ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1351

DetSPose: Detection of suspicious behavior by analyzing changes in people's body posture

Jorge L. Charco, Dr.[1,2], Francisco Alvarez-Solis, Msc.[1], Angela Yanza-Montalván, Dra.[1],
Johanna Zumba-Gamboa, Msc.[1], Jose Franco-Vera, Ing.[1], Lesley Caicedo-Guamán, Ing.[1]
[1]Universidad de Guayaquil, Ecuador
[2]Artificial Intelligence Research Group, Universidad Bolivariana del Ecuador, Campus Durán, Ecuador
jorge.charcoa@ug.edu.ec, docente.software1@ug.edu.ec, angela.yanzam@ug.edu.ec,
johanna.zumbag@ug.edu.ec, jose.francove@ug.edu.ec, lesley.caicedog@ug.edu.ec

Abstract*—This research work develops an automatic system for the detection of suspicious attitudes through computational analysis of body postures, using the MediaPipe library to extract and process the anatomical landmarks of the human skeleton. The study starts from the fundamental premise that certain postural patterns, movement sequences, and biomechanical configurations can reveal potentially hostile intentions before they materialize into explicit actions, all with the aim of complementing current surveillance systems. The proposed system uses the MediaPipe algorithm for obtaining human skeletal landmarks in both two- and three-dimensional spaces, considering three experimental scenarios: the detection of a hidden hand behind the body, the detection of a hand hidden under clothing at the front of the torso, and the recognition of suspicious gaze patterns. The system is evaluated using accuracy, precision, recall, and F1 score metrics. The results demonstrate high performance in all metrics and scenarios using three-dimensional spaces (95.6%, 95.7% and 97.3% respectively for accuracy metric). It suggests that the proposed system offers significant potential for applications in real security environments.

***Keywords*—human pose estimation, suspicious attitudes, behavioral analysis, flashing, classification of abnormal movements**

I. Introduction

In recent years, public safety has become one of the most pressing concerns for modern societies, leading to the widespread deployment of surveillance systems in public and private spaces [1]. Traditional surveillance systems, while widespread, often rely heavily on human operators whose performance may be affected by fatigue, subjectivity, or limited attention span [2]. Most existing automated security solutions focus primarily on facial recognition or object detection, leaving an important but less explored dimension underutilized: the analysis of human body posture.

Body language is a powerful form of non-verbal communication that can reveal emotions, intentions, and states of mind before they are translated into explicit actions [3]. Subtle postural variations, such as sudden movements, concealed hands, or evasive gestures, may serve as early indicators of suspicious or potentially aggressive behavior [4], [5]. Detecting these signs in real time could significantly enhance preventive security and improve response times in critical situations.

This research addresses the development of a module for detecting suspicious attitudes through computational analysis of body posture. Using tools from computer vision and artificial intelligence, particularly MediaPipe for extracting skeletal landmarks, the system aims to recognize patterns that may indicate hostile intentions [6]. By integrating machine learning techniques with conventional surveillance cameras [7]–[10]. This approach provides a low-cost and efficient alternative to complement existing monitoring methods.

Beyond technical contributions, the study also highlights the social and ethical implications of implementing such systems. While posture-based detection can strengthen security, it raises important concerns regarding privacy, data usage, and public acceptance [11], [12]. For this reason, any technological solution must be accompanied by transparent governance and strict ethical guidelines.

Fig. 1. Example of suspicious posture detected using MediaPipe.

II. Material and methods

A. Dataset and Input Sources

For the experimental stage, we worked with video recordings and real-time captures using standard web cameras, a common approach in vision-based surveillance research due to its low cost and ease of deployment [1]. The focus was on three specific suspicious behaviors that can be detected through posture and visual cues:

- **Hidden hand behind the body:** evaluated in both 2D and 3D perspectives.
- **Hidden hand under clothing (front):** also tested in 2D and 3D dimensions to analyze occlusion challenges.
- **Suspicious gaze:** analyzed as an additional behavioral indicator complementing postural cues.

These scenarios were chosen because they are common in situations of potential threat or concealment of objects and are difficult to detect with traditional surveillance systems [2], [4].

B. Pose Estimation Algorithm Used

The core of the proposed system is the MediaPipe Pose algorithm [13], which was selected due to its ability to perform real-time estimation of human skeletal landmarks in both two-dimensional and three-dimensional spaces [6]. Unlike traditional surveillance methods that rely solely on visual observation, this algorithm enables the extraction of anatomical key points such as shoulders, elbows, wrists, hips, and facial orientation, allowing posture inference through computational analysis [14]–[17]. By interpreting these landmarks, it becomes possible to infer human postures and identify anomalies that may indicate suspicious behavior.

In this research, the algorithm was applied specifically to three experimental scenarios: the detection of a hidden hand behind the body, the detection of a hand hidden under clothing at the front of the torso, and the recognition of suspicious gaze patterns. These cases were chosen because they represent common behaviors in situations where individuals may attempt to conceal objects or display attitudes of potential aggression. MediaPipe provided a practical advantage in these experiments, as its landmark models are robust to partial occlusions and allow analysis of subtle variations between normal and anomalous postures.

The estimation process works by capturing video frames through OpenCV [18] and passing them to MediaPipe, which returns the coordinates of the detected landmarks [19]. These coordinates are then analyzed to determine whether the observed body position corresponds to one of the defined suspicious behaviors. For example, the consistent absence of wrist landmarks when positioned behind the torso, or the lack of visibility of hand points in the frontal region, were interpreted as indicators of hidden hands. Similarly, the relative orientation of head landmarks with respect to the torso was used to characterize suspicious gaze, as supported by previous posture and behavior analysis studies [5].

The strength of this algorithm lies in its efficiency and adaptability, as it can operate on conventional hardware without requiring high-performance equipment, making it suitable for integration into standard surveillance infrastructures [6]. However, its limitations must be acknowledged, since accuracy can be affected by lighting conditions, overlapping individuals, or complex environments [20]. Nevertheless, for the experimental scenarios defined in this study, MediaPipe Pose provided reliable results that support the feasibility of using pose estimation as a complementary tool in automated security systems.

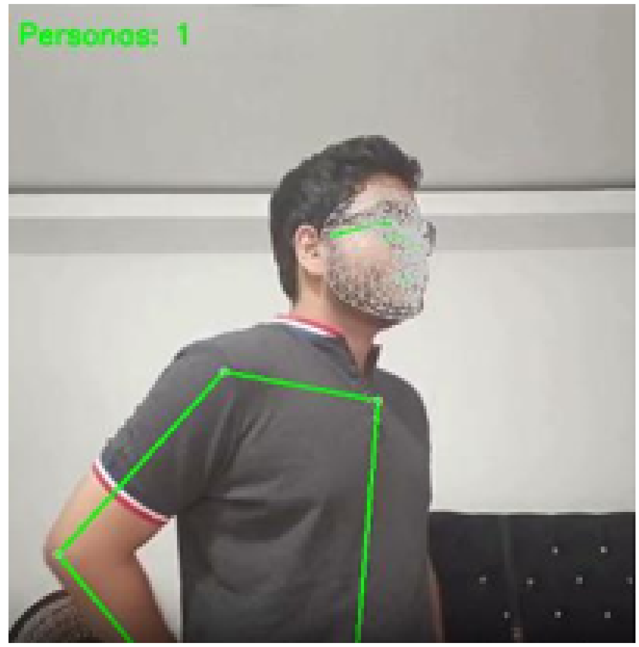

Fig. 2. Hidden hand behind

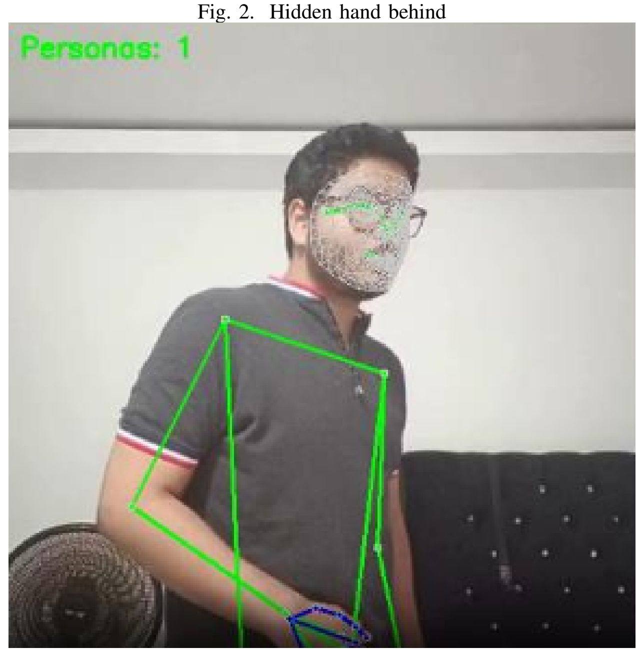

Fig. 3. Hand hidden under clothing

C. Interactive Module

An essential component of the system is the interactive module, which was designed to facilitate user interaction with the detection algorithm in a simple and intuitive manner. Interactive visualization tools have been shown to improve usability and interpretability in computer vision-based surveil-

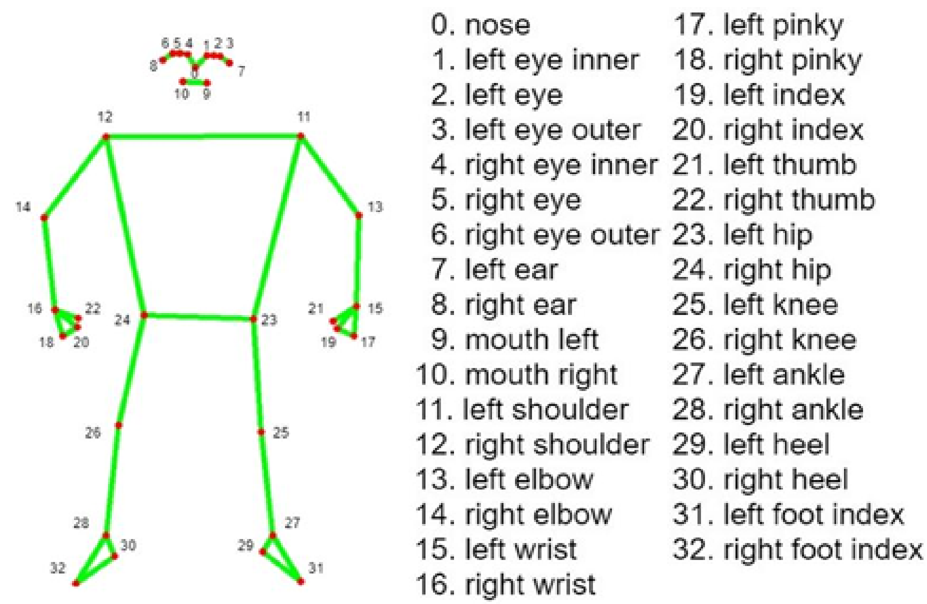

Fig. 4. Key points of the human body detected by Mediapipe

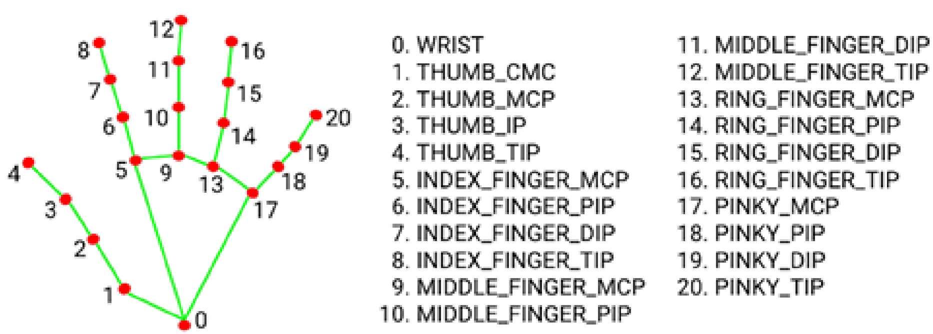

Fig. 5. Key points of the hand detected by Mediapipe

lance systems [21]. This module provides two main views. The first view allows the user to upload a pre-recorded video file, which is then processed by the system to detect suspicious postures within the selected footage. The second view enables the use of a live camera feed, allowing real-time monitoring and analysis of human behavior through posture estimation.

In both views, two configurable options were incorporated to give the user control over the type and depth of analysis. The first option allows selecting the dimension of analysis, either two-dimensional (2D) or three-dimensional (3D). While 2D analysis relies on pixel-based landmark detection, 3D analysis incorporates normalized depth information, which has been shown to improve discrimination in occlusion-prone scenarios such as hidden hands or frontal concealment [14], [20].

The second option defines the frame-skipping parameter, with values ranging from one to five. This strategy is commonly employed in real-time vision systems to balance computational efficiency and detection accuracy [19]. Lower values provide more continuous analysis, whereas higher values reduce computational load, making the system suitable for real-time applications on limited hardware.

Through this interactive module, the system not only performs the technical task of detecting suspicious behaviors but also allows adaptation to different operational constraints, whether for forensic analysis of recorded material or preventive monitoring in live environments [1]. The module serves as the practical bridge between the algorithmic core and its real-world application.

III. Experimental Results

A. Dataset and experimental setup

The experiments used a controlled dataset constructed from three video recordings specifically designed to represent three suspicious postural behaviors: *excessive gaze (suspicious look)*, *hand hidden behind the body*, and *hand under clothing*.

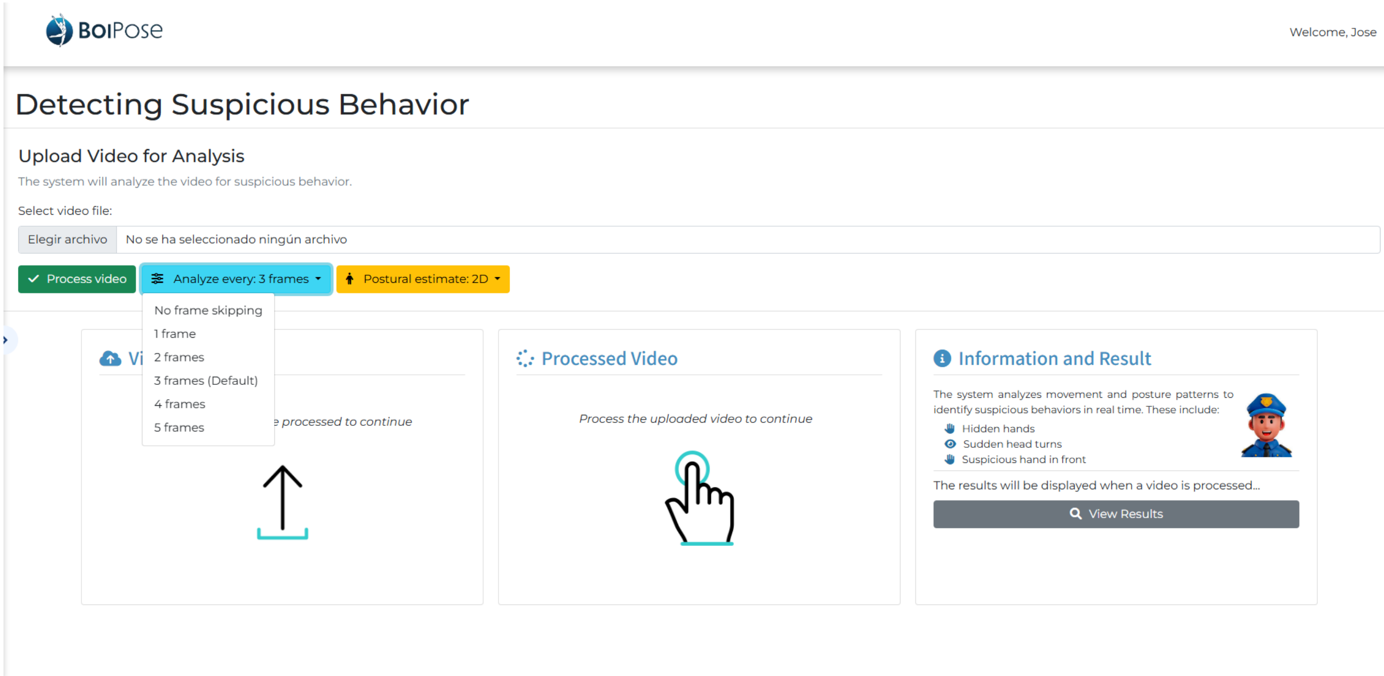

Fig. 6. Video upload detection screen.

Each video was recorded at 60 FPS and annotated to produce a ground truth. The complete dataset comprises **1,656 frames** distributed as follows: 600 frames for hand-under-clothing, 540 frames for hand-behind, and 516 frames for suspicious gaze.

The proposal was implemented in Python using OpenCV for frame extraction and MediaPipe for landmark estimation. The system supports two detection modes: 2D pose analysis (X,Y coordinates) and 3D pose analysis (X,Y,Z coordinates with depth approximation). The analysis pipeline includes a configurable frame-skip parameter (1–5 frames) to trade off detection latency and computational load, and the user may select either 2D or 3D processing for an experiment run.

B. *Evaluation metrics and procedure*

Performance was evaluated per-category using Precision, Recall and F1-score, computed from the confusion matrices obtained after processing all frames in the ground truth. Comparative experiments contrasted a 2D-only processing pipeline against a 3D-enabled pipeline that leverages depth information. In addition to per-class metrics, confusion matrices were inspected to identify the most frequent cross-category errors.

C. *Quantitative results*

Table II shows the main reported metrics (Precision, Recall, F1) for the three target behaviors, comparing 2D and 3D processing and reporting absolute improvements (Δ) observed when switching to 3D.

These results show an average F1 improvement of ≈ 7.6 percentage points when using 3D information.

D. *Confusion analysis*

A deeper inspection of confusion matrices revealed that the most problematic confusions in the 2D pipeline involved **hand-under-clothing** vs **hand-behind-body**. The counts reported in the experiments are:

- **Hand under clothing**: 2D correctly detected 514/600 (85.7%); it misclassified 85 cases as *hand behind*. In 3D, correct detections rose to 574/600 (95.7%) and misclassification fell to 14 cases — an 83.5% reduction in this error type.
- **Hand behind body**: 2D correctly detected 464/540 (86.0%); it misclassified 74 cases as *hand under clothing*. In 3D, correct detections rose to 516/540 (95.6%) and the misclassification dropped to 12 cases — an 83.8% reduction.
- **Suspicious gaze**: 2D precision was high (93.2%) but produced 35 false negatives (classified as none). The 3D system reduced false negatives from 35 to 14 (approx. 60% reduction) and improved precision to 97.3%.

In general, the bidirectional confusion between the two hand-related classes decreased from a total of 159 misclassified cases (74 + 85) in 2D to 26 cases (12 + 14) in 3D, a relative reduction of 83.6%. This strongly supports the hypothesis that depth information disambiguates occlusion/proximity scenarios that are inherently ambiguous in 2D.

E. *Results*

1) **3D improves detection of occlusion-related behaviors.** Incorporating depth (3D) reduced ambiguous cross-class

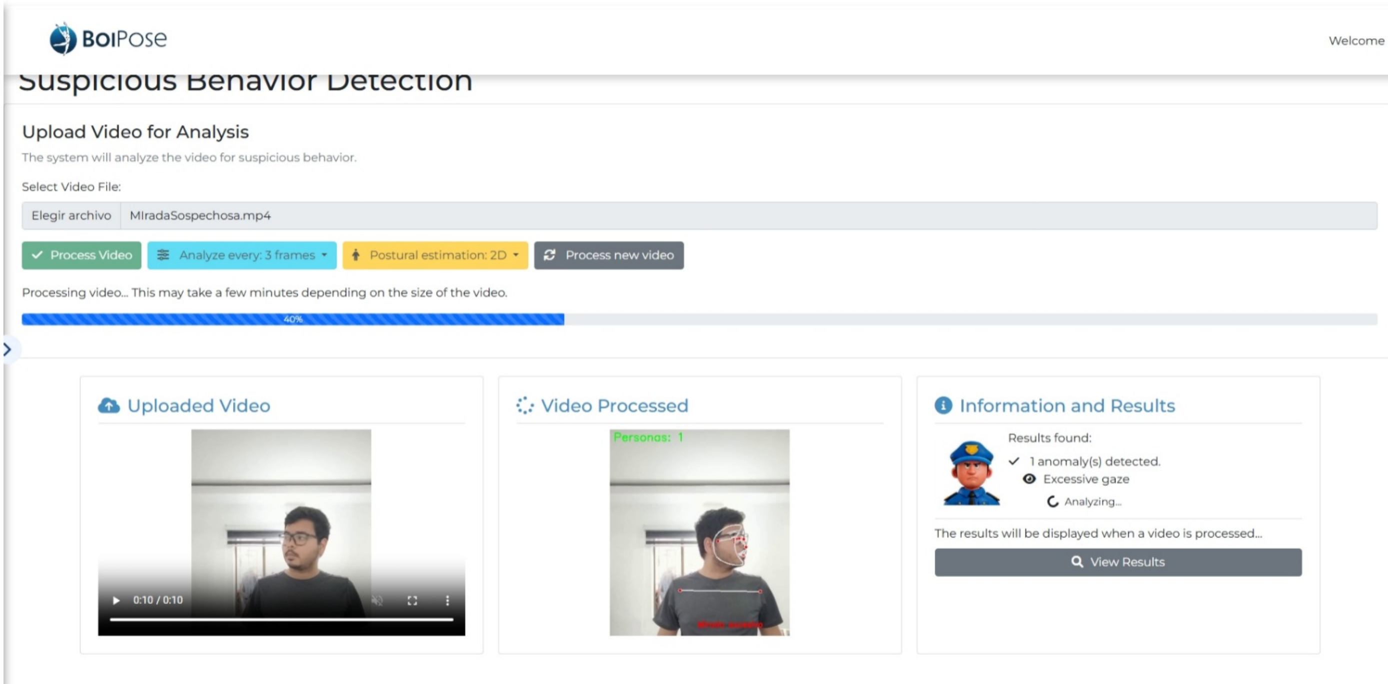

Fig. 7. Real-time analysis and detection process.

TABLE I
GROUND TRUTH USED IN THE EXPERIMENTS (FRAMES PER BEHAVIOUR)

Behavior	Frames	% of total
Hand under clothing	600	36.2%
Hand hidden behind body	540	32.6%
Suspicious gaze	516	31.2%
Total	1,656	100%

TABLE II
PERFORMANCE COMPARISON (2D VS 3D SPACES)

Category	Precision 2D (%)	Recall 2D (%)	F1-score 2D (%)	Precision 3D (%)	Recall 3D (%)	F1-score 3D (%)	ΔF1
Hand under clothing	85.7	87.3	86.5	95.7	98.0	96.8	+10.3
Hand behind body	86.0	85.9	86.0	95.6	97.4	96.5	+10.5
Suspicious gaze	93.2	93.2	96.5	97.3	97.3	98.6	+2.1
Average	88.3	88.8	89.7	96.2	97.6	97.3	+7.6

errors (hand under clothing vs hand behind body) by over 80% and produced an average F1 improvement of $\approx$ 7.6 points across the three behaviors. This validates the theoretical motivation to include spatial depth for disambiguating poses.

2) **2D remains useful for constrained, low-resource deployments.** Where real-time throughput or limited compute is critical, 2D processing remains a viable choice, trading some accuracy for speed. The pipeline's frame-skip parameter further helps balance latency and resource usage.
3) **Behavior-specific performance varies.** Some behaviors (e.g., suspicious gaze) were already detected with high recall in 2D; depth gave smaller but still important improvements. The biggest gains of 3D are for behaviors that involve occlusion or depth ambiguity.

IV. CONCLUSIONS

This work demonstrated the viability of using artificial intelligence to detect suspicious behavior based on body posture analysis. The system developed integrates Mediapipe and OpenCV to identify key points on the body and analyze both pre-recorded and real-time video, with support for local and remote cameras, giving it versatility in different surveillance environments.

Detection of three postures associated with suspicious behavior hand under clothing, hands behind back, and excessive staring was implemented using 2D and 3D analysis. The results show that, while the two-dimensional approach

TABLE III
SELECTED CONFUSION COUNTS (2D VS 3D)

Error type	2D misclassified	3D misclassified	Reduction (%)
Hand under clothing → Hand behind	85	14	83.5%
Hand behind → Hand under clothing	74	12	83.8%
False negatives (suspicious gaze)	35	14	60.0%

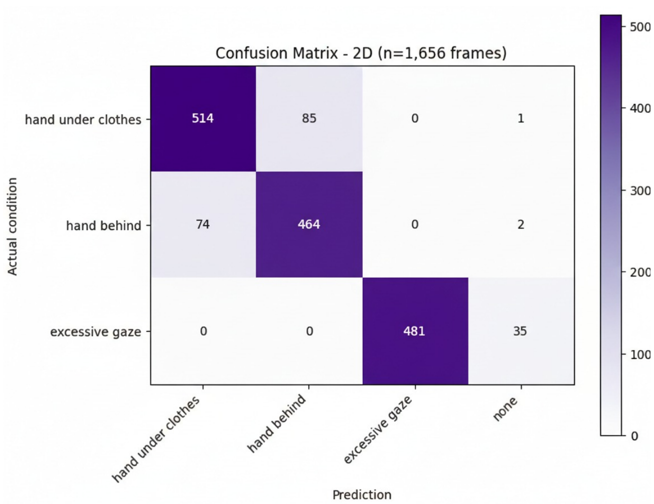

Fig. 8. 2D confusion matrix

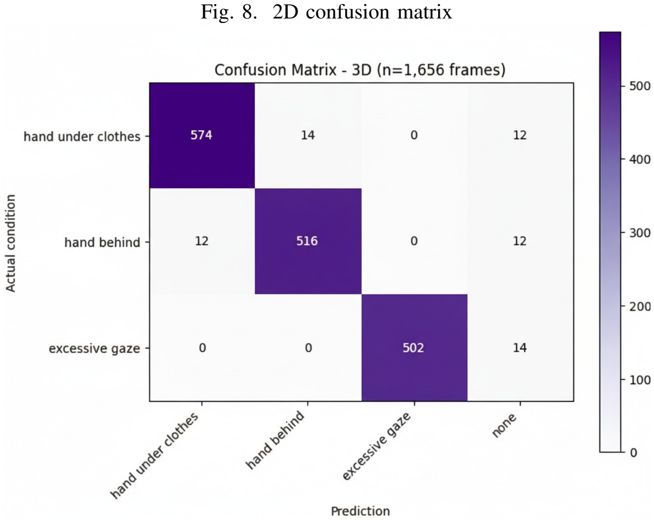

Fig. 9. 3D confusion matrix

offers greater speed and computational efficiency, three-dimensional analysis improves the accuracy of the system by considering the depth of the limbs, thus reducing false positives.

Taken together, the findings demonstrate the potential of computer vision and machine learning to strengthen intelligent surveillance systems, paving the way for future research aimed at optimizing computational efficiency, expanding the number of postures detected, and evaluating system performance in real-world and more complex scenarios.

REFERENCES

[1] Visionplatform, “How computer vision enhances cctv surveillance systems,” Online article, 2024, accessed 2024.

[2] A. Pereiro, “Reconocimiento automático de comportamientos anómalos mediante videovigilancia,” *Revista de Seguridad y Tecnología*, 2022.

[3] V. Kumar, “Body language and non-verbal communication,” *International Journal of Communication Studies*, 2021.

[4] M. Zloteanu *et al.*, “Sitting in judgment: How body posture influences deception detection,” *Psychological Science*, 2021.

[5] S. Wainio-Theberge and J. Armony, “Natural standing posture differences are associated with antisocial personality traits,” *Personality and Individual Differences*, 2024.

[6] F. Zhang *et al.*, “Mediapipe: A framework for building perception pipelines,” *arXiv preprint arXiv:2006.10204*, 2020.

[7] D. Chaudhary, S. Kumar, and V. S. Dhaka, “Video based human crowd analysis using machine learning: a survey,” *Computer methods in biomechanics and biomedical engineering: imaging & visualization*, vol. 10, no. 2, pp. 113–131, 2022.

[8] K. K. Verma, B. M. Singh, and A. Dixit, “A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system,” *International Journal of Information Technology*, vol. 14, no. 1, pp. 397–410, 2022.

[9] H. Ahn and H.-J. Cho, “Research of multi-object detection and tracking using machine learning based on knowledge for video surveillance system,” *Personal and Ubiquitous Computing*, vol. 26, no. 2, pp. 385–394, 2022.

[10] N. Choudhry, J. Abawajy, S. Huda, and I. Rao, “A comprehensive survey of machine learning methods for surveillance videos anomaly detection,” *IEEE Access*, vol. 11, pp. 114 680–114 713, 2023.

[11] D. Ly, “Ai and surveillance: Striking the balance for ethical deployment,” *Forbes Technology Council*, 2023.

[12] A. Omid and K. Soren, “Ethical challenges in artificial intelligence surveillance systems,” *AI and Society*, 2025.

[13] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee *et al.*, “Mediapipe: A framework for perceiving and processing reality,” in *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*, vol. 2019, 2019.

[14] Y. Lin *et al.*, “3d human pose estimation and depth ambiguity analysis,” *Computer Vision and Image Understanding*, 2023.

[15] G. Dibenedetto, S. Sotiropoulos, M. Polignano, G. Cavallo, and P. Lops, “Comparing human pose estimation through deep learning approaches: An overview,” *Computer Vision and Image Understanding*, p. 104297, 2025.

[16] J. L. Charco, A. D. Sappa, and B. X. Vintimilla, “Human pose estimation through a novel multi-view scheme.” in *VISIGRAPP (5: VISAPP)*, 2022, pp. 855–862.

[17] J. L. Charco, A. Cruz-Chóez, A. Yanza-Montalván, J. Zumba-Gamboa, and M. Galarza-Soledispa, “Litenetpose: A lightweight neural network for human pose estimation using attention modules,” *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, 2024.

[18] G. Bradski, A. Kaehler *et al.*, “Opencv,” *Dr. Dobb’s journal of software tools*, vol. 3, no. 2, pp. 1–81, 2000.

[19] A. Singh *et al.*, “Real-time video processing using opencv and python,” *International Journal of Computer Vision Applications*, 2021.

[20] F. Roggio *et al.*, “Limitations of pose estimation systems in complex environments,” *Sensors*, 2024.

[21] N. Intriago Santana and N. Martínez Ochoa, “Creación de un dataset para la estimación de postura de las personas con actitudes sospechosas en ambientes externos,” Master’s thesis, Universidad de Guayaquil, Guayaquil, Ecuador, 2025, proyecto FCI.