International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 11 | Nov 2026 www.irjet.net p-ISSN: 2395-0072

Alert System for CCTV Surveillance: Night-Time Human Anomaly Detection with Face-Whitelist Filtering and Identity-Based Alerting

Atmaram Subhash Toraskar[1], Snehal Naik[2], Akshata Godkar[3]

[123]*Department of Information Technology, Sindhudurg Sub Campus, University of Mumbai, Sawantwadi 416510, Maharashtra, India*

Abstract - *Conventional CCTV systems record continuously but rely on a human to notice unusual activity, so night-time intrusions are often discovered only when footage is reviewed long after the event. This paper presents a low-cost, privacy-aware alerting pipeline that converts recorded CCTV video into timely alerts. The system combines YOLOv8 person detection, ByteTrack multi-object tracking, a user-configurable time-window rule (daily and one-time windows set through a desktop GUI), a face-recognition whitelist that keeps enrolled household members silent, and a Telegram bot that sends exactly one photo alert per newly confirmed unknown person. Flagged incidents are grouped into timestamped events and shown on a browser-based timeline so a reviewer can jump straight to the relevant moments. We report incremental, evidence-backed tests: doubling the ByteTrack track buffer removed an identity-fragmentation fault (2 IDs to 1 for one person); CLAHE enhancement gave no measurable detection gain (91.9% vs 91.7%); identity-based deduplication replaced a cooldown design that produced over 30 duplicate alerts; and in a combined clip an enrolled person stayed silent while an unknown person triggered one alert. We also isolate a limitation, detector failure on dark clothing against dark backgrounds, and show it is not a tracker problem.*

Index Terms - CCTV surveillance, anomaly detection, YOLOv8, ByteTrack, face recognition, real-time alerting.

I. INTRODUCTION

CCTV cameras are widely installed in homes and shops, yet most installations are passive: video is stored and examined only after an incident is reported. During the night nobody watches the live feed, so an intruder can be recorded clearly and still go unnoticed until the opportunity to respond has passed. This gap between "the camera saw it" and "someone noticed in time" motivates this work, which was inspired by a local theft in which the suspects were captured on camera but the footage was not reviewed in time.

Motion-based alarms are a common remedy but generate false alerts for residents moving around their own premises, and fixed quiet-hour rules are impractical for shop owners whose "unusual" periods change (for example, stepping out for two hours). We therefore target four requirements: (1) detect and persistently track people; (2) let a non-technical user define when presence is unusual; (3) suppress alerts for known household members; and (4) notify the owner once per new person and keep a log that shortens later review.

Our contributions are an integrated, locally runnable pipeline meeting these requirements; empirical evidence for each design decision, including negative results; and a root-cause analysis of a low-contrast failure case.

II. RELATED WORK

Single-stage detectors of the YOLO family [1], [2] reformulate detection as one regression pass and suit real-time surveillance. Detection must be linked across frames: SORT [3] uses Kalman filtering and Hungarian matching, and ByteTrack [4] additionally associates low-confidence boxes, helping tracks survive brief occlusion. Contrast Limited Adaptive Histogram Equalization (CLAHE) [5] is a standard low-light pre-processing step. For identity filtering, deep metric learning maps a face to a compact embedding (128-d in dlib [6]) so that recognising a new household member needs only a distance comparison. Prior work usually studies these parts in isolation; we combine them with a configurable schedule and an alert policy, and report where the combination fails.

III. SYSTEM DESIGN

Fig. 1 shows the pipeline. Each stage is an independently testable module and the whole system runs offline on recorded video, so test footage never leaves the local machine.

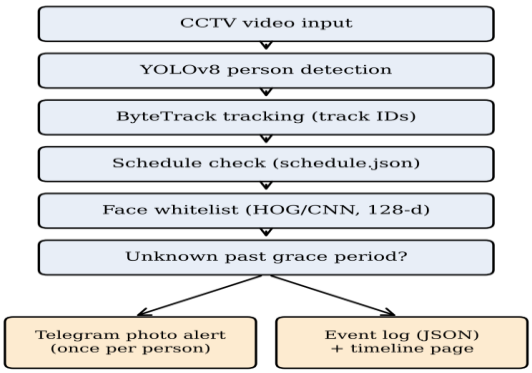

Fig. 1. Proposed alerting pipeline.

A. Detection and tracking

YOLOv8n detects the person class (confidence ≥ 0.4). ByteTrack, through a custom configuration, assigns persistent IDs; the key change is track_buffer = 60 (default 30) so a person who is briefly undetected, e.g. in a doorway, keeps the same ID.

B. Schedule-driven anomaly rule

A Tkinter GUI writes schedule.json containing recurring daily windows (e.g. 23:00–06:00, with midnight wrap-around) and one-time "away" windows with a date and time range. The pipeline reads this file at run time, so the owner changes behaviour without editing code. A person is only "unusual" inside an active window.

C. Face whitelist with grace period

Enrolled faces are stored as 128-d embeddings. For each still-unknown track, every 10th frame the head region (top 35% of the box) is enlarged 3× and passed to a HOG face detector, falling back to a CNN detector; a match requires Euclidean distance ≤ 0.5. A recognised track ID is cached ("sticky") so the label survives head turns. A grace period of 75 frames (about 3 s) lets recognition run before a track may be treated as a confirmed unknown.

D. Identity-based alerting and logging

A confirmed unknown track inside an active window triggers a Telegram Bot API photo message with timestamp, and its ID is added to an alerted set so each person alerts at most once. Per-frame flags within 2 s of each other are merged into one event (start, end, duration, IDs) written to a JSON log. A self-contained HTML page loads the video and log locally, draws clickable markers on a timeline, and reports the share of footage flagged.

TABLE V. KEY PIPELINE PARAMETERS

Parameter	Value
Detector / class	YOLOv8n / person
Confidence threshold	0.4
track_buffer	60 (default 30)
Face match tolerance	0.5
Face check interval	every 10 frames
Grace period	75 frames (~3 s)
Event merge gap	2.0 s
Default quiet window	23:00–06:00

Each logged event stores an ID, start and end time, duration, number of unknown people and the track IDs involved, which is what the timeline page reads. All thresholds were set empirically from the tests in Section IV rather than learned.

IV. EXPERIMENTS AND RESULTS

Experiments were incremental, each phase tested before integration, on a consumer Windows machine using Python, Ultralytics, OpenCV and face_recognition. Test material was a short locally recorded household clip (processed only locally) plus publicly available stock night-vision clips. The study is small-scale and preliminary; figures below are from these clips, not a benchmark.

A. Low-light enhancement

TABLE I. CLAHE VS RAW DETECTION (1,535 FRAMES)

Input	Frames with detection	Rate
Raw	1,410	91.9%
CLAHE	1,408	91.7%

The 0.2-point difference is within frame-to-frame variation, so CLAHE brought no measurable benefit and is not required by the final pipeline.

B. Tracking stability

TABLE II. IDENTITY FRAGMENTATION FIX

Setting	IDs for one person
Default (buffer 30)	2 (frames 48–243, 256–265)
Tuned (buffer 60)	1 (frames 42–265)

C. Face whitelist

TABLE III. ENROLMENT AND MATCHING

Test	Outcome
Close-up phone photos	2/7 then 3/7 faces found
CCTV-scale screenshots (CNN)	4/4 faces found
Enrolled person in CCTV clip	Matched, distance 0.42
Unenrolled person, 19 checks	Unknown, closest 0.56

Enrolling from CCTV-scale images worked far better than close-ups, indicating a scale and angle mismatch. The 0.42 versus 0.56 separation around the 0.5 threshold is workable but thin and needs more subjects.

D. Alert policy

A first design used a 30 s cooldown and sent over 30 duplicate photos for one continuing presence. Identity-

based deduplication fixed this: one alert per unique track ID. Face recognition and alerting were then tested together.

TABLE IV. INTEGRATED PIPELINE TESTS

Clip	Result
Enrolled person only	Recognised; no alert; 0 events
Unknown person only	1 alert after ~3 s; event logged
Both in one scene	Enrolled silent; unknown: 1 alert

The combined clip directly exercises the motivating requirement: silent for the household, alert for a stranger beside them. On a separate shop clip the timeline page placed the marker correctly and reported 1 event covering 4.0 s of 0:06.

E. Failure case: dark clothing

On a stock clip with two people in all-black clothing in a dark room, the tracker produced six IDs. Lowering the confidence threshold to 0.25 and raising track_buffer to 120 did not help. Frame inspection showed YOLOv8 emitted no box in several frames despite a visible person, so the cause is detector sensitivity to low-contrast silhouettes, not tracking.

V. PRIVACY AND ETHICS

Surveillance analytics process biometric data, so the design is local-first: videos, face encodings and snapshots stay on the user's machine; the only outbound traffic is the alert message to the owner's own Telegram chat; bot credentials are held in an environment file and never in source code. Only consenting household members are enrolled, and the whitelist reduces, rather than increases, the number of people who are photographed and reported. Testing used the authors' own recordings and public stock footage; footage of real incidents or third parties was not used.

VI. DISCUSSION AND LIMITATIONS

Three design lessons emerge. First, tracker defaults need tuning for occlusion-heavy scenes. Second, alert deduplication should follow identity rather than time, since a cooldown either spams or risks suppressing a short appearance. Third, face enrolment must match deployment scale. Limitations: evaluation uses few clips and subjects and reports no precision/recall or latency; processing is on recorded files, not a live stream; the detector fails on dark-on-dark scenes; the schedule file must exist before a run; and loitering and group-size rules are not implemented.

VII. CONCLUSION

We presented an integrated, free-tier pipeline that turns passive CCTV footage into per-person alerts that respect a user-defined schedule and a household whitelist, with an event log that cuts review time. Future work includes live RTSP input, a night-tuned or thermal detector, larger multi-subject evaluation with precision, recall and false-alert rates, and behavioural rules such as loitering.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE CVPR, 2016.
[2] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLOv8,” 2023. [Online]. Available: https://docs.ultralytics.com
[3] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” in Proc. IEEE ICIP, 2016.
[4] Y. Zhang et al., “ByteTrack: Multi-object tracking by associating every detection box,” in Proc. ECCV, 2022.
[5] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in Graphics Gems IV, Academic Press, 1994.
[6] D. E. King, “Dlib-ml: A machine learning toolkit,” J. Mach. Learn. Res., vol. 10, pp. 1755–1758, 2009.