creator. As a result, the system consistently captures excellent images through precise pan-tilt movements, promising improved visual storytelling. The initial results confirm the system's effectiveness in content detection, camera control, and content tracking. This advancement has the potential to impact user-generated content across various domains, providing an accessible way to enhance content quality without the high costs associated with traditional cameraman services.

Keywords: AI; ML; YOLOv3; arduino; autonomous camera control; robot camera; jetson AGX xavier; videography; content creation; broadcasting



Citation: Fteiha, B.; Altai, R.; Yaghi, M.; Zia, H. Revolutionizing Video Production: An AI-Powered Cameraman Robot for Quality Content. *Eng. Proc.* **2024**, *60*, 19. <https://doi.org/10.3390/engproc2024060019>

Academic Editor: Renato Filjar

Published: 15 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the realm of artificial intelligence, the growing influence of machine learning and deep learning is causing a significant transformation in the field of information technology. This transformation is bringing about profound changes, creating a new paradigm where automated AI systems can mimic and sometimes even replace existing roles [1,2]. The increasing availability and adaptability of AI-based technology align with its surging demand. The 21st century has introduced a wide range of new needs, and innovative solutions are emerging to meet them [3,4]. As technology advances on social media platforms and channels, there is a growing need for more advanced devices. In response to this, there has been a notable increase in demand. This shift has led to changes in how modern devices cater to the needs of media influencers and content creators are marketed [5,6]. The widespread use and advancement of artificial intelligence (AI) have made it applicable across various digital devices. In the field of media, the integration of modern equipment has resulted in significant advancements, driving progress in the industry. One notable innovation involves an automated cameraman with the ability to recognize and efficiently capture its subject, creating a more immersive experience for viewers [7]. In the era of remote education, media is undergoing a significant transformation, blending both traditional and modern forms. Social media influencers play a crucial role in this shift by continually sharing evolving content, with video blogging becoming a widespread practice.

Camera operators, essential for both cinematic and standard video production, experience improved accuracy in tracking and detection thanks to the integration of AI. The boundaries between capturing, enhancing, and altering images become less distinct due to artificial intelligence. Enhancing video cameras with AI technology introduces an intelligent cameraman—an entity capable of following user commands, resulting in professional-quality content. Videography is a challenging endeavor for both those behind and in front of the camera, which grapples with issues arising from equipment changes and repetitive recording. To address these challenges, AI-driven algorithms are poised to reduce inefficiencies and enhance outcomes, potentially revolutionizing media production [8,9]. This paper presents the concept of an AI-based videographer to simplify high-quality filmmaking using intelligent robotic assistance. The proposed system is designed to record and track moving objects, enabling detailed analysis of individual subjects using artificial intelligence techniques and Convolutional Networks to precisely identify their boundaries. The framework aims to reduce the labor and time burden on professional media practitioners by incorporating advanced technologies like tracking, facial recognition, and auto-focus into their production processes, laying the groundwork for the seamless integration of future advancements. After identifying objects, the system conducts a thorough statistical analysis using specific methods such as YOLOv3 along with a Kalman filter to achieve the desired results. The primary aim of this system is to provide robotic guidance, enhancing precision and accuracy while expanding accessibility to a broader user base. With its suitability for both personal and business settings, this innovation broadens its consumer appeal, offering benefits such as improved structural integrity, stability, and the inclusion of modern features [10,11].

2. Related Works

A thorough examination of the existing systems was conducted, carefully evaluating their advantages and disadvantages. To address the common camera operator challenge, the central question we are exploring is whether artificial machinery can match the skills of a human virtual camera operator. We have documented the key characteristics of these systems. Our study delves into this question by investigating camera automation. We have discovered that this automation can create panoramic views and extract virtual perspectives. These automated mechanisms control aspects like zoom, pan, and tilt. This system is impressively well-structured and supported by commendable research. However, it is crucial to note that it has a specific focus on capturing viral sports events. This specialization limits its suitability for everyday users [12–14]. On a different note, the Pivo Pod X offers an alternative for enhancing mobile filming through AI. It combines smooth tracking and LED lighting to make content creation more appealing. It operates with both cinematic AI and composition AI but relies on the user's smartphone, lacking an integrated camera. Despite its technical capabilities, the Pivo Pod X's tracking performance has faced criticism, particularly when users approach too close, and the content quality may not consistently meet user expectations [15–17]. In a separate context, the BBC effectively utilizes artificial intelligence and machine learning to expand its coverage of live events.

The integration of these technologies has expanded the broadcasting capabilities in various event and television programming areas. However, it is crucial to recognize a limitation in that their focus is specialized to the BBC's specific needs. This specialization, combined with its limited commercial availability and high costs, makes this solution impractical for public use [18]. In a different context, the paper titled "Automatic tracking of laparoscopic instruments for autonomous control of a cameraman robot" presents a new method for guiding a cinematographer robot during laparoscopic surgeries. This method uses a unique marker-free segmentation algorithm to locate surgical instruments in laparoscopic images. While this innovation is impressive for surgical use, it is important to note its narrow focus, primarily designed for capturing surgical procedures and not suitable for broader educational or general user applications [19]. Similarly, there is the concept of an "Autonomous Robot Photographer" designed for event photography. It autonomously

moves around, takes pictures of attendees, and sends them to the designated recipients. However, it is important to note that this innovation is specialized in photography and does not extend its usability to educational or commercial purposes. It is also considered outdated in terms of modern technology standards, limiting its relevance in the 21st century, despite its basic human interaction capabilities [20].

In comparison, “Human tracking robotic camera based on image processing for live streaming of conferences and seminars” is a relevant example. This innovation detects seminar speakers and tracks their movements in both horizontal and vertical directions, improving the quality of live streams. However, it has a limitation: it is partially stationary and can only be placed in a fixed corner position, restricting its movement to redirection [21]. Moreover, “Robotic Cinematography: The Construction and Testing of an Autonomous Robotic Videographer and Race Simulator” introduces an autonomous cinematic robot with a car chassis, a camera system, and a pan-tilt mechanism. This robot efficiently tracks runners on a predefined racecourse. Yet, its application is limited, mainly suitable for running-related contexts, with minimal potential for broader educational or commercial use [22]. In the study “Autonomous Robots for Deep Mask-Wearing Detection in Educational Settings during Pandemics”, a face mask detector was developed using the MobileNetV2 architecture and YOLOv2 object detector classification. Our findings indicate that our robot can move through an educational setting, identify violations, and evade obstacles effectively. Furthermore, it achieved an average precision rate of 91.4% during testing [23]. In the study “Real-time Contact Tracing During a Pandemic using Multi-camera Video Object Tracking”, a device designed to facilitate safe distancing between objects using surveillance cameras was introduced. The system employs various algorithms, including background subtraction to isolate foreground objects, morphological operations to eliminate noise, and blob analysis to detect connected regions within the foreground video. Additionally, Kalman filters are utilized to estimate the motion of objects in the video, and the Euclidean distance between objects is computed to track their interactions [24]. Our incentives in the paper are mainly summarized in the following:

1. The robot should be able to function in different environments, including its ability to track the subject of the video, and its ability to move around the environment in which it is filming.
2. The robot’s motor movement should be tightly integrated with the movement of the camera pan-tilt so that the tracking is accurate, smooth, and highly reliable.
3. The system should be simple in terms of usage for the average user.
4. The system should be versatile and accommodate most user applications.

The remainder of the paper is organized as follows: Section 3 introduces the materials and methods of our proposed system. Section 4 discusses the testing results. Finally, conclusions are drawn in Section 5.

3. Materials and Methods

This unique new AI-based camera tracking system includes important components like Jetson AGX Xavier, an NVMe SSD, a robot chassis, motors with encoders, a 222 Wh rechargeable battery pack, a camera sensor, camera sensor lenses, a camera pan-tilt kit, stepper motors, an Arduino UNO Rev3, an Arduino Mega, a camera tripod, and L293D motor controllers. The system architecture comprises three main parts, each playing a significant role in the overall framework. The primary part, known as the control system, relies on Jetson AGX Xavier as its core component. It facilitates seamless communication between the camera and the other robot components. When detection is confirmed, a specific signal is sent, triggering the registration of commands that govern the robot’s subsequent movement. The Jetson module then sends precise commands to the pan-tilt mechanism and the robot’s base, depending on the desired action. These commands are executed by stepper motors integrated into the system’s hardware, enabling precise movements as per the user’s instructions. The second part is the detection system, which operates within the computer-vision-based image processing pipeline. The video stream

goes through three stages. In the first step, we use the YOLOv3 (You Only Look Once) object detection algorithm to create a box around the person using the device, called the “frame target”. After that, we move to the next step, where the Kalman filter estimation algorithm comes into play. This algorithm calculates the necessary movements to resize and reposition the box so that the person inside it is centered in the frame. To explain this further, imagine a camera scanning its surroundings and using trained models to detect a user’s face by drawing a box around it, signifying that this is the item of interest. Once we confirm this initial detection, we start recording the video. Lastly, the third part of the system involves tracking the user. When we identify the user using the trained model for precise detection, we create a bounding box around them using computer vision algorithms, such as OpenCV, which detects objects. As a result, based on the camera’s detected movements, the robot repeats precise actions to keep the user’s face centered in the image. Figure 1 illustrates the overall working of the system.

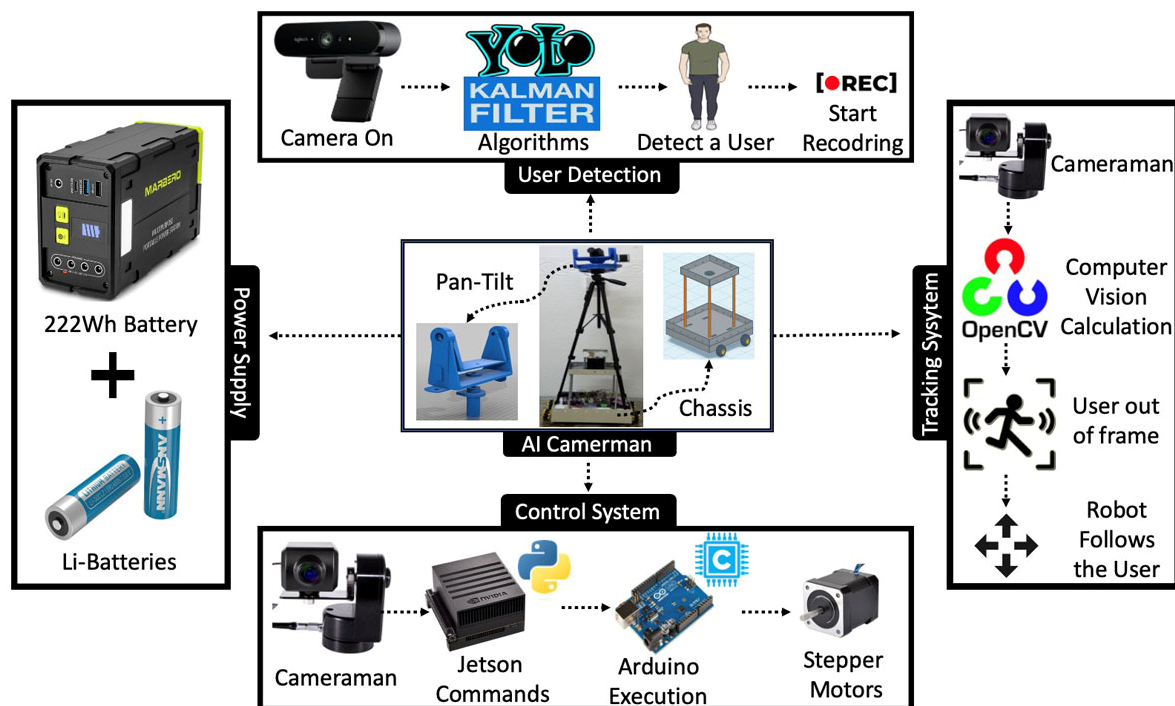


Figure 1. Overall system overview.

3.1. Proposed AI Cameraman Subsystems

3.1.1. System Hardware

This system relies on pre-trained YOLOv3 deep neural networks and depends on Jetson AI computation support. Additionally, the robot’s operation necessitates an independent power source, separate from the motor driver batteries. To implement this proposed system, we require specific hardware components. This includes a pan-tilt Arduino equipped with two stepper motors, which respond to commands from Jetson Xavier and adjust accordingly. Similarly, we need a robot Arduino with four stepper motors designed similarly to receive commands from Jetson Xavier and move in all four directions. Figure 2 illustrates the connections for the pan-tilt subsystem, with the robot connections differing only by having four stepper motors instead of two.

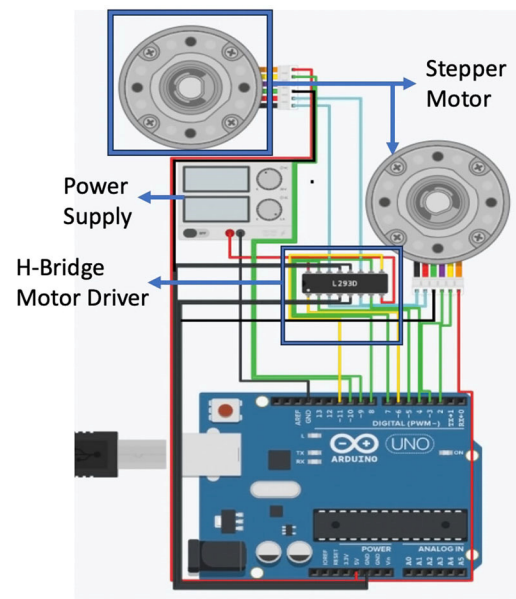


Figure 2. Pan-tilt circuit connections.

3.1.2. Pan Tilt Design

To allow the system to be versatile, a custom pan-tilt was designed and 3D-printed using Ultimaker; the design entailed two adjustable parts driven by motors connected to the Arduino. These two metrics represented the directions in which the pan-tilt would be responsible for steering the camera based on the Jetson Xavier commands, whether they were up, down, left, or right. Figures 3 and 4 show the execution of the pan-tilt design.

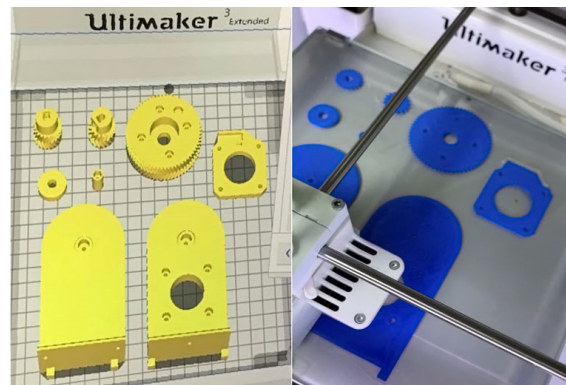


Figure 3. Pan-tilt 3D design.

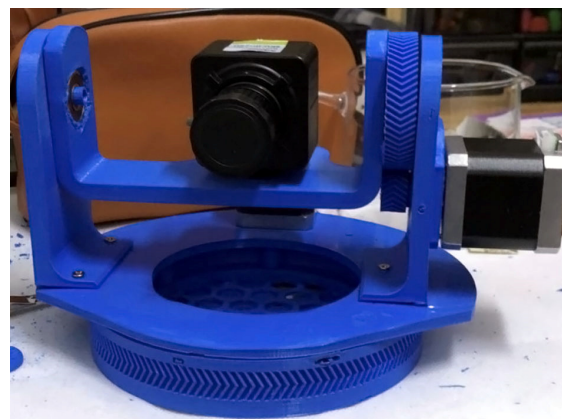


Figure 4. Pan-tilt assembled.

The pan-tilt Arduino utilizes four pins to alter the pan-tilt's positioning. Using the DIRH and STEPH pins, the pan-tilt can hover horizontally on the x-axis. Similarly, the pan-tilt gains its vertical movement through the STEPV and DIRV pins. In the pan-tilt, both pins of each axis are linked independently to two stepper motors. The pan-tilt may be changed horizontally with one motor and vertically with the other. This is undertaken via the use of a message buffer of size 3. The pins on the Arduino are set to OUTPUT in the empty setup section to operate the two stepper motors in the pan-tilt. This i2c library is also linked to the Arduino, allowing it to read i2c messages and run the code that moves the pan-tilt. To interact with the pan-tilt and control it, Jetson Xavier calls the Arduino address 0X11. Depending on which direction the pan-tilt is meant to go, the motors are set to HIGH or LOW.

3.1.3. Robot Chassis Design

Due to the originality of the innovation, the robot chassis in Figure 5 had to be custom-designed to make sure it served the needs of the project with two compartments; the upper section holds the pan-tilt.

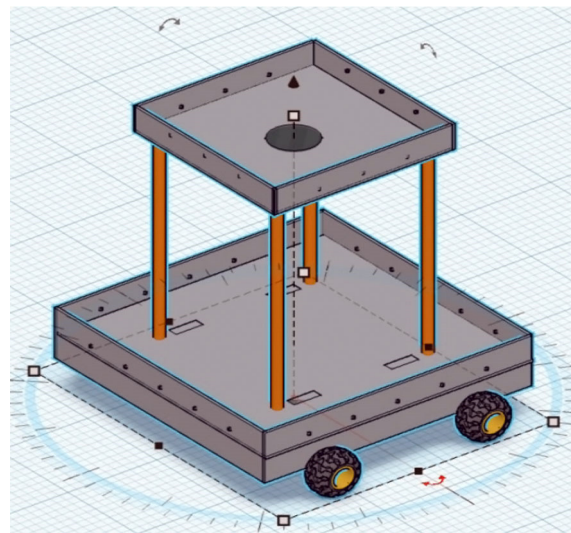


Figure 5. Robot chassis design.

The robot circuit connections were performed in a manner similar to the pan-tilt circuit connections: the only difference is that the robot required four stepper motors rather than two for the pan-tilt. The lower compartment is designed to withhold all the electrical components required to operate the invention.

3.1.4. System Calculations

To obtain high quality, the system frame resolution is set at 1280×720 . We have noticed that the experiment camera covers only 60° of the scene, giving the camera a 30° deviation horizontally. To cover 360° , several calculations must be obtained. The pan-tilt gear ratio is obtained by dividing the number of teeth in the base gear by the motor base gear, as shown below.

$$\text{Ratio} = \frac{\text{BaseGear}}{\text{MotorBaseGear}} = \frac{144}{17} = 8.74 \text{ rev.} \quad (1)$$

The number of steps needed to cover 360° is calculated by multiplying the revolutions with the steps as shown below:

$$\text{Steps} = 200 \times \text{Revolutions} = 1694 \text{ steps.} \quad (2)$$

Each degree moved corresponds to a number of steps calculated in the following manner:

$$\text{Steps per Degree} = \frac{\text{Steps}}{\text{Degree}} = \frac{1694}{360} = 4.7 \frac{s}{d}. \quad (3)$$

Thus, the maximum adjustment for every command that must be less than 255 is calculated as follows:

$$\text{Steps} = 4.7 \times 30 = 141 \text{ steps}. \quad (4)$$

We connected the degrees to the pixels by calculating the difference between the center of the recorded frame and the center of the face identified in the bounding box, the result of which is the required iteration in pixels:

$$\text{NeededAdjustinPixels} = \text{FaceCenter} - \text{FaceBoxCenter}. \quad (5)$$

The needed adjustment in pixels is then linked with the number of angles that the camera covers.

$$\text{Pixel Per Degree} = \frac{\text{Frame Width}}{\text{Camera Angle}} = \frac{1280}{60} = 21.33. \quad (6)$$

Finally, to calculate the adjustment in the degree, to obtain the total steps, the adjustment in pixels is divided by the pixel per degree and multiplied by 4.7.

For a full 360° rotation, the ratio of the diameter of the chassis and the wheel must be obtained as shown below:

$$\text{Ratio} = \frac{D. \text{ Chassis}}{D. \text{ Wheel}} = \frac{106.8}{15.2} = 6.8 \text{ rotations}. \quad (7)$$

which results in the following steps:

$$\text{Steps} = 6.8 \times 200 = 1360 \text{ steps}. \quad (8)$$

The same calculations followed for the pan-tilt x-axis are undergone for vertical adjustment; the only difference is that the camera coverage is 34° with 17° deviation from the center. The ratio is obtained according to the following:

$$\text{Ratio} = \frac{\text{BaseGear}}{\text{MotorBaseGear}} = \frac{64}{17} = 3.05 \text{ rev}. \quad (9)$$

while the steps are said to be:

$$\text{Steps} = 200 \times \text{Revolutions} = 610 \text{ steps}. \quad (10)$$

which, in return, means that every degree required the movement of 1.7 steps where the maximum steps can be calculated as:

$$\text{Maximumsteps} = 1.7 \times 30 = 29 \text{ steps}. \quad (11)$$

where pixels per degree will be found as:

$$\text{Pixel Per Degree} = \frac{\text{Frame Width}}{\text{Camera Angle}} = \frac{720}{34} = 21.18. \quad (12)$$

The z-axis adjustment is required to keep a proper distance from the content creator. Considering that there are no sensors in use, the system will reply on the bounding box size, which is said to be 20 cm.

$$\text{PixelperDegree} = \frac{\text{FrameWidth} \times 360 \times \text{HeadSize}}{\text{HeadFrameRatio} \times 2560 \times \pi \times \text{CameraAngle}}. \quad (13)$$

Finally, the difference between the calculated distance and the observed distance is obtained. The required distance iteration is obtained through the wheel diameter being 5 cm and the wheel rotation being 15.7. Thus, the steps per cm are:

$$\text{PixelperDegree} = \frac{200}{15.7} = 12.75 \text{ steps.} \quad (14)$$

3.1.5. System FSM

Figure 6 shows the proposed system flowchart, clarifying the states of each subsystem as well as the communication stages between the subsystems and Jetson Xavier. The FSM elucidates the various operational states initiated by Jetson AGX Xavier, impacting both the pan-tilt and the robot. It commences with the “idle” state, where essential parameters such as the file format, path, resolution, and filename configuration are established. Simultaneously, the computer vision model is loaded, and video streaming is initiated. These prerequisites lay the foundation for commencing the recording process. As the number of frames accumulates during the recording, the system initiates the detection of a human face. This detection process is triggered every 10 frames. Subsequently, a signal is transmitted to adjust the distance between the user and the robot chassis, taking into account any specified movements that need to be executed. Moreover, at intervals of every 5 frames, the pan-tilt mechanism receives instructions to commence its adjustment process, which can be either on the horizontal or vertical axis. Communication primarily relies on the I2C protocol to fulfill these operational requirements. The pan-tilt operates using discrete steps, while the robot chassis maneuvers based on specific angles and distances. The messages conveyed to the pan-tilt contain designated letters indicating the intended direction of movement; a similar operational protocol is adhered to by the robot chassis.

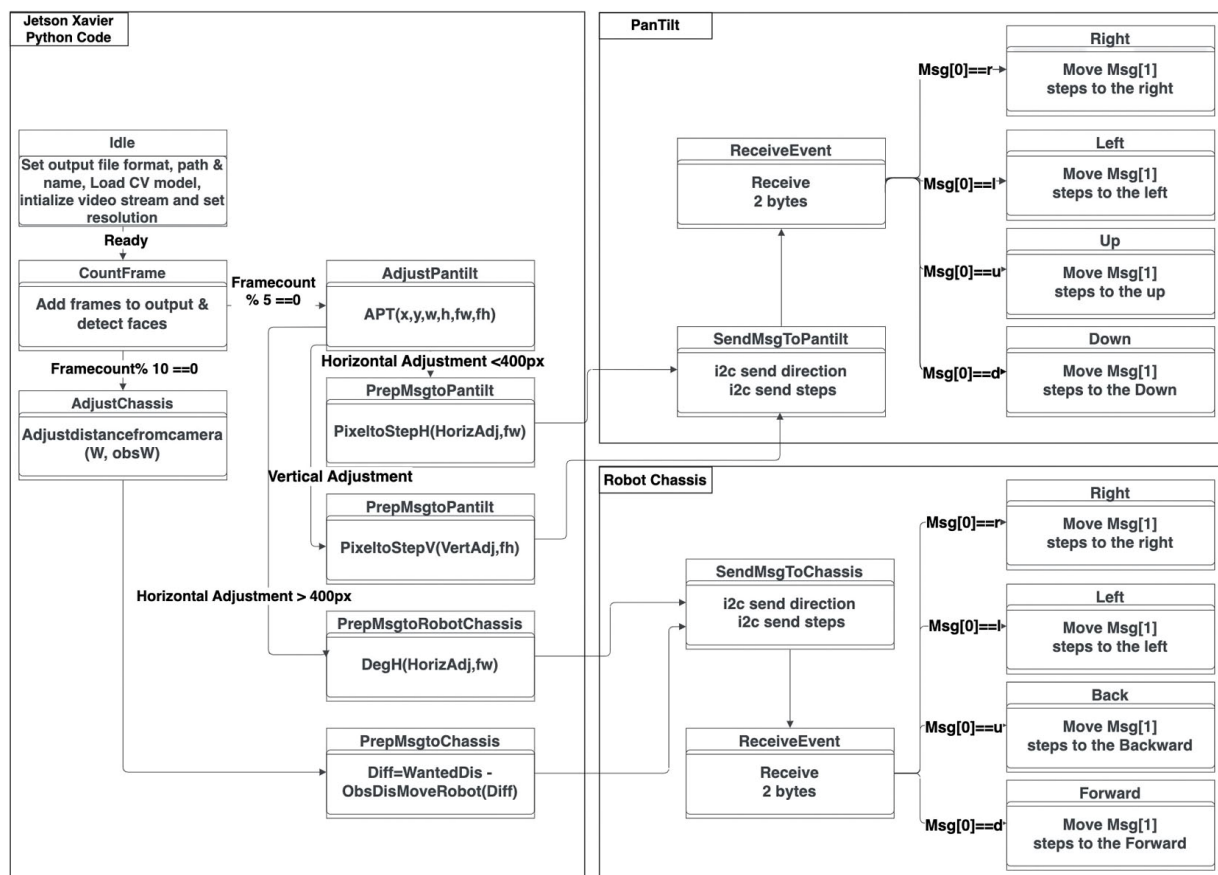


Figure 6. System flowchart.

4. Results

When Jetson AGX Xavier is activated, it establishes communication links to initiate the AI-based camera operation. The person using the system should be approximately 1.5 m away from the camera. When a face is detected by YOLOv3, the pan-tilt system adjusts every five frames to ensure the user remains in the initial frame by utilizing the Kalman filter. This keeps the content creator precisely focused, and a bounding box is drawn around their face to indicate the area that will be controlled. This information is continually communicated with Jetson during filming. The camera's readings are sent to Jetson, which processes them and makes decisions. If adjustments are needed to the pan-tilt or the robot's movements, the corrected instructions are promptly provided and implemented. The robot functions as a cameraman and can be directed to perform specific tasks while maintaining an appropriate distance from the user. We designed our robot to be roughly the same height as the average user to achieve results comparable to those of a human cameraman. Regardless of the user's movements, the AI robot can follow and adjust the camera's perspective to keep the user centered. The robot not only moves in parallel with the user but also synchronizes with the pan-tilt system, whether the movement is vertical or horizontal. Figure 7 illustrates the testing setup with the user initiating the robot and standing in front of it, while Figure 8 displays the camera view with the generated face bounding box, which aids in the centering and tracking of the content creator.



Figure 7. User detection test.

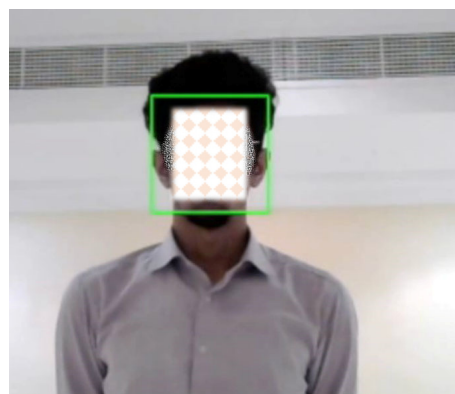


Figure 8. Face bounding test.

Figure 9 illustrates the user coming closer to the robot and changing his position to the left; the response of the robot was simultaneously keeping a 1.5 m distance from the user, as well as fixing the frame to the left as the user moved to the left.



Figure 9. Testing changing user position.

The ensuing hyperlink—<https://www.youtube.com/watch?v=Dd3nVK-toBs> (accessed on 16 September 2023)—leads to a video carefully created to help explain and show the AI cameraman’s functionality and showcase the observed results.

5. Conclusions

In this paper, an AI cinematography robot that recognizes users and captures high-quality videos was implemented. This robot makes it easy for users to shoot professional-grade footage. The camera records the video and sends it to a connected computer system for processing. The YOLOv3 method is used to create a bounding box around the user, then center the user in the frame, and estimate the necessary camera movements. The PID control processes these data and controls the robot’s motors to move it in the intended direction. The brain of the robot is Jetson Xavier, which sends instructions to the pan-tilt and robot’s Arduinos, guiding their movements to ensure accurate filming. Both the robot and pan-tilt respond to Jetson Xavier’s commands, keeping the user within the bounding box. This innovation has great potential for improvement. For example, it could be made more environmentally friendly by using recyclable materials. Additionally, alternative power sources like solar panels could be used to sustain the cameraman. The robot focuses solely on one person and is not intended to handle multiple individuals within the frame. Adding more people would complicate matters unnecessarily. The robot can automatically adjust its angle to accommodate people of different heights, providing a wide coverage range, yet an option to manually adjust the tripod’s height could be added. It is important to note that the robot has not been tested with significant obstacles, as it is designed for smooth surfaces due to its weight limitations and chassis design constraints. The robot uses mecanum wheels, allowing it to move smoothly to the right or left without rotating to avoid obstacles, while adjusting the pan-tilt horizontally based on the base’s movement to sidestep any obstacle. To enhance the user’s experience, we consider adding digital impact lenses, portrait mode, and night mode to the camera.

Author Contributions: Conceptualization, R.A.; methodology, B.F. and R.A.; software, R.A. and B.F.; validation, R.A., B.F. and M.Y.; formal analysis, R.A.; investigation, R.A.; resources, M.Y.; data curation, R.A.; writing—original draft preparation, B.F.; writing—review and editing, B.F., M.Y. and H.Z.; visualization, B.F.; supervision, H.Z. and M.Y.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Abu Dhabi University's Office of Research and Sponsored Programs, the United Arab Emirates.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Acemoglu, D.; Restrepo, P. Robots and jobs: Evidence from US labor markets. *J. Political Econ.* **2020**, *128*, 2188–2244. [CrossRef]
2. King, T.M.; Arbon, J.; Santiago, D.; Adamo, D.; Chin, W.; Shanmugam, R. AI for Testing Today and Tomorrow: Industry Perspectives. In Proceeding of the 2019 IEEE International Conference on Artificial Intelligence Testing (AITest), Newark, CA, USA, 4–9 April 2019; pp. 81–88. [CrossRef]
3. Heo, J.; Kim, Y.; Yan, J. Sustainability of Live Video Streamer's Strategies: Live Streaming Video Platform and Audience's Social Capital in South Korea. *Sustainability* **2020**, *12*, 1969. [CrossRef]
4. Evans, M.; Kerlin, L.; Lerner, O.; Campbell, R. Feels like being there: Viewers describe the quality of experience of festival video using their own words. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*; CHI EA 2018; ACM: New York, NY, USA, 2018; pp. LBW029:1–LBW029:6. [CrossRef]
5. Carvalho, D.; Silva, N.; Cardoso, A.; Fazzion, E.; Pereira, A.; Rocha, L. Understanding Users-Contents Interaction in Non-Linear Multimedia Streaming Services. In Proceedings of the 24th Brazilian Symposium on Multimedia and the Web (WebMedia '18), Salvador, Brazil, 16–19 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 229–232. [CrossRef]
6. Veloso, E.; Almeida, V.; Meira, W.; Bestavros, A.; Jin, S. A hierarchical characterization of a live streaming media workload. In Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement (IMW '02), Marseille, France, 6–8 November 2002; Association for Computing Machinery: New York, NY, USA, 2002; pp. 117–130. [CrossRef]
7. Gaddam, V.; Eg, R.; Langseth, R.; Griwodz, C.; Halvorsen, P. The Cameraman Operating My Virtual Camera is Artificial: Can the Machine Be as Good as a Human? *ACM Trans. Multimed. Comput. Commun. Appl.* **2015**, *11*, 1–20. [CrossRef]
8. Pivo Pod x: Your Pocket-Sized Cameraman. Available online: <https://www.kickstarter.com/projects/getpivo/pivo-pod-x-your-pocket-sized-cameraman> (accessed on 15 September 2023).
9. Wright, C.; Allnutt, J.; Campbell, R.; Evans, M.; Forman, R.; Gibson, J.; Jolly, S.; Kerlin, L.; Lechelt, S.; Phillipson, G.; et al. AI in production: Video analysis and machine learning for expanded live events coverage. *SMPTE Motion Imaging J.* **2020**, *129*, 36–45. [CrossRef]
10. Khoiy, K.; Mirbagheri, A.; Farahmand, F. Automatic tracking of laparoscopic instruments for autonomous control of a cameraman robot. *Minim. Invasive Ther. Allied Technol.* **2016**, *25*, 121–128. [CrossRef] [PubMed]
11. Byers, Z.; Dixon, M.; Goodier, K.; Grimm, C.M.; Smart, W.D. An autonomous robot photographer. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), Las Vegas, NV, USA, 27–31 October 2003; Volume 3, pp. 2636–2641. [CrossRef]
12. Rehman, A.U.; Khan, Y.; Ahmed, R.U.; Ullah, N.; Butt, M.A. Human tracking robotic camera based on image processing for live streaming of conferences and seminars. *Heliyon* **2023**, *9*, e18547. [PubMed]
13. Ike, R. Robotic Cinematography: The Construction and Testing of an Autonomous Robotic Videographer and Race Simulator. Doctoral Dissertation, University of Huston, Houston, TX, USA, 2021.
14. Zia, H.; Alhalabi, M.; Yaghi, M.; Barhoush, A.; Farag, O.; Alkhedher, M.; Khelifi, A.; Ibrahim, M.A.; Ghazal, M. Autonomous Robots for Deep Mask-Wearing Detection in Educational Settings during Pandemics. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 5626764. [CrossRef]
15. Yaghi, M.; Basmaji, T.; Salim, R.; Yousaf, J.; Zia, H.; Ghazal, M. Real-time Contact Tracing During a Pandemic using Multi-camera Video Object Tracking. In Proceeding of the 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 8–9 November 2020; pp. 872–876. [CrossRef]
16. Fang, H.; Zhang, M. Creatism: A deep-learning photographer capable of creating professional work. *arXiv* **2017**. [CrossRef]
17. Thorburn, E.D. Social media, subjectivity, and surveillance: Moving on from Occupy, the rise of live streaming video. *Commun. Crit./Cult. Stud.* **2014**, *11*, 52–63.
18. Li, Z.; Kaafar, M.A.; Salamatian, K.; Xie, G. Characterizing and Modeling User Behavior in a Large-Scale Mobile Live Streaming System. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 2675–2686. [CrossRef]
19. Rejaie, R.; Magharei, N. On performance evaluation of swarm-based live peer-to-peer streaming applications. *Multimed. Syst.* **2014**, *20*, 415–427. [CrossRef]
20. Vrontis, D.; Christofi, M.; Pereira, V.; Tarba, S.; Makrides, A.; Trichina, E. Artificial intelligence, robotics, advanced technologies, and human resource management: A systematic review. *Int. J. Hum. Resour. Manag.* **2022**, *33*, 1237–1266. [CrossRef]
21. Rubio, F.; Valero, F.; Llopis-Albert, C. A review of mobile robots: Concepts, methods, theoretical framework, and applications. *Int. J. Adv. Robot. Syst.* **2019**, *3*, 3–8. [CrossRef]
22. Lokanath, M.; Sai, G.A. Live video monitoring robot controlled by web over internet. In *Proceeding of the 14th International Conference on Science, Engineering the Technology 14th (ICSET2017)*; Vellore, India, 2–3 May 2017; IOP Conference Series: Materials Science and Engineering; IOP Publishing: Bristol, UK, 2017; Volume 263. [CrossRef]

23. Sagheer, A.; Aly, S. An Effective Face Detection Algorithm Based on Skin Color Information. In Proceeding of the 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems, Sorrento, Italy, 25–29 November 2012; pp. 90–96. [[CrossRef](#)]
24. Zhang, Z.; Zhong, Y.; Guo, J.; Wang, Q.; Xu, C.; Gao, F. Auto Filmer: Autonomous Aerial Videography Under Human Interaction. *IEEE Robot. Autom. Lett.* **2023**, *8*, 784–791. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.