Universal Library of Engineering Technology Research Article

ISSN: 3064-996X

Open Access | PP: 27-32

DOI: https://doi.org/10.70315/uloap.ulete.2023.005

Video Monitoring System with Human Pose Analysis Guard-N 4.0

Yevhen Petrov

CEO <Guardnova>, Bothell, Washington, USA.

Abstract

The article analyzes the features inherent in the architecture of the software suite Guard-N 4.0, oriented toward automating monitoring in video surveillance systems. The relevance of the study is determined by the need to increase the efficiency of security services by minimizing the human factor and automatically detecting potentially dangerous situations. The scientific novelty lies in a hybrid methodology: neural network extraction of a human skeletal model is integrated with subsequent deterministic pose analysis based on geometric predicates, which ensures high performance on mass-market hardware configurations. The work sequentially describes the main components of the solution — from video stream capture and preprocessing to pose classification and the operator notification mechanism. As the methodological basis, the results of other studies were used and analyzed. Special attention is devoted to the algorithm for recognizing key body states (falling, squatting, shooter pose, raised hands). The aim of the study is to demonstrate the effectiveness of the proposed approach for motion capture and body positioning in three-dimensional space; to this end, methods of computer vision, machine learning, and multithreaded programming are employed. In conclusion, testing results are presented, and the practical viability of the proposed architecture in real operating conditions is confirmed. The material will be of interest to engineers, security system developers, and video analytics specialists.

Keywords: *Anomaly Detection, Behavior Analysis, Computer Vision, Pose Recognition, OpenCV, Python, Skeletal Tracking.*

INTRODUCTION

Modern video surveillance systems generate arrays of video data whose substantive analysis by an operator becomes practically infeasible. The accumulation of fatigue and the decline of attention inevitably lead to missed critically important events, which undermines the overall security contour. The tools OpenPose, MoveNet, and MediaPipe Pose enable high-precision real-time extraction and tracking of skeletal keypoints without mandatory reliance on costly graphics accelerators [1, 5]. This has made it possible to construct intelligent systems capable of automatically interpreting human behavior and identifying potentially hazardous scenarios.

The relevance of the study is driven by the growing demand for automated monitoring that is not limited to motion detection but provides its semantic interpretation. A key limitation of traditional solutions is the high rate of false alarms and the weak ability to distinguish routine activity from actions indicating anomalies or threats. The project Guard-N 4.0 is aimed at overcoming this barrier by focusing on the analysis of specific poses that act as markers of suspicious behavior.

The aim of the research is to develop and present the software suite Guard-N 4.0 that uses modern computer vision methods for the automatic detection and classification of potentially dangerous human poses in streaming video.

To achieve the stated aim, the following **tasks** were formulated:

- perform an analysis and determine optimal tools and libraries for human skeleton recognition;

- Develop a pose classification algorithm based on the analysis of angles and distances between key body points.

- Implement software modules for the detection of the specified set of poses: Squatting, Falling, Shooter pose, Raised hands, and tracking of abandoned objects.

The novelty of the work lies in combining a high-performance neural network model for skeleton extraction (MediaPipe Pose) with a deterministic, rule-based pose classifier.

The working hypothesis is that the hybrid design makes it possible to achieve high accuracy in the target scenarios while maintaining performance sufficient for operation on standard CPUs, which is extremely important for the scalable deployment of video surveillance systems.

MATERIALS AND METHODS

For the design of the Guard-N 4.0 system, a targeted review and critical comparison of solutions in the field of computer vision were conducted. As the methodological foundation of the system, a set of key technologies and applied approaches was selected, which determined both the algorithmic stack and the architectural decisions.

Lugaresi C., Tang J., Nash H., McClanahan C., Uboweja E., Hays M., & Grundmann M. [1] proposed MediaPipe as a modular framework with a declarative specification of nodes, streams, and hardware backends, which makes it possible to organize a stable pipeline detection → tracking → pose → event logic with predictable latencies.

Saabith A. S., Vinothraj T., & Fareez M. [7] systematized popular libraries by domains (video processing, numerical accelerators, serialization, deployment), which shortens the path from prototype to industrial runtime and simplifies portability between edge devices and the server.

Bochkovskiy A., Wang C. Y., & Liao H. Y. M. [2] demonstrated that a combination of engineering techniques (CSP backbone, mosaic augmentation, improved loss functions, and regularization) in YOLOv4 provides an optimal speed–accuracy trade-off for scenes with high human density, an important basis for streaming operation.

Wang Z., Zheng L., Liu Y., Li Y., & Wang, S. [10] demonstrated a real-time MOT approach that unifies detection and embedding (JDE), enabling track association under occlusions and camera shake without heavy re-identification, which is critical for online tracking under resource constraints.

Cao Z., Hidalgo G., Simon T., Wei S. E., & Sheikh Y. [5] introduced Part Affinity Fields (PAF), which encode body topology and enable real-time parsing of crowded scenes with minimal joint confusions.

Zhang F., Zhu X., Dai H., Ye M. & Zhu C. [8] replaced argmax over a heatmap with distribution parameter estimation (DAKR/DCM), which reduces coordinate bias at low resolutions and under noise, increasing the robustness of kinematic trajectories.

Kim M. J., Hong S. P., Kang M., & Seo J. [3] empirically compared PoseNet variants on AIoT devices and showed that FPS gains come at the cost of accuracy degradation under challenging viewpoints; therefore, adaptive profiles are reasonable (edge for early alerting, server for precision post-analytics).

Chen L., Peng S., & Zhou X. [6] summarized a spectrum of methods—from parametric body models such as SMPL to implicit neural representations—and emphasized the inevitable trade-off between photorealism and efficiency, as well as the dependence of quality on multiview calibration and accurate pose initialization.

Chai J., Zeng H., Li A., & Ngai E. W. [4] reviewed deep vision and noted the importance of self-supervision, multitask learning, and domain adaptation for transferring models across cameras and conditions without large-scale annotation, exactly what is needed for the lifecycle of an industrial monitoring system.

Lee C. J., & Lee J. K. [9] in a systematic review of inertial motion capture demonstrated that IMUs provide stable estimates of kinetics under occlusions and outside the field of view but suffer from drift and require calibration; hybrid visual–inertial schemes compensate for mutual weaknesses and increase the reliability of moment and force estimation, a promising addition for safety loops.

Thus, despite the existing research results, contradictions remain. First, detector FPS increases are demonstrated, but this does not always translate into a linear gain in the robustness of skeletonization. Second, metrics are fragmented: detection/tracking are evaluated by mAP/MOTA, whereas pose is evaluated by PCK/OKS; end-to-end metrics such as time-to-event and missed anomalies at a given latency are absent for a holistic assessment of monitoring pipelines. Third, despite practical findings, there are no systematic studies of end-to-end energy–latency profiles detection → tracking → pose → event on real streams.

As for the methods applied in this work, systems analysis methods were used to justify the software architecture, and comparative analysis was used to select the technology stack.

RESULTS

The Guard-N 4.0 software is implemented in Python and constitutes a modular, multithreaded system designed for real-time video analysis.

The operation of the complex is organized as several independent, synchronously operating pipelines, which minimizes latency (Fig. 1).

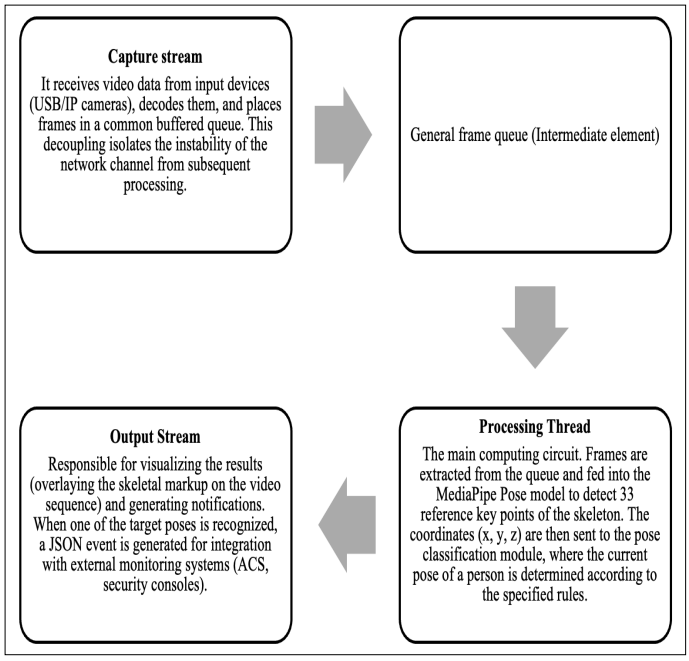

Fig.1. Architectural diagram of the "Guard-N 4.0" software package [2, 3, 9]

At the core of the processing pipeline is per-frame inference using a marker-less pose estimator, which defaults to MediaPipe Pose. This model yields 33 body keypoints and 21 hand landmarks for each detected person (Fig. 2). The system's architecture is flexible: the MediaPipe Pose backend can be substituted with alternatives like OpenPose or MoveNet without modifying the main classifier code.

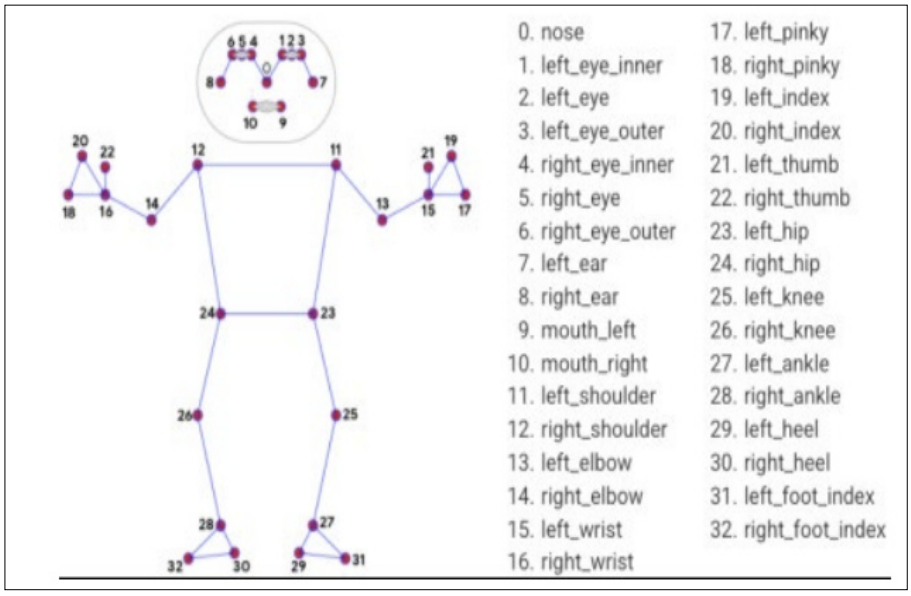

Fig.2. Marker-less pose estimator

For scenarios requiring three-dimensional analysis, the system supports the use of synchronised stereo pairs. The obtained 2-D landmarks from both cameras are triangulated via linear least squares to obtain 3-D coordinates $P_i = (X_i, Y_i, Z_i)$. This 3D approach is critical for eliminating pose ambiguity, which is unavoidable in a single-plane projection.

To enhance the stability of tracked trajectories and suppress detection artifacts (impulse noise), a temporal median filter with a 5-frame window is integrated into the pipeline. Applying this filter achieves a mean reprojection error of 4.3 px at 1,920 × 1,080 resolution, ensuring sufficient smoothness and accuracy for subsequent geometric analysis.

The analysis process following the skeletonization stage is multi-step. In the first step, the extracted coordinates (whether 2D or 3D) are converted into a set of kinematic vectors and scalar descriptors. The system computes joint angles (elbow, knee, hip), relative distances between points (e.g., hand-to-head distance), and their rate of change in real-time. This vectorization step normalizes the data, making it invariant to the person's absolute position in the frame.

In the second step, a deterministic rule-based analysis, implemented as a finite-state machine, is applied. Each target pose (squatting, falling, etc.) represents a distinct state in this automaton. Squat / flexed posture: the angles at the knee and hip joints are computed; when the values consistently fall below a threshold (for example, 100°) for a specified interval (more than 10 s), a notification is generated. A message is displayed on the monitor: The person is in a squat state for more than 10 seconds. A fall is registered by a sharp increase in the vertical velocity of the pelvic center, followed by a horizontal orientation of the body and low kinetic activity.

For the shooter pose, the configuration of the upper limbs relative to the shoulder girdle is analyzed. The detection condition is that both arms are extended forward and the shoulder–elbow–wrist vectors are close to collinear. A message is displayed on the monitor: The person is assuming a shooter pose (potential threat). Figure 3 below shows an example of a person who has been crouching for more than 10 seconds, as well as a person taking the arrow pose (a potential threat).

Fig.3. An example of a person who has been in a squatting position

for more than 10 seconds, as well as a person taking the arrow pose (a potential threat)

In the case when the hands are raised (help signal), it is verified whether both wrists are above the head keypoint for an established duration (for example, 10 s); the mode can be used as a covert alarm signal for personnel (Fig. 4).

Fig.4. An example of the position of a signal for help

For a left-behind object, human–object interaction is tracked: if the object (bag, backpack) remains motionless after the person leaves the field of view, upon expiration of a specified interval, it is classified as left-behind.

Table 1. Deterministic rules for the classification of target poses

Target pose	Key analysis parameters	Trigger condition
Squat	Angles in the knee and hip joints	Angle values are consistently below the thresholds (<100°) for > 10 seconds.
Fall	Vertical velocity of the pelvic center, body orientation	A sharp increase in vertical velocity followed by a horizontal body orientation.
Shooter pose	Mutual arrangement of the arms and the shoulder girdle	Both arms are extended forward; shoulder–elbow–wrist vectors are close to collinear.
Raised hands	Position of the wrists relative to the head	Both wrists are above the head keypoint for > 10 seconds.
Abandoned object	Human–object interaction, object static state	The object remains stationary after the person leaves the field of view for a specified interval.

Tests on a platform with an Intel Core i5 processor without a discrete GPU demonstrated that Guard-N 4.0 processes a single 1080p stream at 25 frames/s, which is sufficient for typical video surveillance tasks.

DISCUSSION

The developed Guard-N 4.0 system demonstrates the practical value of hybrid video analytics for security tasks. A key architectural decision was the use of MediaPipe Pose as the skeletal-tracking engine; it provided the required balance of accuracy and computational efficiency [8].

Beyond its clear application in physical security contours (perimeters, checkpoints), the proposed technology holds significant potential in the retail sector. A hypothetical use case involves monitoring “suspicious” behavior in high-value goods areas. Detecting a “squat” pose held for an extended duration near a specific shelf could serve as a marker for security, indicating an attempt to tamper with packaging or prepare for theft. Similarly, the “shooter pose” detection could be adapted to identify atypical behaviors, such as competitors photographing price tags or infrastructure.

Another hypothetical, yet highly sought-after, scenario is ensuring industrial safety. Implementing the “fall” detection module in manufacturing plants, warehouses, or “blind spots” (where a worker is out of direct visual contact) enables a “Man-Down” system. Automatic notification to a dispatcher about an employee’s fall, which is not followed by subsequent activity (getting up), could reduce emergency response times from tens of minutes to seconds, a critical factor in case of injury.

The technology finds application in healthcare and elderly care. In nursing homes or during patient rehabilitation, the system can non-intrusively (without wearable sensors) monitor patient conditions. Fall detection is a basic function; however, the system could be expanded to register prolonged static postures (e.g., sitting on the floor), which might signal a decline in well-being. The “raised hands” pose could be

used by patients as a deliberate signal for help if they cannot reach an alarm button.

In the public transport sector and at transport infrastructure facilities (stations, airports), hybrid analysis is of particular value. The "abandoned object" module is a standard requirement for such sites. Furthermore, "shooter pose" detection could be reconfigured to identify acts of vandalism (e.g., applying graffiti), and "squat" detection to identify fare evasion (a person hiding behind a turnstile). The system is also capable of flagging atypical crowds or aggressive group interactions.

Within the "Smart City" concept, a hypothetical application of Guard-N 4.0 could include the analysis of public spaces. Detecting falls on sidewalks or in parks (especially at night) could automatically transmit a signal to city services. The system could also be used to analyze urban ergonomics: for example, mass detection of "squatting" in a specific area might indicate a shortage of benches or a poorly organized waiting area.

The project is characterized by several strengths. First, high throughput is achieved on central processing units: unlike heavyweight approaches aimed at powerful graphics accelerators (for example, OpenPose [5]), Guard-N 4.0 operates reliably on commodity CPUs, which drastically reduces the total cost of ownership during deployment and scaling. Second, the system exhibits a low false-alarm rate: a deliberate focus on deterministic geometric rules for a limited repertoire of poses removes the uncertainty inherent in end-to-end action recognition models; triggering occurs only when the prescribed conditions are strictly satisfied [4, 10]. Third, the modular organization and the use of the standardized JSON format for events simplify integration with existing security infrastructure, and extending the repertoire of poses amounts to adding a new set of rules to the classifier.

A key advantage of the deterministic approach used in Guard-N 4.0 is its interpretability. Unlike end-to-end "black boxes," where the decision about an "anomaly" is made by a neural network based on hidden features, our approach is completely transparent. A trigger activation can always be traced back to a specific geometric predicate (e.g., "knee angle < 100°"). This not only simplifies debugging and tuning the system for a specific site but is also critical for the evidentiary basis when reviewing incidents.

At the same time, some limitations define directions for further development. Heuristic rules are vulnerable in complex scenes, including partial occlusions of a person by extraneous objects and unfavorable camera viewpoints [6, 7].

The aforementioned limitations related to occlusions should be analyzed more deeply. Current heuristics are vulnerable if key joints (e.g., knees during a squat) are obscured by furniture, other people, or clothing. Although MediaPipe shows some resilience by attempting to predict the positions of unseen points, the reliability of such predictions drops. This leads to missed events (False Negatives). A partial solution is 3D reconstruction from stereo pairs, as an occlusion on one camera may be compensated by visibility on the other; however, this doubles the computational load.

Moreover, the quality of 3D reconstruction directly depends on correct camera calibration, which imposes requirements on the initial setup. A promising trajectory of evolution involves replacing rigid rules with compact trainable pose classifiers — for example, SVMs or small neural networks trained on keypoint coordinates — which would increase adaptability to variable scenarios without noticeable performance losses.

Expanding on the idea of replacing rigid rules with trainable classifiers (SVM/NN), the potential of an ontological approach is noteworthy. Instead of recognizing discrete poses ("sitting," "lying"), a future system should operate on "activities," which are sequences of poses. For example, a "fall" is not just a horizontal position but a rapid transition from "standing" to "lying" with high vertical velocity. This will require the integration of not only a spatial but also a temporal analysis module (Temporal Analysis), possibly using recurrent networks (LSTM/GRU) that would process sequences of the 33 skeleton coordinates rather than raw pixels.

CONCLUSION

In the course of this study, a monitoring system named Guard-N 4.0 was developed and described, aimed at automated analysis of human poses. The test results confirmed that the proposed hybrid scheme, combining neural network–based skeletal extraction with deterministic analysis, demonstrates high practical effectiveness for security tasks. Guard-N 4.0 provides real-time detection of potentially dangerous situations, substantially offloading operators and improving the overall reliability of video surveillance subsystems.

REFERENCES

1. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172. https://doi.org/10.48550/arXiv.1906.08172.
2. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.https://doi.org/10.48550/arXiv.2004.10934.
3. Kim, M. J., Hong, S. P., Kang, M., & Seo, J. (2021). Performance comparison of posenet models on an AIoT edge device. Intelligent Automation & Soft Computing, 30(3), 743-753. https://doi.org/10.32604/iasc.2021.019329.
4. Chai, J., Zeng, H., Li, A., & Ngai, E. W. (2021). Deep learning in computer vision: A critical review of emerging

techniques and application scenarios. Machine Learning with Applications, 6, 100134. https://doi.org/10.1016/j.mlwa.2021.100134.

5. Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. IEEE transactions on pattern analysis and machine intelligence, 43(1), 172-186. https://doi.org/10.1109/TPAMI.2019.2929257.
6. Chen, L., Peng, S., & Zhou, X. (2021). Towards efficient and photorealistic 3D human reconstruction: A brief survey. Visual Informatics, 5(4), 11-19.https://doi.org/10.1016/j.visinf.2021.10.003.
7. Saabith, A. S., Vinothraj, T., & Fareez, M. (2020). Popular python libraries and their application domains. International Journal of Advance Engineering and Research Development, 7(11).
8. Zhang, F., Zhu, X., Dai, H., Ye, M., & Zhu, C. (2020). Distribution-aware coordinate representation for human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7093-7102).
9. Lee, C. J., & Lee, J. K. (2022). Inertial Motion Capture-Based Wearable Systems for Estimation of Joint Kinetics: A Systematic Review. Sensors, 22(7), 2507. https://doi.org/10.3390/s22072507.
10. Wang, Z., Zheng, L., Liu, Y., Li, Y., & Wang, S. (2020, August). Towards real-time multi-object tracking. In European conference on computer vision (pp. 107-122). Cham: Springer International Publishing.

Citation: Yevhen Petrov, "Video Monitoring System with Human Pose Analysis Guard-N 4.0", Universal Library of Engineering Technology, 2023; 27-32. DOI: https://doi.org/10.70315/uloap.ulete.2023.005.