IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

IoT-Based Smart Video Doorbell with Real-Time Monitoring and Two-Way Audio Communication

Utkarsha Mahadev Mane[1], Mr. M. D. Shinde[2], Prof. J. R. Mahajan[3]
Student, Department of Artificial Intelligence and Data Science[1]
Guide, Department of Artificial Intelligence and Data Science[2]
Co-Guide, Department of Artificial Intelligence and Data Science[3]
Adsul Technical Campus, Chas, Ahilyanagar, India

Abstract: *Conventional doorbells provide only a local audible indication that someone is waiting at the entrance. They do not provide remote visual awareness or a convenient way for the resident to communicate with the visitor. This paper presents an IoT-based smart video doorbell that combines a network camera, edge processing, real-time video monitoring, event notifications, and full-duplex audio communication in a single residential security unit. A compact controller such as an ESP32-class camera board can acquire visitor video and detect a doorbell event, while an edge or gateway processor can perform optional face verification and event filtering. The system forwards selected events through a secure Internet connection to a mobile or web interface. The resident can view the entrance remotely and communicate with the visitor through a microphone and speaker without opening the door. A modular architecture is proposed so that face recognition can be enabled when sufficient computing resources and an appropriate privacy policy are available. The design also considers authentication, encrypted communication, local fallback operation, and data-minimization practices. The proposed system offers an affordable and extensible approach for smart-home entry monitoring while addressing limitations of traditional wireless doorbells.*

Keywords: Internet of Things, smart doorbell, video monitoring, two-way audio, ESP32-CAM, edge computing, smart home security

I. INTRODUCTION

The growth of connected home technologies has changed the role of ordinary household devices from isolated appliances to networked services. A doorbell is a simple example: the traditional device only produces a sound inside the house, whereas a connected doorbell can provide visual information, remote notifications and interaction with a visitor. This capability is useful when residents are away from home, when the entrance is not directly visible, or when a visitor needs to communicate before access is granted.

Recent smart-doorbell research has explored several combinations of IoT connectivity, camera-based monitoring, face recognition and remote alerts. An IJSET publication describes an ESP32-CAM-oriented system incorporating face detection, visitor alerts, live video and two-way audio. A related IJARST publication demonstrates a wireless IoT doorbell concept, while a WJARR study integrates face recognition, cloud messaging and remote door-control functions. These studies show the feasibility of connected visitor monitoring, but they also indicate a need for a balanced design that treats live video, audio interaction, event handling and security as coordinated functions rather than independent features.

The objective of this work is therefore to design a practical IoT-based smart video doorbell with two-way audio. The proposed system focuses on five functions: (i) doorbell-event detection, (ii) entrance video acquisition, (iii) remote video access, (iv) full-duplex visitor communication, and (v) secure event notification. Face recognition is treated as an optional enhancement rather than the sole access decision, reducing the risk of unsafe automatic entry.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40358

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

II. OBJECTIVES AND CONTRIBUTIONS

- Develop a modular smart-doorbell architecture combining video, IoT connectivity and two-way audio.
- Provide a remote interface through which a resident can observe the entrance and communicate with a visitor.
- Support optional edge-based visitor/face analysis without requiring every video frame to be uploaded to the cloud.
- Introduce event-driven operation to reduce unnecessary network traffic and processing.
- Incorporate practical security measures including device authentication, encrypted transport, access control and limited data retention.
- Define measurable evaluation parameters such as notification delay, video latency, audio round-trip delay, recognition performance and system availability.

III. RELATED WORK

The first reference paper considered in this work presents an IoT-enabled smart doorbell using an ESP32-CAM and combines face recognition, real-time alerts, live video and two-way audio. Its main contribution is the integration of embedded vision with remote visitor interaction. The present paper differs by emphasizing a layered, event-driven architecture in which video, audio, recognition and notification services can be scaled independently [1].

The IJARST study on a wireless IoT doorbell demonstrates how wireless communication can transform a conventional doorbell into a connected smart-home component [2]. The proposed design extends that concept by treating the camera and audio path as first-class communication services rather than adding them only as supplementary features.

The WJARR publication reports a Raspberry Pi-based smart doorbell with face recognition, Telegram notifications and Blynk-based remote door control. The reported evaluation shows that recognition performance is sensitive to illumination and face orientation, which supports the decision in this paper to keep automated unlocking separate from recognition and to provide a human-in-the-loop interaction model [3].

Other research cited in the WJARR study covers contactless doorbells, smart-home wireless doorbells, IoT security, face authentication and video processing [4]-[8]. Collectively, these works motivate a system that prioritizes secure communication, low response time and privacy-aware processing.

IV. RESEARCH GAP

Existing emphasis	Observed limitation	Proposed approach
Wireless doorbell connectivity	Limited visitor context	Video + event notification
Face recognition	Sensitive to lighting/pose	Optional edge verification + human review
Remote alerts	Alert may be one-way	Alert + live video + two-way audio
Remote door control	Security risk if fully automated	Manual authorization; recognition is advisory
Cloud processing	Privacy and bandwidth concerns	Event-driven edge/cloud split

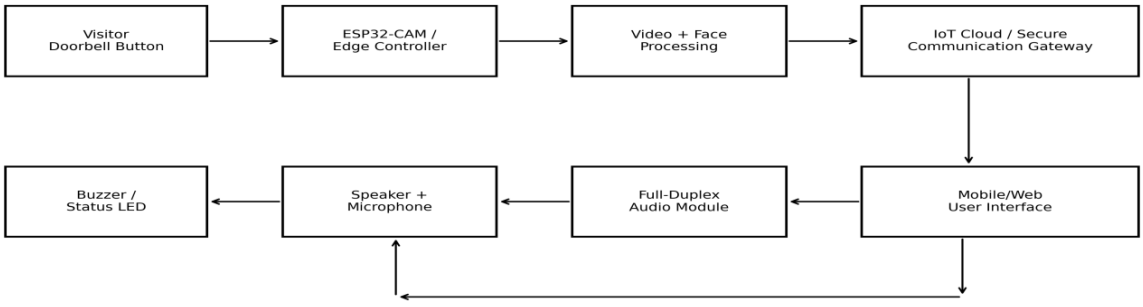
V. PROPOSED SYSTEM

The proposed smart doorbell is organized into four functional layers: sensing, edge processing, IoT communication and user interaction. At the entrance, a push button or touch sensor generates the visitor event. A camera captures the visitor view, while a microphone and speaker provide bidirectional audio. An edge controller manages the peripherals and performs basic filtering. When required, a gateway or local server performs computationally heavier video analysis such as face embedding and matching.

After an event is generated, the system creates a short-lived session. The user receives an event notification and can open the live video interface. Audio is then established as two simultaneous streams so that the resident and visitor can speak naturally. The session is terminated after a configurable timeout or user action. This event-driven operation avoids continuous cloud transmission when no visitor is present.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, October 2026 Impact Factor: 8.2

• Proposed layered architecture of the IoT-based smart video doorbell



VI. HARDWARE COMPONENTS

Component	Function	Design consideration
ESP32-CAM / camera controller	Image capture, event control and network connectivity	Compact, low power
Microphone	Captures visitor speech	Noise suppression preferred
Speaker	Plays resident speech	Adequate outdoor loudness
Push button/touch sensor	Triggers visitor event	Debouncing required
Wi-Fi router	Internet/network access	Reliable coverage near entrance
Edge gateway (optional)	Face analysis and session management	Used when local compute is required
Status LED/buzzer	Local feedback	Short and non-intrusive indication
Power supply	Provides regulated power	Surge and voltage protection

VII. SOFTWARE COMPONENTS

• Embedded firmware for GPIO, camera, button and network management.
• Video streaming service for live entrance monitoring.
• Audio capture/playback service supporting simultaneous transmit and receive paths.
• IoT messaging service for event notifications and device status.
• Optional OpenCV/face-embedding module for local visitor verification.
• Mobile/web interface for viewing, answering, muting and ending the visitor session.
• Secure authentication and encrypted transport for device-to-server and user-to-server communication.

VIII. SYSTEM METHODOLOGY

• Visitor Event Detection

The controller continuously checks the doorbell input in a low-power monitoring state. When the input changes to the active state, software debouncing is applied to prevent repeated triggers. A valid event activates the camera and prepares the communication session.

• Video Acquisition and Processing

The camera captures frames at a configurable resolution and frame rate. Instead of transmitting full-resolution video continuously, the system can use a lower-rate preview until a visitor event is confirmed. During an active session, a higher-quality stream may be enabled according to available bandwidth. Optional face detection is performed locally to identify a face region and, when authorized, compare it with enrolled reference embeddings.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

• Two-Way Audio

The audio subsystem uses a microphone for the entrance-side uplink and a speaker for the downlink. Full-duplex operation means that both directions remain active at the same time. Echo cancellation, automatic gain control and noise suppression can be included to improve speech quality. The design should avoid storing raw audio unless there is a clear user requirement and appropriate consent.

• IoT Notification

When a valid doorbell event occurs, the device publishes a compact event message containing a timestamp, device identifier and event type. The server or notification service forwards the event to the resident's application. A snapshot or short preview can be attached where privacy settings permit. The live video and audio session are opened only when the user chooses to respond.

IX. PROPOSED ALGORITHM

1 Initialize camera, audio interfaces, doorbell input and network connection.
2 Authenticate the device with the IoT service.
3 Enter standby mode and monitor the doorbell input.
4 When the button is pressed, debounce and validate the event.
5 Activate the camera and capture the visitor scene.
6 Optionally detect a face and compare it with authorized reference data.
7 Publish a visitor-event notification to the user's application.
8 If the resident accepts the event, establish live video and full-duplex audio.
9 Allow the resident to communicate with the visitor; do not unlock the door automatically based only on face recognition.
10 Terminate the session after user action or timeout, stop unnecessary streams and return to standby.
11 Record only minimum operational logs required for diagnostics and security auditing.

X. SECURITY AND PRIVACY DESIGN

Because a smart video doorbell operates at the boundary of a private residence, security is an essential design requirement. The proposed system uses device authentication, encrypted communication and role-based user access. Default passwords should be disabled, credentials should not be hard-coded into publicly shared source code, and firmware should be updateable. Video and audio should be transmitted only during required sessions or according to an explicit recording policy.

1. Use TLS-protected application communication wherever supported.
2. Use unique credentials or device certificates rather than shared default passwords.
3. Separate visitor recognition from the final door-access decision.
4. Store face templates only when necessary, with explicit enrollment and deletion controls.
5. Apply rate limits and lockouts to reduce unauthorized login attempts.
6. Keep audit logs minimal and avoid unnecessary storage of visitor images or audio.
7. Provide a physical or software mechanism to disable the camera/microphone when privacy is required.

XI. EVALUATION METHODOLOGY

A publication-quality implementation should be evaluated using reproducible experiments rather than unsupported performance claims. The following metrics are recommended. Actual measured values should be inserted after testing the prototype.

Metric	Measurement method	Target interpretation
Event-to-notification delay	Measure time from button activation to app notification	Lower is better

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, October 2026 Impact Factor: 8.2

Video latency	Timestamp a scene change and compare client display time	Suitable for conversational monitoring
Audio round-trip delay	Measure time between spoken signal and remote return	Lower delay improves conversation
Face verification accuracy	Evaluate known/unknown samples under varied conditions	Report precision/recall or accuracy
False acceptance rate	Test unauthorized faces against enrolled identities	Should be minimized
False rejection rate	Test enrolled users across pose/lighting conditions	Should be minimized
Availability	Record successful sessions over repeated trials	High service availability
Power consumption	Measure standby and active operating current	Event-driven design should reduce standby load

XII. TEST SCENARIOS

- Daylight, frontal visitor view.
- Low-light entrance condition.
- Side-facing and partially occluded visitor.
- Weak or temporarily unavailable Wi-Fi.
- Simultaneous speaking by resident and visitor.
- Repeated doorbell presses within a short interval.
- Unauthorized login attempt.
- Long-duration standby followed by an event.

XIII. RESULTS AND DISCUSSION

The proposed architecture is designed to be evaluated as a functional prototype rather than as a purely conceptual doorbell. A complete experimental section should report the measured values from the test scenarios in Section 8. The most important expected trade-offs are between video quality, processing load, network bandwidth and latency. Higher video resolution increases image detail but also increases bandwidth and may raise delay on constrained networks. Similarly, face recognition is affected by illumination, pose, occlusion and the quality of the enrollment images. These factors should therefore be reported separately rather than represented by a single accuracy number.

For audio, the principal usability factor is conversational delay. Echo and acoustic feedback are likely to occur when the speaker and microphone are physically close, so acoustic echo cancellation and careful gain control are important. The event-driven architecture can also reduce unnecessary network activity compared with a system that continuously uploads video. However, cloud dependence introduces a possible failure point; therefore, local doorbell feedback should remain available even when the Internet connection is unavailable.

XIV. ADVANTAGES AND LIMITATIONS

- Advantages

1. Remote visual awareness of visitors.
2. Natural two-way communication without opening the door.
3. Event-driven operation can reduce unnecessary bandwidth and energy use.
4. Optional local face verification can reduce dependence on cloud processing.
5. Modular design permits integration with other smart-home services.

- Limitations

1. Performance depends on Wi-Fi quality and Internet availability for remote access.
2. Face recognition can degrade under poor lighting, occlusion and unusual viewing angles.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

3. Outdoor audio can be affected by wind and environmental noise.
4. Camera and microphone data create privacy obligations.
5. Low-cost embedded hardware may require a separate gateway for computationally intensive recognition.

XV. FUTURE SCOPE

1. On-device neural face embedding optimized for low-power processors.
2. Local video storage with automatic deletion policies.
3. Wake-word or voice-assisted visitor interaction.
4. Integration with smart locks using multi-factor authorization.
5. Offline-first operation with local notifications when Internet access fails.
6. Tamper detection using enclosure sensors and event logging.
7. Adaptive video bitrate selection based on network conditions.

XVI. CONCLUSION

This paper proposed an IoT-based smart video doorbell that unifies entrance video monitoring, remote notifications and full-duplex audio communication in a modular smart-home security architecture. The design separates event detection, video acquisition, optional face verification, IoT messaging and user interaction so that each function can be optimized independently. Unlike a conventional doorbell, the proposed system allows a resident to see and speak with a visitor from a remote location while retaining human control over access decisions. Security and privacy are incorporated through authenticated communication, encryption, limited data retention and optional local processing. The design can be implemented using an ESP32-class camera device with an optional edge gateway for more demanding analysis. Future prototype testing should quantify notification delay, video latency, audio quality, recognition performance, power consumption and reliability under realistic lighting and network conditions. The resulting platform can serve as a foundation for broader smart-home entrance monitoring and secure visitor-management applications.

REFERENCES

1. A.Muni Swamy, Saraswathi, Srikanya, Snehitha, Sai Kumar Reddy, and Dastagiri, “Design and Implementation of an IoT-Enabled Smart Doorbell with Real-Time Face Recognition and Full-Duplex Audio Communication,” International Journal of Science, Engineering and Technology, 2026, doi:10.5281/zenodo.18668435.
2. M. Jagadeeswari Devi, M. DhanaLakshmi, and I. Veera Brahmendra, “Smart Home with Wireless Door bell by using IoT,” International Journal of Advanced Research in Science and Technology (IJARST), vol. 13, issue 3, pp. 408–419, doi:10.48047/IJARST/V13/I03/53.
3. S. Kadam and A. Khan, “IoT based smart doorbell with face recognition and remote alerts,” World Journal of Advanced Research and Reviews, vol. 27, no. 1, pp. 755–763, 2025, doi:10.30574/wjarr.2025.27.1.2580.
4. S. Surla, N. A. Manepalli, N. A. Shaik, and N. S. Gurram, “IoT and Face Recognition based Automated Door Lock System,” 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), pp. 648–651, 2023, doi:10.1109/ICEARS56392.2023.10085609.
5. Shaout and M. Theisen, “State of the Art - Smart Doorbell Systems,” 2021 22nd International Arab Conference on Information Technology (ACIT), pp. 1–8, 2021, doi:10.1109/ACIT53391.2021.9677313.
6. Antzoulis, M. M. Chowdhury, and S. Latiff, “IoT Security for Smart Home: Issues and Solutions,” 2022 IEEE International Conference on Electro Information Technology (eIT), pp. 1–7, 2022, doi:10.1109/eIT53891.2022.9813914.
7. G. Rodrigo and D. De Silva, “IoT-enabled Contactless Doorbell with Facial Recognition,” 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), pp. 1–6, 2023, doi:10.1109/ICITIIT57246.2023.10068625.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

8. M. Khan, H. Anum, S. S. Batool, and B. Bashir, "Smart Home with Wireless Smart Doorbell with Smart Response," 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 1–5, 2021, doi:10.1109/ICECCME52200.2021.9590865.
9. T. P. Rani, S. Sakthy, P. Kalaichelvi, T. Vignesh, and M. Priyadarshan, "Home Security and Anti-Theft System," 2022 1st International Conference on Computational Science and Technology (ICCST), pp. 130–133, 2022, doi:10.1109/ICCST55948.2022.10040423.
10. V. Valarmathi, T. Sathya, J. Nirmala, Sivarajeswari, R. M. Prasanth, and V. Srihari, "Design and Implementation of Secured Contactless Doorbell using IoT," 2022 International Conference on Computer, Power and Communications (ICCPC), pp. 187–191, 2022, doi:10.1109/ICCPC55978.2022.10072073