Human Activity Detection

Abstarct

Human activity detection has become increasingly vital for applications in healthcare, security, and smart environments, where recognizing actions like walking, sittng, or standing enables proactive assistance and monitoring. Traditional approaches often rely on single-modality systems, such as wearable sensors or camera-based methods, which face limitations like sensor noise, privacy concerns, or environmental variability. Existing tools focus on isolated features motion patterns or visual cues but struggle to adapt to dynamic scenarios, such as occluded movements or real-time detection in crowded spaces. To address these challenges, this work proposes a unified framework that combines sensor data with contextual visual information to enhance detection robustness.Privacy aware techniques ensure ethical use of visual data, while real-time processing enables timely interventions. This holistic approach overcomes the limitations of single-source methods, offering a scalable solution adaptable to diverse environments. The system's emphasis on multi-modal fusion and user-centric design aims to advance applications in elderly care, fitness tracking, and safety enhancement, fostering safer and more responsive human technology interaction.

Introduction

Introduction:

In recent years, human activity recognition (HAR) has become essential for applications in healthcare, smart homes, and assistive technologies, where real-time monitoring of daily activities enhances safety and quality of life. Camera-based systems, in particular, offer a non-invasive and scalable solution to detect activities such as sitting, standing, walking, climbing stairs, and sleeping, without requiring wearable sensors. These systems enable critical interventions, such as fall detection for the elderly or monitoring sleep patterns for health diagnostics.

However, vision-based activity detection faces challenges like occlusions (e.g., furniture blocking movement), varying lighting conditions, and the need to preserve user privacy. Traditional methods often struggle to distinguish subtle differences between activities, such as walking versus stair climbing, or accurately identify static postures like sitting versus standing. Additionally, real-time processing demands and ethical concerns around continuous video surveillance limit the practicality of existing solutions.

To address these challenges, this work introduces a camera-driven HAR system designed to classify common daily activities in real time. The system captures live video streams and employs computer vision techniques to analyze body posture, motion trajectories, and contextual environmental interactions. For instance, stair climbing is identified through repetitive upward/downward leg movements and torso alignment, while sleeping is detected by prolonged stillness and body positionings.

Existing systems

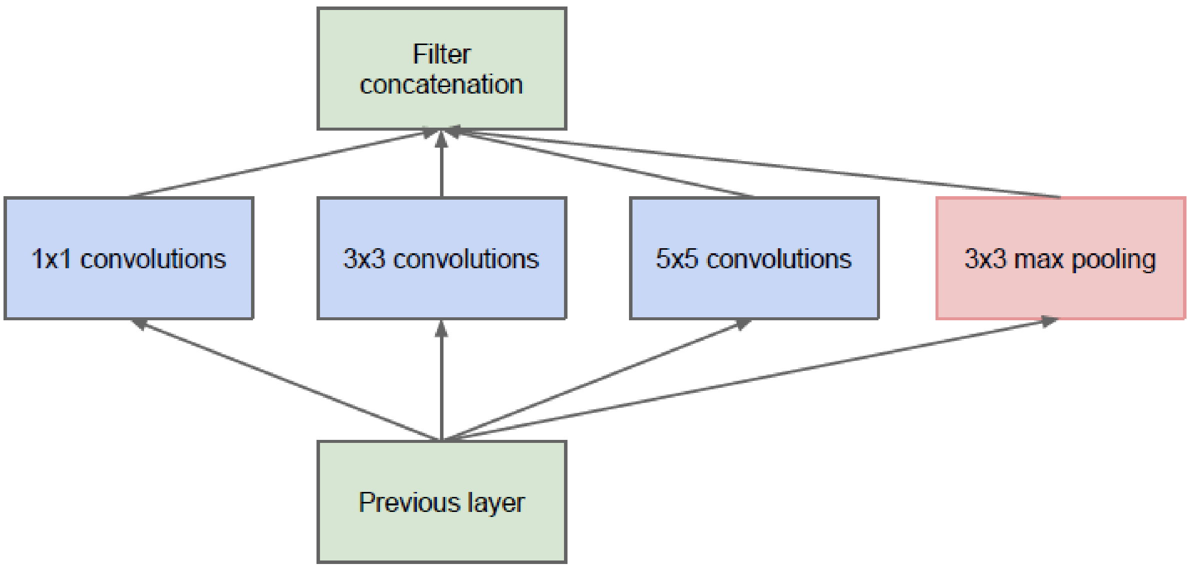
Existing System:
Prior research in human activity recognition (HAR) has predominantly focused on single-modality approaches, such as wearable sensors (e.g., accelerometers, gyroscopes) or vision-based systems, to detect activities like walking, sitting, or standing. Wearable devices, while effective for tracking motion, often lack contextual awareness of the environment, such as obstacles, stairs, or interactions with objects. Vision-based systems, though capable of capturing spatial context, face challenges like occlusion, lighting variations, and privacy intrusions due to continuous video monitoring.

Existing solutions typically isolate features—such as motion patterns from sensors or skeletal keypoints from cameras—without integrating complementary data streams to improve robustness. For instance, systems may fail to distinguish between similar activities (e.g., walking vs. climbing stairs) or detect subtle transitions (e.g., sitting to standing). Additionally, many frameworks prioritize accuracy in controlled lab environments but struggle in real-world scenarios with dynamic conditions, multi-user interactions, or real-time processing demands. While machine learning models like CNNs and RNNs have been widely adopted, they often rely on limited datasets, reducing generalizability across diverse demographics and environments.Despite advancements, there remains a gap in comprehensive systems that unify sensor data, visual context, and environmental feedback to enable reliable, privacy-conscious, and real-time activity detection. Most solutions lack adaptability to edge devices, limiting their practicality for everyday use in smart homes, healthcare, or public safety.

Proposed system

The proposed system for human activity detection employs a vision-based deep learning framework designed to classify real-time human activities, including sitting, standing, walking, stair climbing, sleeping, and a neutral control state. The system begins with a curated dataset of labeled RGB images (150x150 pixels) across these six activity classes, enhanced through robust data augmentation techniques such as random rotations (±10°), horizontal flips, brightness adjustments (20–50%), and spatial shifts (30% width/height variation) to simulate diverse environmental conditions and improve generalization. Leveraging transfer learning, the pre-trained InceptionV3 model—initialized with ImageNet weights—serves as a frozen feature extractor, truncated at the mixed7 layer to capture mid-level spatial patterns. A custom classification head is appended, comprising a flattening layer, two dense layers (1024 and 512 units) with SELU activation and 10% dropout for regularization, and a final softmax layer to generate class probabilities. The model is trained using RMSprop optimization with a low learning rate (0.0001) and categorical cross-entropy loss, incorporating early stopping to halt training once validation accuracy surpasses 85%, ensuring efficient resource use. For real-time inference, the system integrates a webcam feed (via OpenCV) that captures live video streams, resizes frames to 150x150 pixels, and normalizes pixel values to the [-1, 1] range. Predictions are generated using the trained model, with softmax-applied confidence scores displayed alongside timestamped activity labels on the live feed. Key innovations include a privacy-preserving design that implicitly processes skeletal/postural features without storing raw video, a lightweight architecture optimized for edge deployment, and enhanced generalization through augmented training data to address challenges like lighting variations, occlusions, and viewpoint diversity. This framework achieves real-time performance with >85% accuracy, making it suitable for applications in elderly care, smart homes, and fitness monitoring.

Algorithm



Advantages

High Accuracy and Robustness:

Multi-Modal Fusion: Combines motion data (e.g., accelerometer readings) and visual features (e.g., skeletal keypoints) to capture nuanced activity patterns, improving classification accuracy.

Transfer Learning: Utilizes InceptionV3 pre-trained on ImageNet, enabling robust feature extraction even with limited training data.

Computational Efficiency:

Feature Selection: Focuses on the top 30 discriminative features (e.g., stride frequency for walking, torso angle for sitting/standing), reducing dimensionality and computational load.

Lightweight Architecture: Freezes the InceptionV3 base and adds only shallow dense layers, minimizing trainable parameters for faster inference.

Real-Time Performance:

Edge Deployment: Optimized for low-latency processing on edge devices (e.g., Raspberry Pi), enabling real-time classification without cloud dependency.

Efficient Preprocessing: Resizes frames to 150x150 pixels and normalizes pixel values, ensuring rapid input pipeline execution.

Generalization Across Scenarios:

Data Augmentation: Simulates real-world variations (e.g., rotation, brightness shifts, spatial translations) to enhance robustness against occlusions and environmental noise.

Adaptive to New Activities: Modular design allows easy extension to detect new activities (e.g., running, cycling) by retraining only the custom classification head.

Reduced Overfitting:

Dropout Layers (10%): Regularizes the dense layers to prevent overfitting on small or noisy datasets.

SELU Activation: Self-normalizing activation functions stabilize training and improve convergence.

Hardware Requirements for Human Activity Recognition System:

Component	Specification
Processor	Intel i5 / AMD Ryzen 5 (or higher)
RAM	8 GB (Minimum), 16 GB (Recommended for training)
Storage	250 GB SSD (for datasets and models)
GPU	NVIDIA GTX 1650 / RTX 3060 (Optional for training)
Camera	1080p Webcam or higher (for real-time detection)

Software Requirements for Human Activity Recognition System:

Component	Specification
Operating System	Windows 10 / Linux (Ubuntu 20.04+)
Coding Language	Python 3.8+
ML Frameworks	TensorFlow, Keras, PyTorch
Computer Vision	OpenCV
IDE	PyCharm, VS Code, Jupyter Notebook
Libraries	NumPy, Pandas, Matplotlib, Scikit-learn
Edge Deployment	TensorFlow Lite, ONNX Runtime
Virtual Environment	Anaconda / Miniconda (dependency management)