Multi-Sensor Data Fusion for Accurate Human Activity Recognition with Deep Learning

Edmundo Jalon Arias*, Luz M. Aguirre Paz, Luis Molina Chalacan

Universidad Regional Autónoma de los Andes (UNIANDES), Ecuador

Emails: uq.sistemas@uniandes.edu.ec; direccionadmision@uniandes.edu.ec; uq.luismolina@uniandes.edu.ec

Abstract

In the era of pervasive computing and wearable technology, the accurate recognition of human activities has gained paramount importance across a spectrum of applications, from healthcare monitoring to smart environments. This paper introduces a novel methodology that leverages the fusion of multi-sensor data with deep learning techniques to enhance the precision and robustness of human activity recognition. Our approach commences with the transformation of accelerometer and gyroscope time-series data into recurrence plots, facilitating the distillation of temporal patterns and dependencies. Subsequently, a dual-path convolutional network framework is employed to extract intricate sensory patterns independently, followed by an attention module that fuses these features, capturing their nuanced interactions. Rigorous experimental evaluations, including comparative analyses against traditional machine learning baselines, validate the superior performance of our methodology. The results demonstrate remarkable classification performance, underscoring the efficacy of our approach in recognizing a diverse range of human activities. Our research not only advances the state-of-the-art in activity recognition but also highlights the potential of deep learning and multi-sensor data fusion in enabling context-aware systems for the benefit of society.

Keywords: Multi-Sensor Data Fusion; Deep Learning; Human Activity Recognition; Sensor Fusion Techniques; Data Fusion Strategies; Cross-Modal Fusion

1. Introduction

In an era marked by the proliferation of smart devices and the Internet of Things (IoT), the ability to understand and interpret human activities from sensor data has become a pivotal element in the realms of healthcare, security, human-computer interaction, and beyond. With the ever-increasing prevalence of sensors embedded in smartphones, wearables, ambient environments, and even urban infrastructures, the wealth of information generated by these devices holds immense potential for advancing our understanding of human behavior and well-being.

Human activity recognition (HAR) has emerged as a fundamental research area within the fields of artificial intelligence and data science. It aims to automatically identify, classify, and comprehend human activities from the myriad of data streams produced by sensors. Accurate HAR has the potential to revolutionize various domains, from personalized healthcare and assisted living to context-aware computing and intelligent transportation systems. However, the task of HAR is multifaceted and challenging due to the inherent complexities in human behavior, the diversity of sensors, and the variability in data sources. This paper delves into the heart of this challenge by exploring the integration of multi-sensor data and deep learning techniques for the purpose of achieving accurate HAR. Multi-sensor data fusion refers to the harmonious combination of information from diverse sensors, allowing for a more
comprehensive and robust understanding of the activities being performed. Deep learning, on the other hand, has demonstrated remarkable prowess in modeling complex patterns and representations from data, making it a powerful candidate for enhancing the performance of HAR systems.

Our research sets out to bridge the gap between the richness of multi-sensor data and the capabilities of deep learning models. Through a synergistic approach, we endeavor to unlock the full potential of sensor data by leveraging the inherent strengths of various sensor modalities while harnessing the expressive power of deep neural networks. In doing so, we aim to contribute not only to the advancement of HAR but also to the broader field of machine learning and artificial intelligence. In the following sections, we will delve into the key components of our paper as summarized in Table 1.

Table 1: Overview of paper organization

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduction</td>
<td>- Present the context and motivation for the research.</td>
</tr>
<tr>
<td></td>
<td>- Highlight the importance of multi-sensor data fusion and deep learning in HAR.</td>
</tr>
<tr>
<td></td>
<td>- Provide an overview of the paper's structure.</td>
</tr>
<tr>
<td>II. Related Work</td>
<td>- Review existing literature on HAR, sensor fusion, and deep learning in HAR.</td>
</tr>
<tr>
<td></td>
<td>- Identify gaps in the literature that your research addresses.</td>
</tr>
<tr>
<td>III. Methodology</td>
<td>- Describe the data sources and datasets used in the research.</td>
</tr>
<tr>
<td></td>
<td>- Explain the data preprocessing steps.</td>
</tr>
<tr>
<td></td>
<td>- Detail the feature extraction techniques employed.</td>
</tr>
<tr>
<td></td>
<td>- Present the deep learning architectures used for HAR.</td>
</tr>
<tr>
<td>IV. Experimental Configurations</td>
<td>- Outline the experimental setup and equipment used.</td>
</tr>
<tr>
<td></td>
<td>- Describe the training and testing procedures.</td>
</tr>
<tr>
<td></td>
<td>- Specify the evaluation metrics used for performance assessment.</td>
</tr>
<tr>
<td>V. Results and Discussion</td>
<td>- Present the experimental results, including quantitative performance metrics.</td>
</tr>
<tr>
<td></td>
<td>- Discuss the findings and insights gained from the results.</td>
</tr>
<tr>
<td></td>
<td>- Compare your approach to existing methods, if applicable.</td>
</tr>
<tr>
<td>VI. Conclusion</td>
<td>- Summarize the key contributions and findings of the research.</td>
</tr>
<tr>
<td></td>
<td>- Discuss the implications and applications of the proposed approach.</td>
</tr>
<tr>
<td></td>
<td>- Highlight any limitations and suggest areas for future research.</td>
</tr>
</tbody>
</table>

2. Related Works

In this section, we embark on a comprehensive exploration of the existing body of research that forms the foundation upon which our study is built. HAR is a multifaceted domain at the intersection of sensor technology, machine learning, and context-aware computing. Nweke et al. [11] investigated multi-sensor fusion for mobile and wearable sensor based HAR. Their work laid the foundation for considering the fusion of data from different sensors to improve activity recognition accuracy. They explored the challenges and potential benefits of sensor fusion techniques, which aligns with our research focus. In a subsequent study, Nweke et al. [12] extended their exploration by employing multiple classifier systems for human activity identification. Their approach aimed to enhance classification
performance by combining outputs from multiple classifiers. This is relevant to our work as we also seek to improve recognition accuracy through data fusion, albeit with a deep learning perspective. Nafea et al. [13] contributed to the field by applying Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) to multi-sensor HAR. Their research highlights the effectiveness of deep learning techniques in handling multi-sensor data, which aligns with our methodology. Yu et al. [14] introduced the integration of neuromorphic computing with multi-sensing data fusion for HAR. Their work showcases the potential of cutting-edge technologies in improving recognition accuracy. While their focus differs from ours, their approach to sensor fusion is noteworthy. San Buenaventura et al. [15] explored deep learning techniques for smartphone based HAR using multi-sensor fusion. Their work is particularly relevant as it demonstrates the application of deep learning in multi-sensor contexts, mirroring our research objectives.

Gravina and Li [16] ventured into emotion-relevant activity recognition based on smart cushions using multi-sensor fusion. Although their emphasis is on emotional states, their utilization of multi-sensor data fusion techniques may offer insights applicable to our research on general activity recognition. Aguileta et al. [17] conducted a comprehensive survey on multi-sensor fusion for activity recognition, providing a valuable overview of existing approaches and challenges. Their survey serves as a valuable resource for understanding the broader landscape of multi-sensor data fusion in activity recognition. Patil et al. [18] explored data integration based HAR using deep learning models. Their research aligns with our focus on deep learning and data integration for activity recognition, and their findings may offer complementary insights. Cao et al. [19] addressed the optimization of multi-sensor deployment for wearable activity recognition, demonstrating the importance of sensor placement and ensemble pruning. This work may contribute to optimizing our experimental configurations. Finally, Miao et al. [20] proposed a dynamic inter-sensor correlation learning framework for multi-sensor based wearable HAR. Their approach adapts to changing sensor relationships, a concept that could enhance the adaptability of our multi-sensor fusion methodology.

3. Methodology

The methodology section forms the bedrock of our research, offering a detailed blueprint of the procedures, techniques, and frameworks employed to accomplish the core objectives of our study. In this section, we unveil the inner workings of our approach for human activity recognition through the fusion of multi-sensor data using deep learning. Our methodology is meticulously structured, beginning with data collection and preprocessing, followed by feature extraction, model development, and evaluation.

Step 1: Data Transformation with Recurrence Plots

In the initial phase of our methodology, we grapple with the challenge of harmoniously integrating data from two distinct sensor sources: accelerometer and gyroscope. These sensors provide crucial insights into the movements and orientations of the subjects, but they do so in the form of time-series data. To unlock the potential hidden within these time-series, we employ a pivotal technique known as recurrence plots. Recurrence between time \( i \) and time \( j \) is given as follows:

\[
R_{ij} = \Theta(\epsilon_i - \| \bar{x}_i - \bar{x}_j \|), \bar{x}_i \in \mathbb{R}^m, i, j = 1, ..., N
\]  

(1)

This transformation method plays a foundational role in our approach, enabling us to convert the continuous streams of sensor data into a structured format that can be readily processed by deep learning models.

Recurrence plots function as a bridge between raw sensor readings and image-based data representation. By transforming the time-series data into recurrence plots, we effectively capture the temporal dependencies and patterns that are indicative of human activities. The transformation process renders the data amenable to convolutional neural networks (CNNs) and other image-based deep learning architectures, which excel at pattern recognition tasks. This strategic transformation equips our model with the ability to discern nuanced activity patterns from the sensor data, a critical step in achieving accurate human activity recognition. In essence, this initial phase lays the groundwork for the subsequent stages of our methodology, ensuring that our deep learning models can effectively leverage the multi-sensor data for enhanced recognition performance.

Step 2: Dual-Path Convolutional Networks for Sensory Pattern Learning

Having successfully transformed the accelerometer and gyroscope time-series data into recurrence plots in the previous step, our next strategic move involves the utilization of dual-path convolutional networks. This phase marks the convergence of the transformed data streams, where we embark on the intricate process of pattern learning specific to each sensor modality.

Doi: [https://doi.org/10.54216/FPA.130206](https://doi.org/10.54216/FPA.130206)

Received: April 17, 2023 Revised: July 15, 2023 Accepted: September 16, 2023
In our approach, we recognize the distinct characteristics and information nuances offered by the accelerometer and gyroscope data. The use of dual-path convolutional networks allows us to maintain the fidelity of these unique patterns by enabling each sensor's data to follow an independent learning trajectory.

\[ S_{i,j} = (I * K)_{i,j} = \sum_{m} \sum_{n} I_{i,j} \cdot K_{i-m,j-n}. \]  

The dual-path architecture bifurcates the network into two distinct pathways, with one dedicated to processing the recurrence plots generated from accelerometer data and the other specialized in handling the gyroscope-derived recurrence plots.

By doing so, our model excels at learning sensor-specific patterns and extracting nuanced features that are inherent to each data source. This dual-path configuration empowers our deep learning architecture to effectively capture and discriminate between the subtle variations in human activity patterns, which may be manifested differently in accelerometer and gyroscope data. As a result, we achieve a comprehensive understanding of the sensory dimensions of human activities, paving the way for more accurate and robust recognition in the subsequent phases of our methodology. This dual-path approach underscores our commitment to harnessing the unique information embedded within each sensor's data while maintaining their combined synergy for a holistic human activity recognition solution.

We use ReLU activation after each convolution as follows:

\[ f(x) = \max(0, x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \]  

**Step 3: Attention Module for Feature Fusion**

After independently processing the accelerometer and gyroscope recurrence plots through the dual-path convolutional networks, the next pivotal stage in our methodology revolves around the fusion of these distinct features. To achieve this fusion, we employ a sophisticated mechanism known as an attention module. This module plays a crucial role in enabling our model to capture and emphasize the interactions, dependencies, and synergies between the features extracted from the two sensor modalities. The attention module acts as a dynamic information aggregator, allowing our model to adaptively weigh the significance of the learned features from both sensor types. This adaptability is particularly valuable because it permits our deep learning architecture to prioritize features that are most relevant for recognizing the ongoing human activities. By attentively fusing the information from accelerometer and gyroscope pathways, our model excels at discerning subtle interdependencies that may be indicative of specific activities. This can be expressed as follows:

\[ K = Q = V = W \]  

\[ A(K, Q, V) = \text{softmax} \left( \frac{K \cdot Q^T}{\sqrt{d}} \right) V \]  

In the above formula, we use the following terms Query (Q), Key (K), and Value (V). This fusion not only enhances the model's overall predictive power but also ensures that it comprehensively leverages the combined knowledge embedded within the data streams. The result is a more accurate and context-aware representation of human activities, where the model discerns not only the nuances within each sensor's data but also the intricate relationships between them.

**Step 4: Activity Classification with SoftMax Layer**

In the final phase of our methodology, we bring together all the insights and representations garnered throughout the data transformation, feature learning, and fusion processes to perform the ultimate task: human activity classification. This critical classification step occurs within the SoftMax layer, employing categorical cross-entropy as the guiding criterion. The SoftMax layer serves as the ultimate decision-maker in our deep learning architecture, where the model's learned representations and feature interactions are distilled into a probabilistic framework. By applying SoftMax, we generate probability distributions over the different classes of human activities.

\[ f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} (i=1,2,\ldots,N) \]
This distribution allows us to determine the likelihood of each recorded sensor pattern belonging to a particular activity category. The use of categorical cross-entropy loss as the optimization objective ensures that the model's predictions align closely with the ground truth labels, thus facilitating the accurate classification of human activities.

$$L(w) = \frac{1}{N} \sum_{n=1}^{N} H(p_n, q_n) = -\frac{1}{N} \sum_{n=1}^{N} [y_n \log \hat{y}_n + (1 - y_n) \log (1 - \hat{y}_n)],$$ \hspace{1cm} (7)

4. Experimental Setups

In this section, we delve into the intricate details of our experimental configurations, providing a meticulous account of the setup and methodologies employed to evaluate the efficacy of our proposed multi-sensor data fusion approach for HAR. A robust and well-structured experimental framework is the linchpin upon which the credibility and reliability of our findings rest. Our experimental implementation setups were meticulously designed to ensure the reliability and reproducibility of our research findings. Table 2 provides a concise overview of the hardware and software components that constituted our experimental environment. The choice of hardware platforms was driven by the need for computational efficiency and scalability, while software selections were tailored to support deep learning model development and multi-sensor data fusion. This unified configuration served as the foundation for our comprehensive investigation into the fusion of multi-sensor data for HAR.

Table 2: Implementation Setups

<table>
<thead>
<tr>
<th>Component</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hardware Platform</strong></td>
<td>- CPU: Intel Xeon</td>
</tr>
<tr>
<td></td>
<td>- GPU: RTX2060</td>
</tr>
<tr>
<td></td>
<td>- RAM: 32 GB</td>
</tr>
<tr>
<td></td>
<td>- Storage: 1TB</td>
</tr>
<tr>
<td><strong>Software Framework</strong></td>
<td>- Deep Learning Framework: TensorFlow 2.8.0</td>
</tr>
<tr>
<td></td>
<td>- Operating System: Windows 10</td>
</tr>
<tr>
<td></td>
<td>- Programming Language: Python</td>
</tr>
<tr>
<td></td>
<td>- Libraries: NumPy, pandas, SciPy, sk-learn</td>
</tr>
</tbody>
</table>

In our research, we employ the MotionDataset as a compelling case study to evaluate the effectiveness of our multi-sensor data fusion approach for human activity recognition using deep learning. This dataset presents a valuable resource for our investigation due to its diverse and rich collection of time-series data. Captured using accelerometer and gyroscope sensors, specifically attitude, gravity, userAcceleration, and rotationRate, this dataset offers a granular view of human movements and behavior. Collected using an iPhone 6s, securely positioned in the participant's front pocket, the data acquisition process is facilitated by SensingKit, which leverages the Core Motion framework on iOS devices. The MotionDataset encompasses the contributions of 24 participants, spanning a wide range of demographics, including gender, age, weight, and height. These individuals engaged in a series of six distinct activities, each recorded in 15 separate trials. The activities encompass both simple and complex movements, including downstairs and upstairs navigation, walking, jogging, and periods of sitting and standing. This comprehensive variety of activities in a controlled environment allows us to rigorously test the capability of our multi-sensor data fusion method in distinguishing between diverse human activities. Beyond its primary purpose of activity recognition, the MotionDataset offers an intriguing dimension for exploration. It seeks to uncover 'personal attribute fingerprints' within the time-series sensor data. This concept refers to attribute-specific patterns embedded in the data, patterns that could potentially be utilized to infer personal attributes such as gender or personality traits of the participants, extending the dataset's utility beyond traditional activity recognition applications.
5. Results and Discussion

In this pivotal section, we unveil the outcomes of our empirical endeavors and embark on a comprehensive discussion of the results obtained through the fusion of multi-sensor data for HAR using deep learning. The journey through this section is twofold: first, we present the quantitative results and performance metrics acquired from our meticulously designed experiments, elucidating the accuracy, precision, recall, F1-score, and other pertinent indicators. Second, we delve into a qualitative analysis, where we decipher the implications of our findings, elucidate the strengths and limitations of our approach, and explore the broader implications for the field of HAR.

In Figure 1, parts (a) and (b) present captivating visualizations through Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) plots, respectively, for our multi-sensor dataset. These visualizations offer profound insights into the distribution and separability of data points within our feature space. In part (a), the PCA plot elucidates the variance and dimensionality of the data, revealing clusters that correspond to different human activities. This aids in understanding the intrinsic structure of the dataset. Meanwhile, part (b), the t-SNE plot, takes a step further by mapping the data into a lower-dimensional space while preserving local structures and intricate relationships. These visualizations collectively serve as a critical foundation for our subsequent analysis, guiding feature selection, model development, and ultimately, enhancing the efficacy of our multi-sensor data fusion approach for human activity recognition.

Figure 1: Data Visualization with PCA (left) and t-SNE (right)

Figure 2: Recurrence Plots for Human Activity Classes

Doi: https://doi.org/10.54216/FPA.130206
Received: April 17, 2023 Revised: July 15, 2023 Accepted: September 16, 2023
In Figure 2, we present a compelling visualization of recurrence plots specific to each class within our dataset. These recurrence plots offer an intuitive and insightful glimpse into the temporal patterns and dependencies inherent to different human activities. Each recurrence plot serves as a visual representation of the underlying sensor data, with distinct clusters and structures that correspond to unique activity patterns. This visualization not only showcases the complexity of the data but also underscores the potential for our model to capture these intricate patterns through its deep learning architecture. In Figure 3, the visualization of the confusion matrix presents a crucial analytical component that offers a comprehensive overview of the model’s performance in classifying human activities. This matrix, a fundamental tool in evaluating classification models, succinctly illustrates how well the model aligns its predictions with the ground truth labels for each activity class. The visualization of the confusion matrix in Figure 3 serves as a critical component in validating the robustness and effectiveness of our methodology in accurately recognizing a wide array of human activities.

![Confusion matrix](image)

**Figure 3:** Confusion matrix of the proposed fusion model

To rigorously evaluate the effectiveness of our proposed multi-sensor data fusion approach for human activity recognition, we conducted a series of comparative experiments against well-established ML baselines. This comparative analysis aims to shed light on the performance advantages offered by our methodology in harnessing the synergy of multi-sensor data and deep learning techniques. For our baseline models, we considered traditional ML algorithms commonly used in activity recognition tasks, including Decision Trees, Random Forests, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN) [21]. These ML baselines are characterized by their versatility and simplicity, making them suitable for benchmarking against our deep learning-based approach.
Table 3: Comparative Analysis on Different Performance Metrics

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy (%)</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F1-Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Trees</td>
<td>82.4</td>
<td>83.2</td>
<td>82.1</td>
<td>82.6</td>
</tr>
<tr>
<td>Random Forests</td>
<td>87.6</td>
<td>88.2</td>
<td>87.4</td>
<td>87.8</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td>89.3</td>
<td>89.7</td>
<td>89.1</td>
<td>89.4</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td>78.9</td>
<td>79.6</td>
<td>78.7</td>
<td>79.1</td>
</tr>
<tr>
<td>Proposed Approach</td>
<td>96.2</td>
<td>94.5</td>
<td>94.1</td>
<td>94.3</td>
</tr>
</tbody>
</table>

In Table 3, we observe that our proposed approach significantly outperforms traditional ML baselines across all key performance metrics. The higher accuracy, precision, recall, and F1-score values attained by our methodology underscore its superior capability in accurately recognizing diverse human activities. These results validate the efficacy of leveraging deep learning and multi-sensor data fusion, marking a substantial advancement in the field of HAR.

6. Conclusions

This paper presents a novel and effective methodology for human activity recognition through the fusion of multi-sensor data using deep learning. Our approach capitalizes on the intrinsic patterns and interdependencies captured within accelerometer and gyroscope sensor data, harnessing their combined potential for improved accuracy and robustness. Through the transformation of time-series data into recurrence plots, feature learning with dual-path convolutional networks, and the incorporation of an attention module to capture feature interactions, our methodology demonstrates a holistic approach to multi-sensor data fusion. The results, as showcased through comprehensive experiments and visualizations, unequivocally validate the superiority of our approach, with significantly higher accuracy and precision compared to traditional machine learning baselines. Moreover, our model excels in recognizing a wide range of human activities, even in scenarios involving intricate temporal patterns. This research not only advances the state-of-the-art in human activity recognition but also highlights the potential of deep learning and multi-sensor fusion in diverse real-world applications. Our findings pave the way for enhanced context-aware systems, from healthcare monitoring to smart environments, where the accurate recognition of human activities plays a pivotal role. As we move forward, the synergy between deep learning and multi-sensor data fusion promises to unlock new frontiers in human behavior analysis and decision support systems, ultimately enhancing our ability to understand and interact with the world around us.

References


Doi: https://doi.org/10.54216/FPA.130206
Received: April 17, 2023 Revised: July 15, 2023 Accepted: September 16, 2023


