AI Enable Efficient Learning: An Edge Intelligence Driven Monocular Lecture Video Recording System

Rui Wang(✉), Zesen Zou, and Yang Gao

Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China
{wangr,zasen2000,gy_albert}@buaa.edu.cn

Abstract. Currently, smart classrooms in universities generally record live teaching videos based on multiple traditional cameras, which is resource-wasteful and inefficient. To address this problem, this paper presents an edge intelligence driven monocular lecture video recording system based on pan-tilt-zoom (PTZ) components and NVIDIA Jetson TX2. The system uses an AdaBoost-based face detector to identify the teachers' faces, applies a Siamese network to track them, and employs the PID algorithm to control the PTZ camera for active auto-tracking and recording. Results indicate that our intelligent monocular PTZ video capture device performs effectively in terms of stability, real-time and economy, thus having the potential to benefit both teachers and students.

Keywords: Lecture video Record · Face Tracking · PTZ Camera · Edge platform · Automatic

1 Introduction

In recent years, the use of artificial intelligence (AI) has played a significant role in enhancing the educational efficiency. AI, edge computing, and intelligent of teaching technologies have made education smarter [1], more comprehensive and automated, thus have produced a range of outcomes, including smart campus management [2], smart classrooms, teaching robots, smart teaching platforms, and more [3].

Video capture device is a crucial aspect of education nowadays [4]. The recorded teaching videos can be used as a regular recording and broadcasting system, as well as live teaching, connecting online and offline to meet the needs of regular interactive online learning [5] in the post-epidemic era. However, traditional manual-based methods of lecture recording use a course capturing solution that place fixed cameras in multiple locations, which is resource-wasteful and inefficient. Furthermore, even though this solution achieves automatic course shooting, it has disadvantages of limited field of view (FOV), difficulty in arrangement, high cost and complicated operation.

In order to produce high-quality teaching video recordings with monocular vision automatically, visual object tracking (VOT) technology with AI is needed to direct the

D. Kumar et al. (Eds.): IEIT 2023, AHSSEH 10, pp. 233–239, 2023.
https://doi.org/10.2991/978-94-6463-230-9_29

capture like a human camera operator [6]. Compared with fixed camera, pan-tilt-zoom (PTZ) camera, is more conducive to expanding the FOV. In addition, Edge Intelligence, which combines AI and edge computing, can enable teaching videos recording device to be smaller, contain fewer components, and more economical. AI algorithms can run directly on the Edge platform, without the constraints of networking, size, etc. The combination of edge and cloud can further enable efficient connectivity with smart campus and campus cloud, and puts real-time computing at the edge without adding the burden of server computing as well as video transmission.

The primary contribution of this research is the design and implementation of an intelligent lecture video recording system based on edge computing. Our system achieves high-quality automatic recording in teaching scenarios by simulating the behavior of a human camera operator capturing online lectures, and adopting advanced face detection, tracking, and camera control algorithms. The innovative design and implementation of the system provide a low-cost, yet effective and fully automatic way of recording lecture videos for multimedia teaching monitoring management system, thereby bringing convenience to the teaching and supervision work and promoting access to quality education.

2 System Hardware

This paper presents the design of a face target tracking and shooting system using the Nvidia Jetson TX2 platform and Sony EVI-D90P camera, as shown in Fig. 1. The system includes a visual face tracking module and a PTZ camera control module. The visual face tracking module uses a face detector to locate the teacher's face, and the PTZ camera control module sends control signals to the camera to achieve stable tracking of the teacher.

The NVIDIA Jetson TX2 development module is an AI System on Chip utilizing the Pascal™ architecture. It has six CPU cores, an NVIDIA Pascal™ GPU with 256 stream processors, and 8GB of LPDDR4 memory. Its small size, high throughput, and low power consumption make it ideal for embedded applications. Additionally, deep learning models can be trained in the cloud and deployed on TX2 endpoints using the official interface provided by NVIDIA.

The system uses a Sony EVI-D90P camera as the PTZ camera for tracking recording, which supports shooting in dim light environments with a horizontal rotation range of −170 to 170°, and a vertical rotation range of −20 to 90°. The maximum support is 720*540, 25fps video output. The camera outputs video signals using the Video port and receives control data streams from the TX2 platform using the RS422 interface. The VISCA camera control protocol was used to implement the TX2 control signal connection to the camera.

3 Software Design

The system is composed of three main components: face detection, target tracking, and PTZ camera control. The Fig. 2 explain the workflow. The target detection task involves locating the teacher's face in the video frames. Currently, visual target detection can

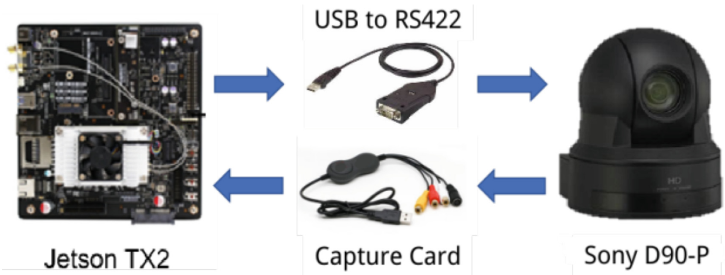


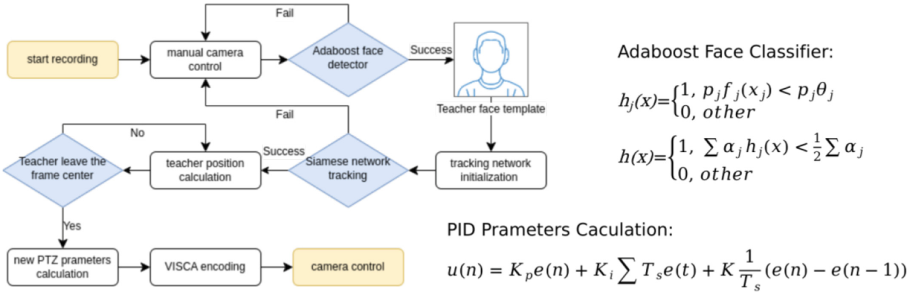
Fig. 1. System hardware connection

be classified into two categories: Convolutional neural networks (CNN) and traditional detection algorithms. Convolutional neural network detection algorithms offer higher accuracy, but require significant computational effort. On the other hand, traditional face detection algorithms are suitable for teaching scenarios with low background complexity, as they can detect teachers' faces accurately while achieving higher real-time performance.

Once the target face has been detected in the images, the tracking module must continuously track its position. Currently, the most advanced real-time target tracking algorithms include the Siamese network [7] and correlation filtering algorithms. Notably, the Siamese network tracking algorithm exhibits outstanding tracking accuracy while maintaining a high inference speed, making it a popular choice among scholars.

The tracking algorithm used in this paper is a Siamese network-based target tracking algorithm, as shown in Fig. 3. A Siamese network is a neural network structure containing two branches and is often used for one-shot learning tasks. The inference process in Siamese networks usually consists of two stages. First, one branch of the network learns from the template samples, and then the other branch processes the input image and outputs the target result based on the learned information. The two branches of the Siamese network generally share parameters and perform a similarity measure between the two inputs to obtain an accurate result.

The final component of the system is the PTZ camera control. Most PTZ camera control methods fall into two categories: PID or preset speed methods. Compared with the latter, PID can provide a stable platform for the PTZ camera to operate from, which


start recording
manual camera control
Fail
Adaboost face detector
Success
Teacher face template
Fail
tracking network initialization
Success
Siamese network tracking
teacher position calculation
No
Teacher leave the frame center
Yes
new PTZ prameters calculation
VISCA encoding
camera control
Adaboost Face Classifier:
PID Prameters Caculation:

$$h_j(x)=\begin{cases}1, & p_j f_j(x_j) < p_j\theta_j \\ 0, & other\end{cases}$$

$$h(x)=\begin{cases}1, & \sum \alpha_j h_j(x) < \frac{1}{2}\sum \alpha_j \\ 0, & other\end{cases}$$

$$u(n) = K_p e(n) + K_i \sum T_s e(t) + K \frac{1}{T_s}(e(n) - e(n-1))$$

Fig. 2. System workflow

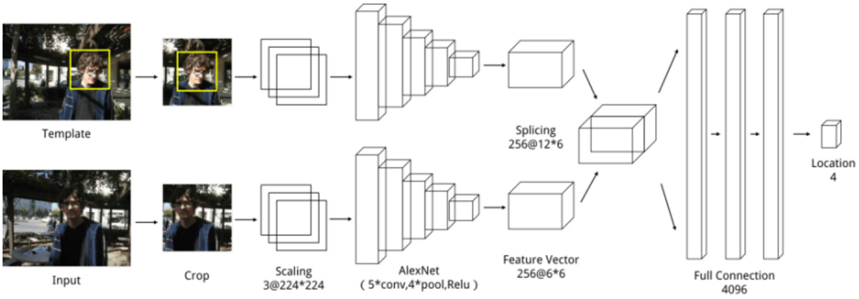


Fig. 3. Siamese Network tracking model

means it can help maintain a stable and precise position even if there are external factors involving.

In this paper, we utilize the superiority of AI tracking and edge computing platforms to solve the problem of unsatisfactory tracking capture and complex hardware system deployment when capturing classroom teaching content. The system consists of three main components: an Adaboost-based face detector, a Siamese network face tracker, and a PID-based control module for the PTZ camera. In this paper, we optimized and deployed the Siamese network algorithm on an edge platform to achieve high tracking speed and accuracy.

4 Experiment and Result

In the experimental scenario, a teacher delivers a lecture in a classroom with sufficient lighting and a display screen placed in the corner. The teacher moves around the podium, making frequent gestures and changes in posture. Despite these movements, the tracking system consistently keeps the teacher's face in the center of the screen in real-time. Screenshots captured during the experiment demonstrate the effectiveness of the tracking system (Fig. 4).

During the experiment, the teacher circumnavigated the podium to test the tracking system's capabilities. As shown in Fig. 5, the system tracked the target smoothly, even when the tester was under or on the podium. Moreover, the system was able to track

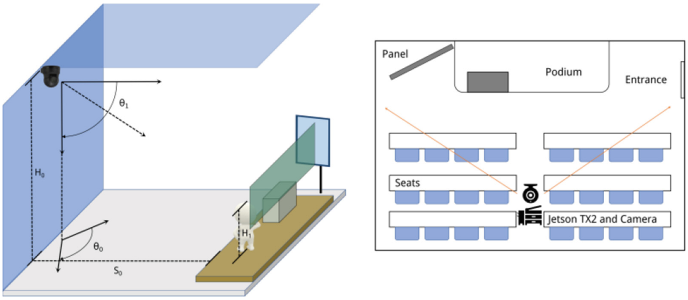
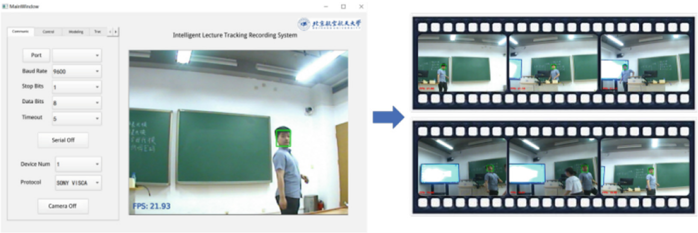


Fig. 4. Experiment scenario



Baud Rate 9600
Stop Bits 1
Data Bits 8
Timeout 5
Device Num 1
Protocol SONY VISCA

Fig. 5. Track and Capture when Moving

Table 1. Comparison of Lecture Recording System

Proposer	Camera Setting	Distance	Tracking Algorithm	Control Algorithm	platform
This paper	1(PTZ)	3 ~ 10 m	Siamese Network	PID	Edge platform
Dong H. [6]	1(PTZ)	5 m	TLD	Preset speed	PC
Mei L. [8]	1(PTZ)	5 m	Kalman Filter	Preset speed	PC
Wang H. [9]	1(PTZ)	2 m	Particle Filter and PDR	Space Positioning	PC
Singh S. [10]	1(PTZ)	5 m	HOG Matching	Preset speed	FPGA

the target steadily even when the person assumed various postures, including showing their front face, side face, and back of the head. The tracking algorithm precisely located the person's face in the image, enabling the tracking system to function correctly. These results demonstrate the system's robustness and effectiveness in diverse real-world scenarios.

The system built in this paper is suitable for recording teaching videos and is capable of achieving real-time, stable tracking recording at an average of 22 FPS in various teaching environments. The system can track without losing the target during one class period under various scenarios including pose changes, low light, human occlusion and object occlusion. Meanwhile, we made a comparison between the proposed system in this paper and the existing intelligent classroom recording systems using PTZ camera, as shown in Table 1. This comparison further demonstrates the superiority of the system for automatic course recording.

5 Conclusions

The research in this papaer demonstrates the practical value of the monocular intelligent lecture recording system based on edge computing. This system aligns with the innovative teaching mode centered on learners and the development of an intelligent education environment proposed by the Chinese government. With digital technology rapidly transforming the landscape of curriculum teaching and interaction, we believe that our work represents a significant step towards the integration of AI technology and education. Our hope is that this integration will enable more effective digitization and informatization of double first-class curriculum construction, empowering education through technology.

Acknowledgments. We would like to express our gratitude to the following funding sources that supported this research: the First-Class Undergraduate Course Construction Project of Beihang University and the Teaching Reform Fund Project of Beihang University.

References

1. C. Peñarrubia-Lozano, M. Segura-Berges, M. Lizalde-Gil, and J. C. Bustamante, (2021) A Qualitative Analysis of Implementing E-Learning during the COVID-19 Lockdown, Sustainability, vol. 13, no. 6: p. 3317, doi: https://doi.org/10.3390/su13063317.
2. L. Min, H. Xie, X. Gu and X. Hu, (2021) Research and Discussion on Key Technologies of Lightweight Smart Campus Based on LAPP, In:2021 International Conference on Internet, Education and Information Technology (IEIT), Suzhou, China, pp. 191–194, doi: https://doi.org/10.1109/IEIT53597.2021.00048.
3. A. Kaur and M. Bhatia, (2022) Smart Classroom: A Review and Research Agenda, IEEE Trans. Eng. Manage. pp. 1–17, 2022, doi: https://doi.org/10.1109/tem.2022.3176477.
4. H. Wang and J. Hu, (2022) Intelligent lecture recording system based on coordination of face-detection and pedestrian dead reckoning, PeerJ Comput. Sci., vol. 8: p. e971, doi: https://doi.org/10.7717/peerj-cs.971.
5. W. Rui, Z. Haoyuan, L. Hui, Q. Xiaolei, S. Jiangtao and X. Yuedong, (2021) Toward the flipped interactive teaching for "signal analysis and processing" to smarter education integrated with modern information technology, In:2021 2nd International Conference on Artificial Intelligence and Education (ICAIE), Dali, China, pp. 583–586, doi: https://doi.org/10.1109/ICAIE53562.2021.00129.
6. R. Wang, H. Dong, T. X. Han, and L. Mei, (2016) Robust tracking via monocular active vision for an intelligent teaching system, Vis Comput, vol. 32, no. 11: pp. 1379–1394, doi: https://doi.org/10.1007/s00371-015-1206-8.
7. Held, D., Thrun, S., Savarese, S. (2016). Learning to Track at 100 FPS with Deep Regression Networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.) Computer Vision – ECCV 2016. Lecture Notes in Computer Science, Springer, Cham. vol 9905. pp. 749-765. https://doi.org/10.1007/978-3-319-46448-0_45
8. Wang, R., Mei, L. (2013). Intelligent Tracking Teaching System based on monocular active vision. In: 2013 IEEE International Conference on Imaging Systems and Techniques (IST). China. pp. 431–436. https://doi.org/10.1109/ist.2013.6729736
9. Wang, H., Hu, J. (2022). Intelligent lecture recording system based on coordination of face-detection and pedestrian dead reckoning. PeerJ Computer Science, 8, e971. https://doi.org/https://doi.org/10.7717/peerj-cs.971
10. Singh, S., R., Saini, R., Saurav, S., Saini, A. K. (2017). Real-time Object Tracking with Active PTZ Camera using Hardware Acceleration Approach. International Journal of Image, Graphics and Signal Processing, 9(2), pp. 55–62. https://doi.org/https://doi.org/10.5815/ijigsp.2017.02.07.