International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 06 June-2026 | Impact Factor: 3.5

A Centroid-Based Pose Estimation Framework for Real-Time Abnormal Activity Detection Using Adaptive KNN

Snehal Mathure, Bhakti Kokate, Vaishnavi Surwade, Snehal Chavan, Mahesh Dhande
Department of Artificial Intelligence and Data Science
Matoshri College of Engineering and Research Centre
Nashik, India
Email: snehalmathure5@gmail.com

Abstract—Real-time abnormal activity detection in surveil-lance systems requires a balance between accuracy, computa-tional efficiency, and privacy preservation. This paper presents a lightweight and efficient framework for detecting human activities using skeletal pose estimation on standard consumer-grade hardware. The proposed system extracts 33 human body landmarks and converts them into normalized spatial feature vectors using a centroid-based approach to achieve translation and scale invariance.

For classification, a K-Nearest Neighbors (KNN) model is employed with a visibility-weighted Manhattan (L1) distance metric to improve robustness against noisy or partially occluded landmarks. Additionally, an adaptive distance threshold derived from intra-class variations and a temporal smoothing mechanism are introduced to enhance prediction stability and reduce false detections.

The system supports multi-person detection using region-based processing and integrates an automated alert mechanism that generates real-time notifications, including snapshot capture and email transmission via SMTP. Experimental results demonstrate that the proposed framework achieves high accuracy with low latency, making it suitable for real-time surveillance and safety-critical applications.

Index Terms—Pose Estimation, K-Nearest Neighbors (KNN), Anomaly Detection, Real-time Surveillance, Multi-person Detec-tion, Edge Computing

I. Introduction

The rapid expansion of surveillance systems in urban en-vironments has resulted in a significant increase in visual data, making continuous manual monitoring impractical and error-prone. Human operators often experience reduced atten-tion over time, leading to missed critical events. Although automated approaches have been developed to address this limitation, many existing solutions rely on computationally intensive deep learning models and require high-end hard-ware, limiting their applicability in real-time and resource-constrained environments.

To overcome these challenges, there is a growing need for efficient and privacy-preserving edge-based solutions. This work proposes a lightweight abnormal activity detection framework that operates on skeletal pose data instead of raw video frames. By utilizing pose estimation techniques, human motion is represented as a set of 33 anatomical landmarks, which are further transformed into normalized feature vectors using a centroid-based approach. This representation reduces computational complexity and enhances privacy by avoiding storage of raw image data.

The classification of human activities is performed using an enhanced K-Nearest Neighbors (KNN) algorithm. Unlike conventional implementations, the proposed system incorpo-rates a visibility-weighted Manhattan (L1) distance metric, adaptive distance thresholding based on intra-class variations, and temporal smoothing across consecutive frames. These improvements significantly increase robustness against noise, occlusion, and unstable predictions.

Furthermore, the system supports multi-person detection through region-based processing and integrates an automated alert mechanism. Upon detection of abnormal activities such as falls or suspicious gestures, the system generates real-time alerts, captures evidence snapshots, and sends notifications via an SMTP-based communication module. This end-to-end integration transforms the system from a passive monitoring tool into an active response mechanism.

The key contributions of this work are as follows:

- A centroid-based normalization technique for efficient and invariant pose representation.
- An enhanced KNN classifier with visibility-weighted distance and adaptive thresholding.
- A temporal smoothing strategy to ensure stable real-time predictions.
- A multi-person detection framework suitable for dynamic environments.
- An automated alert system integrating real-time notifica-tion and evidence capture.

II. Related Work

Human activity recognition has evolved significantly from traditional image processing techniques to advanced deep learning-based approaches. Early methods relied on back-ground subtraction and handcrafted features for motion de-tection; however, these approaches were limited in handling complex and dynamic environments.

© 2026 The Author(s). Published by IJCOPE Journal. Website: https://ijcope.org/

Recent advancements have been dominated by deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are capable of learning spatial and temporal patterns from video data. While these methods achieve high accuracy, they require substantial computational resources and often depend on GPU-based architectures, making them unsuitable for real-time deployment on resource-constrained devices.

To address privacy concerns and computational limitations, several studies have explored the use of skeletal pose estimation as an alternative representation of human activity. Frameworks such as MediaPipe [1] enable real-time extraction of human body landmarks, allowing activity recognition without storing raw image data. These approaches improve privacy preservation while reducing data dimensionality.

Classical machine learning techniques, including Support Vector Machines (SVM) and K-Nearest Neighbors (KNN), have been applied to pose-based activity recognition due to their simplicity and low computational cost. However, many existing implementations rely on basic distance metrics and lack mechanisms to handle noise, occlusion, and temporal instability in real-time scenarios.

In contrast, the proposed work enhances the traditional KNN approach by incorporating a visibility-weighted distance metric, adaptive thresholding based on intra-class variations, and temporal smoothing across consecutive frames. Additionally, unlike most existing systems that focus solely on detection, this work integrates an automated alert mechanism for real-time response, addressing a critical gap in end-to-end surveillance systems [5].

III. Proposed Methodology

The proposed framework is designed for real-time abnormal activity detection using computationally efficient pose-based representations. Instead of processing raw video frames, the system operates on skeletal landmark data, significantly reducing computational overhead while preserving essential motion information. The architecture consists of three main components: data acquisition, feature processing and classification, and the automated alert mechanism.

A. Data Acquisition and Landmark Extraction

The system captures real-time video using a Flask-based streaming interface, with frame processing performed using OpenCV. Human subjects are detected using background subtraction and contour-based region-of-interest (ROI) extraction. This enables multi-person detection in dynamic environments.

MediaPipe Pose framework [1] is used to extract 33 anatomical landmarks from each frame. These landmarks form the basis for subsequent feature processing and classification.

B. Feature Normalization

To achieve translation and scale invariance, the extracted landmarks are normalized using a centroid-based approach. The centroid (c_x, c_y) is computed as the midpoint of the left and right hip coordinates:

$$c_x = \frac{x_{\text{left_hip}} + x_{\text{right_hip}}}{2}, \quad c_y = \frac{y_{\text{left_hip}} + y_{\text{right_hip}}}{2} \tag{1}$$

Each landmark is then translated relative to this centroid and scaled using the maximum distance between landmarks or torso length. This normalization ensures robustness against variations in position and distance from the camera.

C. KNN-Based Classification

The classification is performed using the K-Nearest Neighbors (KNN) algorithm [3]. Each frame is represented as a feature vector containing normalized coordinates and visibility scores of all landmarks.

Instead of using Euclidean distance, the system employs a visibility-weighted Manhattan (L1) distance metric:

$$d(q, p) = \sum_{j=1} w_j \cdot |q_j - p_j| \tag{2}$$

where w_j represents the visibility weight of each feature dimension. This approach reduces the impact of noisy or occluded landmarks.

An adaptive distance threshold is computed based on intra-class variations in the training dataset to improve classification reliability. Additionally, a temporal smoothing mechanism is applied, where predictions are confirmed only after consistent classification across multiple consecutive frames.

D. Automated Alert Mechanism

Upon detecting an abnormal activity, the system triggers an automated response pipeline. The detected frame is captured and stored as visual evidence. Simultaneously, an audio alert is generated using a text-to-speech engine.

An email notification is then sent via an SMTP-based communication module, including the captured image and activity details. A cooldown mechanism is implemented to prevent repeated alerts for the same event within a short time interval, ensuring efficient and reliable alert management.

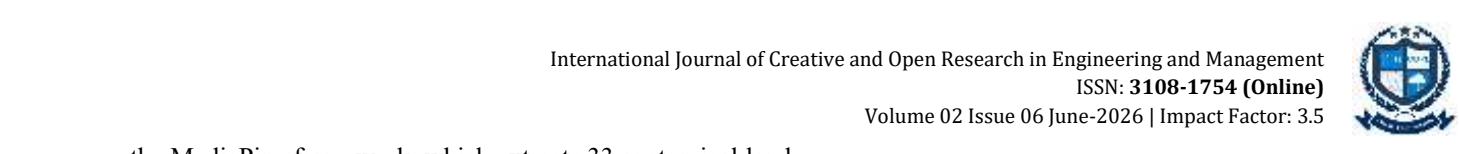
IV. System Implementation and Experimental Setup

To validate the proposed methodology, a real-time anomaly detection system was implemented and evaluated on standard consumer-grade hardware. This section describes the software environment, dataset construction, and integrated alert mechanisms.

A. Development Environment

The system was developed using Python and a set of open-source libraries to ensure efficient real-time performance. The core application and web-based monitoring interface were implemented using the Flask framework, enabling live video streaming through a browser.

Computer vision operations, including frame preprocessing and region-of-interest extraction, were performed using OpenCV [2]. Human pose estimation was carried out using



the MediaPipe framework, which extracts 33 anatomical landmarks per frame.

Feature processing and numerical computations, including normalization and distance calculations, were implemented using NumPy, while Pandas was used for dataset management and CSV-based storage. The system operates in real-time without requiring GPU acceleration, making it suitable for deployment on low-resource devices.

B. Dataset Generation

A lightweight, custom dataset was created to train the activity classification model. During system calibration, multiple samples of different activities such as sitting, standing, and abnormal gestures were recorded.

Each sample consists of normalized landmark coordinates and visibility scores, resulting in a feature vector stored in a structured CSV file (*activity.csv*). The dataset is incrementally updated during runtime, enabling continuous learning and adaptation without requiring large-scale pre-training.

C. Real-Time Classification Pipeline

During operation, skeletal landmarks extracted from each frame are normalized and converted into feature vectors. These vectors are classified using a K-Nearest Neighbors (KNN) model based on a visibility-weighted Manhattan (L1) distance metric.

An adaptive distance threshold, computed from intra-class variations in the dataset, is used to filter uncertain predictions. Additionally, a temporal smoothing mechanism ensures that classifications are confirmed only after consistent detection across multiple frames, improving robustness in dynamic environments.

The system also supports multi-person detection using background subtraction and contour-based region-of-interest extraction, allowing simultaneous monitoring of multiple individuals.

D. Automated Alert Mechanism

The system integrates both local and remote alert mechanisms for immediate response. Upon detection of an abnormal activity, the system generates an audio alert using a text-to-speech engine and captures the corresponding video frame as evidence.

An automated email notification is then sent via an SMTP-based communication module, including the captured image and activity details. A cooldown mechanism is implemented to prevent repeated alerts for the same event within a short time interval, ensuring efficient alert management and reducing redundancy.

V. Results and Discussion

The performance of the proposed system is evaluated using real-time experiments and classification metrics. The system demonstrates efficient detection and classification of human activities under varying conditions.

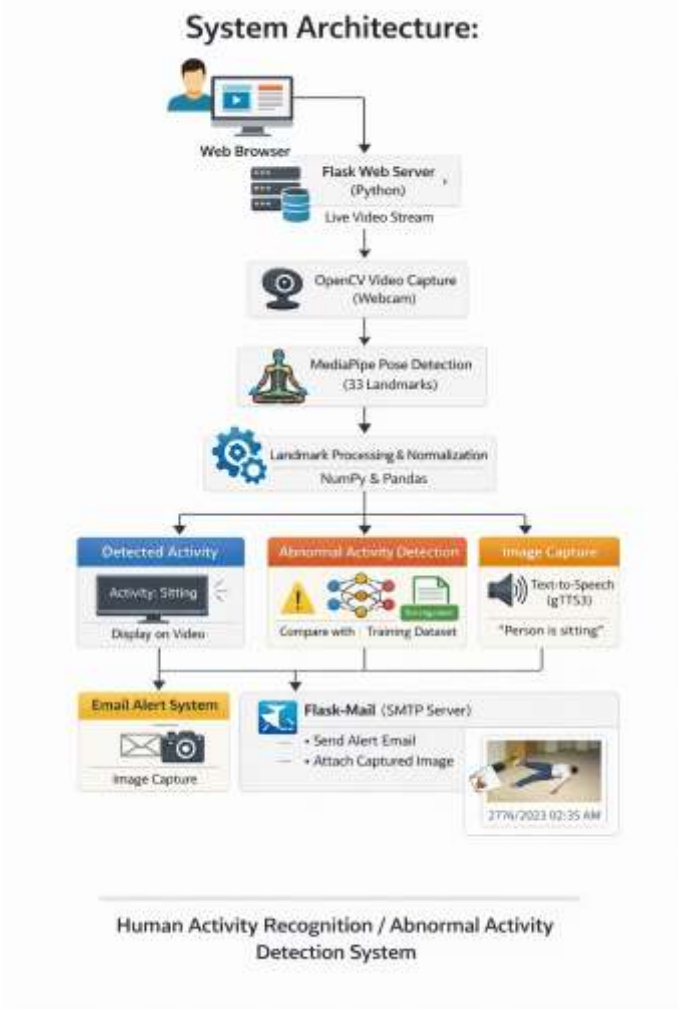


Fig. 1. Proposed system architecture illustrating the complete pipeline from video acquisition to anomaly detection and automated alert generation.

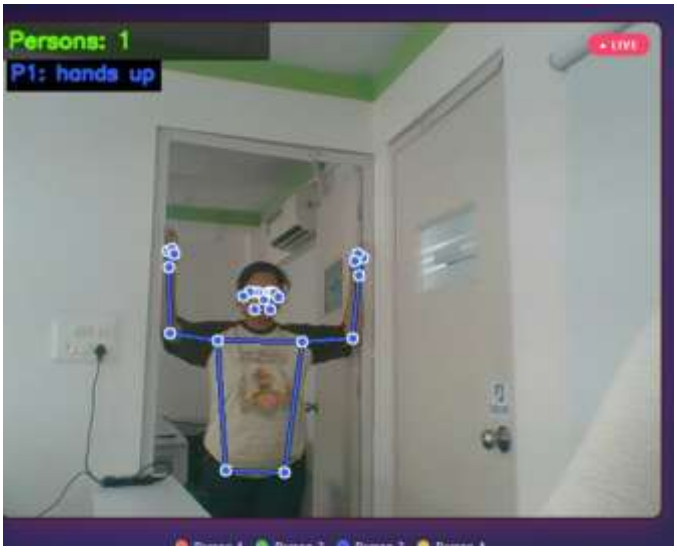


Fig. 2. Real-time pose detection using MediaPipe, where skeletal landmarks are used to classify human activities such as the 'hands-up' gesture.

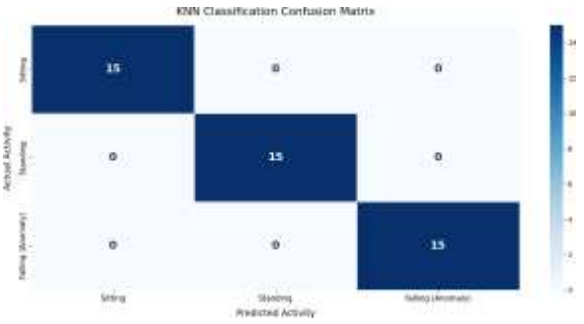


Fig. 3. Confusion matrix representing classification performance across activity classes.

As shown in Fig. 1, the system follows a structured pipeline integrating pose estimation, feature processing, classification, and alert generation.

Fig. 2 demonstrates the real-time capability of the system to accurately detect and classify human posture using skeletal landmark data.

The classification performance is evaluated using the confusion matrix in Fig. 3. The results indicate that the model achieves near-perfect classification accuracy across all activity classes, with minimal or no misclassification observed during testing.

The high accuracy can be attributed to the use of normalized pose features, visibility-weighted distance metrics, and temporal smoothing, which collectively improve robustness against noise and variations in human posture. The model demonstrates zero misclassification on the collected test dataset, indicating strong separability of feature representations; however, performance may vary under more complex real-world conditions.

VI. Conclusion

This paper presents a lightweight and efficient framework for real-time abnormal activity detection using skeletal pose estimation. By leveraging MediaPipe for 33-point landmark extraction and an enhanced K-Nearest Neighbors (KNN) classifier, the system effectively distinguishes between normal and anomalous human activities.

Unlike conventional approaches, the proposed method incorporates centroid-based normalization, a visibility-weighted distance metric, adaptive thresholding, and temporal smoothing to improve classification robustness. Additionally, the system supports multi-person detection and integrates an automated alert mechanism for real-time response.

Experimental results demonstrate that the system achieves near-perfect accuracy (approximately 100%) on the collected dataset while maintaining low computational overhead. This makes it suitable for deployment in real-time surveillance environments without requiring specialized hardware.

Future work will focus on extending the system to 3D pose estimation, increasing dataset diversity, and improving generalization for more complex real-world scenarios.

References

[1] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. Yong, J. Lee, W. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A Framework for Building Perception Pipelines," arXiv:1906.08172, 2019.

[2] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.

[3] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[5] M. Mubashir, L. Shao, and L. Seed, "A Survey on Fall Detection: Principles and Approaches," Neurocomputing, vol. 100, pp. 144–152, 2013.

[6] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "A Comprehensive Survey on Human Pose Estimation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[7] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep Learning for Sensor-Based Activity Recognition: A Survey," Pattern Recognition Letters, vol. 119, pp. 3–11, 2019.