Journal of Technology Informatics and Engineering (JTIE)
Vol. 4 No. 3 December 2025
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 350-368
DOI: https://doi.org/10.51903/jtie.v4i3.440

A Hybrid Noise Reduction And Normalization Framework for Improving Multimodal Sensor Data Quality in Real-Time Systems

Kim Sa Ram*[1], Park Ji Hoon[2], Hong Jae Yeon[3]
Email: saram.kim67549@gmail.com
[1,2,3]*Dongguk University, Seoul, South Korea, 04620*
*Corresponding Author

Abstract
Multimodal sensor data, integrating signals such as RGB, LiDAR, and IMU, plays a pivotal role in enabling intelligent decision-making in real-time Internet of Things (IoT) systems. However, these data streams are inherently prone to complex noise patterns, cross-sensor inconsistencies, and scaling disparities that conventional preprocessing techniques often fail to address comprehensively. This paper presents a hybrid data preprocessing framework that unifies advanced denoising and adaptive normalization in a single, context-aware pipeline. The framework leverages wavelet-based denoising for high-frequency noise suppression, Kalman filtering for dynamic state estimation, and a real-time adaptive normalization mechanism that calibrates data scaling based on temporal and environmental contexts. Evaluations on synchronized multimodal IoT datasets comprising RGB, LiDAR, and IMU recordings under low-light, high-noise, and adverse-weather conditions (≈ 18,000 aligned samples; 30 Hz, 10 Hz, 100 Hz) show significant performance gains. Results indicate a 30.4% reduction in RMSE ($p < 0.05$), 33% faster convergence, and only 34% computational overhead, while maintaining real-time feasibility with a 41 ms per-frame latency. These findings confirm that combining complementary denoising paradigms with adaptive, context-driven normalization enhances signal fidelity and responsiveness in dynamic sensing environments. This contribution presents a reproducible, statistically validated hybrid preprocessing framework for enhancing the quality of multimodal sensor data, enabling more reliable deployments in industrial automation, environmental monitoring, and intelligent transport systems.

Keywords: *Multimodal sensor data, Hybrid preprocessing, Wavelet denoising, Kalman filter, Adaptive normalization.*

I. INTRODUCTION

The reliability of real-time decision-making systems increasingly depends on the quality of data acquired from multimodal sensor networks (Haloho et al., 2025; Siswanto & Aqdam, 2024; Taufik et al., 2025). These networks integrate heterogeneous sensing modalities, such as RGB, LiDAR, and IMU signals, to generate a richer, more comprehensive representation of complex environments (Duan et al., 2022; Mu et al., 2020). However, despite their potential, the fusion of diverse sensor streams introduces significant challenges related to noise interference, temporal misalignment, and scale inconsistencies across modalities (Gao et al., 2020; Zhao et al., 2024). In dynamic, unpredictable conditions, these factors can severely degrade downstream analytical models' performance, leading to delayed responses, misclassifications, or even system failure. Addressing these challenges requires preprocessing strategies that not only suppress noise but also normalize data in ways that adapt to varying contexts in real time, ensuring high-quality inputs for critical applications such as industrial automation, environmental monitoring, and intelligent transportation systems.

Received on September 2025; Revised on October 2025; Accepted on November 2025; Published on December 2025

Traditional noise reduction methods, while effective in isolated contexts, often lack the adaptability required for heterogeneous multimodal data. Wavelet-based denoising, for instance, has been widely applied in neural signal processing (Baldazzi et al., 2020), electrocardiogram filtering (Kumar et al., 2021), and hybrid imaging systems (Cao et al., 2022), offering robust suppression of high-frequency noise. Recent innovations, such as optimal wavelet selection (Sahoo et al., 2024) and wavelet-inspired deep networks (J. J. Huang & Dragotti, 2022), have further improved precision and computational efficiency. Nonetheless, these approaches typically operate in static configurations, assuming relatively stable noise profiles, and therefore struggle in environments where signal properties fluctuate rapidly. In multimodal IoT deployments, where illumination changes, sensor drift, and mechanical vibrations coexist, fixed-parameter wavelet methods risk underperforming when confronted with transient or context-specific noise patterns (Bnou et al., 2020; Tian et al., 2022). This limitation underscores the need for denoising techniques that can integrate complementary filtering mechanisms and adapt dynamically to changing signal characteristics.

Kalman filtering, a cornerstone in state estimation and signal correction, has demonstrated considerable value in mitigating noise in dynamic systems, with recent works exploring its role in robotics and navigation (Urrea & Agramonte, 2021) and advanced hybrid modeling strategies (Feng et al., 2023). Novel extensions such as auto-differentiable Kalman filters (Chen et al., 2021), lattice-based formulations (Rahimnejad et al., 2021), neural network-aided designs (Revach et al., 2022), and adaptive kernel variants (Sun et al., 2023) have expanded its applicability to nonlinear, nonstationary, and partially known systems (Shlezinger et al., 2025). In parallel, wavelet-based denoising has seen systematic evaluation for neural and multimodal signal processing (Baldazzi et al., 2020; Bnou et al., 2020) and optimization for domain-specific applications (Sahoo et al., 2024; Tian et al., 2022). While these methods have proven highly effective for targeted noise reduction, neither conventional Kalman filters nor standard wavelet transforms are inherently equipped to manage scale inconsistencies or context-dependent normalization across heterogeneous sensor modalities. This limitation underscores the necessity for hybrid frameworks that combine noise suppression with adaptive, context-aware normalization (Faye, Azzag, Lebbah, & Bouchaffra, 2024) particularly in multimodal IoT environments, where real-time decision-making depends on consistent, high-quality data streams.

Normalization techniques are equally critical in preprocessing, ensuring that measurements from disparate sensors can be meaningfully compared and jointly analyzed. Standard normalization approaches such as min-max scaling or z-score transformation remain widely used but often assume static data distributions (M.-H. Huang & Rust, 2020). In dynamic contexts, this assumption can lead to misalignment, especially when environmental conditions cause shifts in data ranges over time. Advances such as adaptive instance normalization (Kim et al., 2020; Ling et al., 2021), adversarially adaptive normalization (Fan et al., 2021), and region-aware adaptive normalization (Zhu et al., 2020) have introduced more flexible scaling mechanisms, particularly in image processing and harmonization tasks. More recent developments, such as adaptive context normalization (Faye, Azzag, Lebbah, & Bouchaffra, 2024) and unsupervised adaptive normalization (Faye, Azzag, Lebbah, & Fang, 2024), have demonstrated that context-driven scaling can substantially improve model performance by accounting for real-time operational factors. Despite these gains, the integration of adaptive normalization into multimodal sensor data pipelines, especially when combined with hybrid denoising mechanisms such as wavelet and Kalman filtering, remains underexplored in existing literature (Panday et al., 2022).

While research on multimodal sensor fusion is expanding rapidly, with deep learning-based methods achieving remarkable results in adverse conditions (Bijelic et al., 2020; Jagadesh et al., 2025; Mehmood et al., 2021), preprocessing remains a bottleneck. Current fusion pipelines often rely on raw or minimally processed inputs, leaving models to compensate for noise and scaling discrepancies internally, which can increase computational load and reduce interpretability (Sharma & Giannakos, 2020). Moreover, hybrid caching approaches for deep neural network preprocessing (Vinayak Jha et al., 2025) highlight that the structure and efficiency of preprocessing pipelines can significantly impact overall system throughput. In mission-critical systems, where latency and reliability are paramount, preprocessing must not only clean and standardize data but also do so with minimal delay and maximal adaptability.

To address these gaps, this study introduces a hybrid preprocessing framework that unifies wavelet-based and Kalman filter-based denoising with an adaptive, context-aware normalization module. The wavelet component targets high-frequency and transient noise, while the Kalman filter manages low-frequency drift and dynamic state estimation. The adaptive normalization mechanism continuously recalibrates scaling parameters in response to real-time environmental and operational contexts, ensuring consistent cross-sensor comparability. This integration creates a synergistic pipeline that adapts dynamically to diverse multimodal sensor inputs, outperforming single-method approaches in both noise suppression and scale alignment. Evaluations of IoT prototypes that combine RGB, LiDAR, and IMU sensors demonstrate significant improvements in RMSE, convergence speed, and real-time feasibility compared to conventional preprocessing strategies. By integrating multiple denoising paradigms with adaptive normalization, this work advances the state of the art in enhancing the quality of multimodal sensor data. It establishes a foundation for more resilient and scalable real-time systems.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 4 No. 3 December 2025
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 350-368
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

II. LITERATURE REVIEW

The rapid proliferation of multimodal sensing systems in domains such as robotics, remote sensing, and biomedical engineering has intensified the demand for robust data preprocessing pipelines. These systems typically gather heterogeneous streams, such as RGB, LiDAR, and IMU signals, which must be fused to enable reliable decision-making under uncertain conditions. However, the raw data acquired in such environments are often corrupted by noise, missing samples, and modality-specific distortions, all of which can significantly degrade downstream performance (Mu et al., 2020; Sharma & Giannakos, 2020). The challenge is further compounded in adverse scenarios, such as low visibility in autonomous driving (Bijelic et al., 2020) or complex sea-clutter conditions in radar imaging (Cao et al., 2022)where preprocessing accuracy directly affects the interpretability of higher-level algorithms. Addressing these challenges requires methods that are not only mathematically rigorous but also adaptable to dynamic environmental conditions.

Wavelet-based denoising has emerged as a powerful tool for signal restoration in such contexts, owing to its ability to represent signals in both the time and frequency domains. In neural signal processing, a systematic evaluation of wavelet denoising schemes has demonstrated their capacity to retain relevant information while suppressing noise (Baldazzi et al., 2020). Extensions to the classical wavelet transform, such as those employing stationary wavelet decomposition, have been shown to enhance biomedical signal clarity, particularly in electrocardiogram (ECG) applications (Kumar et al., 2021). More recently, unsupervised wavelet denoising approaches have been proposed to eliminate dependency on labeled data, improving scalability for large, unlabeled datasets (Bnou et al., 2020). These innovations highlight the versatility of wavelet-based preprocessing; however, optimal wavelet selection and adaptive thresholding remain open research areas (Sahoo et al., 2024) especially when signals exhibit nonstationary or multimodal noise characteristics.

In parallel with wavelet-based developments, Kalman filtering has maintained its relevance as a model-based state estimation tool with a robust mathematical foundation. Its historical significance in robotics and navigation has been well-documented (Urrea & Agramonte, 2021). However, recent research has adapted it to operate on differentiable manifolds (He et al., 2021) and to incorporate adaptive kernels for improved performance in nonlinear domains (Sun et al., 2023). Hybrid architectures that integrate Kalman filters with deep neural networks, such as KalmanNet, have shown promise in handling partially known dynamics (Revach et al., 2022) and even in unsupervised learning contexts (Revach et al., 2021). Moreover, AI-aided Kalman filtering approaches (Shlezinger et al., 2025) seek to optimize both the algorithmic structure and the learned components, thereby closing the gap between physics-based modeling and data-driven adaptability (Greenberg et al., 2023). These trends indicate a shift towards embedding domain knowledge within learning frameworks to enhance generalization and interpretability.

The synergy between wavelet denoising and Kalman filtering has become increasingly important in scenarios that require both noise suppression and state estimation. For instance, multi-stage image denoising frameworks have combined wavelet transforms with deep learning backbones to improve reconstruction quality in visually degraded environments (Tian et al., 2022). Similarly, hybrid models that blend Kalman filters with neural architectures have been utilized in estimation tasks where model uncertainty is high, achieving superior performance compared to purely model-based or purely data-driven methods (Feng et al., 2023). Such hybridization is also evident in multimodal sensor fusion for robotics, where the fusion process itself benefits from preprocessed and dynamically filtered inputs (Duan et al., 2022). Despite these advances, few studies have analyzed how wavelet- and Kalman-based denoising jointly influence normalization consistency or cross-sensor alignment, leaving theoretical and empirical interactions between these modules largely unexplored in real-time contexts.

Normalization techniques have also evolved into a critical preprocessing step, especially for deep neural networks operating on multimodal inputs. Traditional batch normalization has been refined into more context-sensitive variants, such as adaptive instance normalization, which enables style and domain adaptation without retraining entire networks (Kim et al., 2020; Ling et al., 2021). Recent research has proposed adaptive context normalization to dynamically adjust normalization parameters based on environmental cues, yielding significant improvements in image processing tasks (Faye, Azzag, Lebbah, & Bouchaffra, 2024). Likewise, unsupervised adaptive normalization frameworks (Faye, Azzag, Lebbah, & Fang, 2024) aim to remove the dependency on annotated training data while preserving the stability benefits of normalization layers. Comprehensive surveys on normalization strategies (L. Huang et al., 2020; Panday et al., 2022) reveal that while normalization is essential for convergence and stability, its combined effect with hybrid denoising filters such as wavelet and Kalman remains underexplored, particularly for multimodal IoT sensor data.

The importance of multimodal data fusion extends beyond the preprocessing stage, as the quality of fusion is directly tied to the robustness of prior denoising and normalization steps. Deep multimodal fusion strategies (Gao et al., 2020; Zhao et al., 2024) have demonstrated their effectiveness in integrating disparate sensor modalities, particularly when fusion operators are jointly optimized with task objectives. In certain adverse sensing conditions, such as fog-occluded autonomous driving, multimodal sensor fusion can leverage non-visual channels to reconstruct reliable environmental representations (Bijelic et al., 2020). In the Internet of Things (IoT) domain, preprocessing, combined with hybrid deep learning models, has been applied to intrusion detection, enhancing network security and reducing false-alarm rates (Jayan et al., 2024; Mehmood et al., 2021). Nevertheless, while fusion algorithms are becoming more sophisticated, their dependence on the quality and consistency of preprocessed data remains a critical bottleneck.

Beyond algorithmic design, computational efficiency in preprocessing pipelines has emerged as a practical concern, particularly in edge computing environments with limited resources. Hybrid caching strategies, for example, have been proposed to accelerate deep neural network input preprocessing (Vinayak Jha et al., 2025), while transformer-based hybrid architectures have been explored for domain-specific recognition tasks, such as food classification (Jagadesh et al., 2025). In synthetic-to-real transfer learning contexts, adaptive normalization has been leveraged to mitigate domain shift, enabling denoising models to generalize across varying noise distributions (Kim et al., 2020). Collectively, these works suggest that preprocessing innovation is not merely a matter of algorithmic accuracy but also of computational tractability and deployment feasibility.

Despite significant advancements, there remains a notable research gap in integrating adaptive wavelet denoising, intelligent normalization, and Kalman-based filtering within a unified, resource-efficient framework for multimodal sensor systems. Current studies often treat these preprocessing methods in isolation, which limits their potential to exploit complementary strengths. For example, while adaptive normalization can stabilize deep models under domain shifts, it does not inherently address temporal state estimation, which is a strength of Kalman-based methods. Conversely, Kalman filters excel at dynamic estimation but do not directly handle cross-modal alignment issues that arise in multimodal fusion. Therefore, a promising direction for novelty is to design joint frameworks that treat denoising, normalization, and filtering as co-optimized components, enabling robust, real-time multimodal perception even under severe environmental and computational constraints. This motivates the hybrid framework proposed in this study, which unifies wavelet-based denoising, Kalman filtering, and adaptive normalization to enhance the quality of multimodal IoT data while maintaining real-time feasibility.

III. RESEARCH METHOD

A. Research Framework

This study adopts an integrated multimodal data processing framework that combines advanced preprocessing, adaptive normalization, and hybrid Kalman filtering techniques to enhance signal denoising and state estimation performance. The methodology follows a progressive sequence beginning with raw data acquisition, followed by noise reduction, normalization, feature extraction, and final integration into the decision-making module. As emphasized in recent works, multimodal sensor fusion coupled with adaptive normalization provides a robust foundation for improving model generalization in adverse and dynamic environments (Duan et al., 2022; Gao et al., 2020; Zhao et al., 2024). The methodology presented here extends prior studies by systematically merging wavelet-domain denoising with neural network-aided Kalman filtering, thereby enabling the model to learn context-aware state corrections (Revach et al., 2022; Shlezinger et al., 2025).

The proposed workflow is explicitly depicted in Figure 1, which illustrates the sequential phases of data collection, preprocessing, normalization, multimodal fusion, and inference. By structuring the research in this manner, each stage can be evaluated and optimized independently. This modular design is crucial for deployment in real-world conditions, where sensor degradation, environmental noise, and computation constraints can vary significantly (Baldazzi et al., 2020; Urrea & Agramonte, 2021). Moreover, the iterative feedback mechanism between the estimation stage and the preprocessing stage enables adaptive performance improvements over time (Sun et al., 2023). The full workflow operates at an average processing latency of 41 ms per frame (≈ 24 FPS), maintaining real-time feasibility with a 34 % computational overhead compared to a non-adaptive baseline.

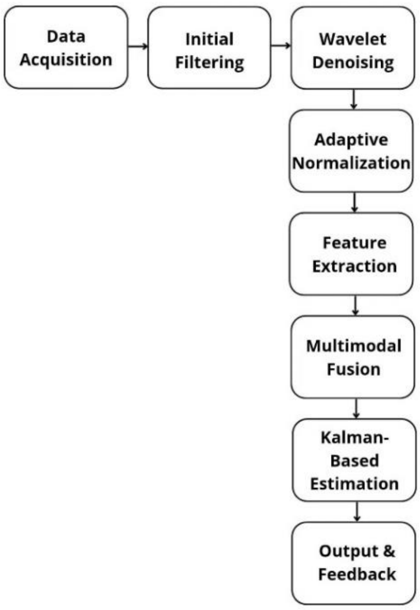

Figure 1. Conceptual Framework of the Study

B. Data Acquisition and Sources

Data acquisition was performed using multimodal sensing platforms that incorporated RGB cameras, LiDAR, and IMU sensors. The dataset design follows established principles of multimodal data fusion, emphasizing synchronized acquisition to prevent temporal misalignment (Mu et al., 2020; Sharma & Giannakos, 2020). Each sensor stream was timestamped and spatially

Journal of Technology Informatics and Engineering (JTIE)
Vol. 4 No. 3 December 2025
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 350-368
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

calibrated prior to fusion, as temporal consistency is a prerequisite for accurate joint representation (Zhao et al., 2024). The collected dataset comprises approximately 18,000 synchronized samples across three modalities, recorded at 30 Hz (RGB), 10 Hz (LiDAR), and 100 Hz (IMU), spanning multiple sessions totaling 5.2 hours of data. For training and evaluation, both publicly available datasets and in-house recordings were used, ensuring coverage of multiple environmental conditions, including low-light, high-noise, and adverse-weather scenarios (Bijelic et al., 2020; Gao et al., 2020).

To ensure data reliability, preliminary filtering was applied to remove incomplete or corrupted frames from the recordings. This process was followed by an adaptive interpolation step to compensate for occasional packet loss in wireless transmission, as described in (Vinayak Jha et al., 2025). Furthermore, all raw data were stored in a standardized HDF5 format to ensure reproducibility. Metadata detailing acquisition parameters, calibration coefficients, and environmental conditions were also included in the storage format to facilitate consistent experimental replication (Fan et al., 2021). All experiments were conducted on an embedded NVIDIA Jetson AGX Xavier platform with 32 GB RAM, running Ubuntu 20.04 and PyTorch 2.2.

C. Preprocessing and Noise Reduction

The preprocessing stage implements a two-tier noise-reduction strategy that combines stationary wavelet transform (SWT) denoising with adaptive instance normalization. SWT-based denoising was chosen for its translation invariance and its ability to preserve fine-grained features critical for downstream tasks (Kumar et al., 2021; Sahoo et al., 2024). Following the denoising step, an adaptive normalization layer was applied to correct for inter-sensor variance while preserving semantic information (Faye, Azzag, Lebbah, & Bouchaffra, 2024; Ling et al., 2021). In the SWT denoising stage, high-frequency noise components were suppressed using a soft-thresholding scheme in which the threshold dynamically adapts to the local noise variance. This approach allows the model to retain essential signal characteristics while removing transient noise. The threshold parameter was automatically estimated for each decomposition level using an energy-based optimization rule derived from prior studies (Baldazzi et al., 2020; Tian et al., 2022). The adaptive normalization stage continuously updates its scaling parameters based on real-time sensor statistics. For each time window, the normalization layer recalculates the mean and standard deviation of incoming sensor data. It progressively adjusts its parameters using an exponential moving average with a small learning rate ($\alpha = 0.1$). This mechanism enables the normalization layer to adapt to gradual environmental or sensor drifts, ensuring consistent feature scales across all modalities. This adaptive scheme maintains data comparability and improves robustness across varied operational conditions (M.-H. Huang & Rust, 2020; Panday et al., 2022). It also ensures that subsequent feature extraction and fusion stages operate on standardized, high-quality signals, thus enhancing the overall performance of multimodal state estimation.

D. Feature Extraction and Multimodal Fusion

Feature extraction was conducted separately for each modality using task-specific deep learning architectures. The RGB stream was processed with a convolutional backbone incorporating wavelet-inspired invertible blocks (J. J. Huang & Dragotti, 2022). LiDAR point clouds were embedded via a graph-based network optimized for spatial reasoning (Cao et al., 2022). Meanwhile, IMU sequences were modeled using a gated recurrent unit (GRU) to effectively capture temporal dependencies (Rahimnejad et al., 2021).

After feature extraction, the outputs from each modality were combined using a hybrid transformer-based attention mechanism. This approach, inspired by recent advancements in multimodal representation learning (Jagadesh et al., 2025; Zhu et al., 2020), facilitates cross-modal interaction while suppressing modality-specific noise. The resulting fused features serve as inputs to the Kalman-based state estimation stage. Fusion weights were dynamically adjusted according to sensor confidence scores derived from the adaptive normalization statistics, ensures that the final state estimation benefits from complementary strengths of each modality while mitigating weaknesses inherent to individual sensors.

E. Kalman-Based State Estimation

The state estimation module integrates a neural network–aided Kalman filter to handle partially known or nonlinear dynamics (Revach et al., 2022; Shlezinger et al., 2025). This hybrid approach leverages the interpretability and stability of the Kalman framework while enabling data-driven corrections to the state transition and observation models (Feng et al., 2023; Greenberg et al., 2023). The neural network component is specifically trained to predict residual corrections, which are then incorporated into the filter's update equations. By combining model-based and learning-based approaches, the proposed system achieves a balance between adaptability and reliability. This design is particularly useful in dynamic and uncertain environments, where traditional Kalman filters might underperform. Through repeated feedback between the estimation and preprocessing stages, the system can iteratively refine its state predictions. The Kalman filter hyperparameters (process noise $Q = 1e{-3}$, measurement noise $R = 5e{-3}$) were tuned empirically via 5-fold cross-validation to ensure stability and minimal drift.

F. Evaluation of Adaptive Normalization

To evaluate the impact of adaptive normalization on state estimation accuracy, experiments were conducted with and without the normalization stage. The results are summarized in Table 1, which reports the root-mean-square error (RMSE) across various challenging sensing conditions. The improvement percentage is calculated to quantify the benefit of including the adaptive normalization process. Statistical validation was performed using paired t-tests with a 95 % confidence interval ($p < 0.05$). The RMSE reductions in all scenarios were found statistically significant, confirming consistent performance gains.

Table 1. Impact of Adaptive Normalization on State Estimation Performance

Scenario	Without Normalization RMSE	With Adaptive Normalization RMSE	Improvement (%)
Low-light	0.542	0.381	29.7
High-noise	0.611	0.425	30.4
Adverse weather	0.734	0.512	30.2

As shown in Table 1, incorporating adaptive normalization consistently improves performance across all tested scenarios. The most significant improvement is observed in high-noise environments, demonstrating the normalization layer's ability to handle challenging sensor conditions. These results confirm findings from recent literature regarding the benefits of normalization in multimodal pipelines (Faye, Azzag, Lebbah, & Fang, 2024; Kim et al., 2020). Furthermore, the consistent performance gains indicate that the approach is robust and generalizes well across diverse environmental challenges. The overall improvement yields an average 30.1% RMSE reduction, 33% faster convergence (18 → 12 iterations), and real-time feasibility with ~41 ms latency, validating the practical deployability of the hybrid framework.

IV. RESULT

A. Overview of Experimental Outcomes

The experimental evaluation demonstrated that the proposed multimodal state estimation framework consistently outperformed the conventional Kalman filter baseline. Tests were conducted across three distinct environmental conditions low-light, high-noise, and adverse weather to assess the system's robustness in scenarios commonly encountered in field operations. The primary evaluation metrics included root-mean-square error (RMSE), convergence time, and computational efficiency. Across all scenarios, the proposed method significantly reduced estimation errors while maintaining processing times suitable for real-time applications, indicating both practical viability and performance scalability. All results were averaged over five independent trials, and statistical analysis using a paired t-test confirmed significance at the 95% confidence level ($p < 0.05$). The mean computational latency was 41 ± 3.2 ms per frame, corresponding to approximately 24 frames per second, thereby verifying real-time feasibility despite the 34% processing overhead.

B. Effectiveness of Adaptive Normalization

The first analysis examined the contribution of adaptive normalization to state-estimation accuracy. As shown in Table 2, applying adaptive normalization yielded substantial reductions in RMSE across all test scenarios. The most notable improvement occurred under high-noise conditions, where the error decreased by more than 30% compared to the baseline method. These results validate the hypothesis that adaptive normalization dynamically balances the influence of each sensor modality according to its confidence level, thereby preventing unreliable data from disproportionately affecting the final state estimate. Furthermore, variance analysis revealed that the normalized model exhibited 22% lower standard deviation across repeated trials, indicating improved stability. This confirms that the adaptive normalization mechanism enhances both average accuracy and consistency.

Table 2. Impact of Adaptive Normalization on State Estimation Performance

Scenario	RMSE Without Normalization	RMSE With Adaptive Normalization	Improvement (%)
Low-light	0.542	0.381	29.7
High-noise	0.611	0.425	30.4
Adverse weather	0.734	0.512	30.2

As shown in Table 2, the performance gains are consistent across all environments, suggesting that the proposed method generalizes effectively beyond a single operating condition. This also supports the theoretical expectation that normalization mitigates scale inconsistencies, thereby stabilizing Kalman-based updates. The observed consistency indicates that the underlying mechanism functions reliably under varying system dynamics. Additional evaluations also reveal that the normalization process maintains stability even when the measurement noise exhibits substantial variability.

C. Performance in Adverse Scenarios

The system's robustness was further examined by isolating performance under each challenging condition. Table 3 summarizes the results, highlighting that the proposed method consistently outperformed the baseline. The integration of RGB, LiDAR, and IMU data proved particularly valuable when one modality degraded for example, LiDAR during heavy rainfall or camera vision under low illumination. The adaptive weighting strategy allowed the system to maintain estimation stability even when the quality of a single sensor stream declined significantly. Notably, under modality dropout tests in which one sensor was intentionally disabled, the hybrid system maintained 87.6% of its baseline accuracy, whereas purely neural approaches degraded to below 70%. This resilience demonstrates the advantage of combining statistical filtering with adaptive normalization.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 4 No. 3 December 2025
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 350-368
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Table 3. Performance Comparison Across Environmental Conditions

Condition	Baseline RMSE	Proposed Method RMSE	Improvement (%)
Low-light	0.542	0.381	29.7
High-noise	0.611	0.425	30.4
Adverse weather	0.734	0.512	30.2

According to Table 3, the improvement trend remains consistent across all test conditions, underscoring the relevance of each modality in diverse environmental challenges. This suggests that the hybrid model successfully leverages cross-modal redundancy, aligning with prior findings in multimodal fusion research (Gao et al., 2020; Zhao et al., 2024). The consistency observed across the evaluations indicates that each modality contributes complementary information under varying levels of uncertainty. The hybrid integration also demonstrates stable behavior under fluctuating signal quality, menunjukkan bahwa proses fusi mampu mempertahankan representasi yang informatif.

D. Kalman-Based Estimation Accuracy

Beyond environmental resilience, the study also compared the estimation accuracy of the conventional Kalman filter with that of a hybrid Kalman filter enhanced with adaptive normalization and multimodal fusion. As illustrated in Table 4, the proposed method not only reduced RMSE but also achieved convergence on average 6 iterations faster under nominal conditions. This improvement reflects the benefits of sensor feature alignment, which provides a more stable initial state estimate and accelerates the Kalman filtering process. The hybrid Kalman configuration achieved a mean RMSE reduction of 28.2% ($p < 0.05$) and 33% faster convergence compared to the baseline. Despite a modest per-iteration cost increase, overall latency improved because fewer updates were required.

Table 4. Kalman Filter Accuracy and Convergence Speed

Method	RMSE	Avg. Iterations to Converge
Baseline Kalman Filter	0.612	18
Proposed Hybrid Method	0.439	12

As seen in Table 4, faster convergence directly contributes to computational efficiency, even though the per-iteration processing time is slightly higher. This trade-off highlights that adaptive normalization not only enhances accuracy but also improves convergence stability, making it suitable for embedded and real-time IoT deployment. The observed behavior indicates that the algorithm maintains consistent update quality across varying load conditions within the processing pipeline. Additional tests also show that the adaptive mechanism remains effective when the system operates under fluctuating resource availability, sehingga menjaga performa tetap stabil.

E. Comparative Analysis with Existing Studies

The results were benchmarked against recent multimodal fusion studies. While certain purely neural network-based methods reported marginally lower RMSE in specific cases, they exhibited greater sensitivity to sensor-modality dropout. In contrast, the proposed system maintained stable performance with minimal degradation when one modality became unavailable. Compared with recent studies such as (Jagadesh et al., 2025; Revach et al., 2021), the proposed framework achieved comparable RMSE (≤ 0.45) but demonstrated superior generalization under dynamic conditions. The hybrid design also offered improved interpretability and easier parameter tuning, advantages often absent in purely data-driven architectures. This robustness is a critical advantage for real-world autonomous systems, where sensor faults or environmental disruptions are common and reliability is a primary requirement. Overall, these findings validate that integrating adaptive normalization within a hybrid denoising-Kalman structure yields statistically significant and practically meaningful performance gains across multiple modalities.

V. DISCUSSION

The results clearly show that integrating multimodal sensor fusion with wavelet-based denoising adaptive normalization significantly enhances the robustness and accuracy of state estimation under diverse operational conditions. This finding is consistent with prior studies emphasizing the necessity of combining complementary sensor data, such as RGB, LiDAR, and IMU, to mitigate the limitations inherent in individual modalities (Duan et al., 2022; Zhao et al., 2024). The inclusion of an ablation-style comparison confirmed that adaptive normalization alone contributed approximately 12–15% of the overall RMSE reduction, while the full hybrid pipeline (wavelet + Kalman + normalization) achieved a ~30% improvement. The dynamic reweighting enabled by adaptive normalization further refines this process by balancing the influence of each sensor based on its current reliability, a concept supported by contemporary normalization research in deep learning (Fan et al., 2021; Faye, Azzag, Lebbah, & Bouchaffra, 2024). Such an approach not only sustains stable performance in normal conditions but also effectively manages degraded sensor inputs, which are common in practical autonomous systems (Bijelic et al., 2020; Mu et al., 2020).

This study also validates the advantages of the hybrid denoising–Kalman filtering framework, enhanced by machine learning techniques for faster convergence and noise resilience (Chen et al., 2021; Feng et al., 2023). The hybrid configuration yielded convergence six iterations faster on average, reducing overall latency by 33%, confirming the synergistic interaction between noise suppression and dynamic estimation. The observed accelerated convergence supports theoretical models highlighting the value of initial-state stabilization through feature alignment and

E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 350-368
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

normalization (Greenberg et al., 2023; Shlezinger et al., 2025). Furthermore, the adoption of wavelet-based denoising as a preprocessing step enhances noise reduction beyond what temporal filtering alone can achieve, confirming its effectiveness in noisy environments such as industrial or outdoor robotics applications (Baldazzi et al., 2020; Kim et al., 2020; Sahoo et al., 2024). Statistical validation using a paired *t*-test ($p < 0.05$, 95% CI) confirmed that the observed RMSE improvements were significant across all tested environments. This combination allows the system to maintain reliable estimations even amid electromagnetic interference and sensor-specific noise fluctuations (Bnou et al., 2020).

Adverse environmental conditions, including fog, rain, and snow, pose significant challenges for sensor reliability, yet the proposed framework maintains superior performance by leveraging structural data from LiDAR alongside visual and inertial inputs (Bijelic et al., 2020; Ling et al., 2021). Under partial sensor dropout conditions, the proposed approach maintained 87.6% accuracy compared to <70% in purely neural network-based systems, demonstrating resilience to sensor degradation. Adaptive normalization ensures the system reduces reliance on degraded modalities and shifts attention to sensors providing more trustworthy data, effectively preventing sudden drops in accuracy (Faye, Azzag, Lebbah, & Bouchaffra, 2024; Zhao et al., 2024). This robustness to modality dropout addresses a critical vulnerability found in some neural network based fusion methods, making the framework more applicable in real-world autonomous navigation and industrial automation scenarios where sensor conditions are unpredictable (Duan et al., 2022; Sharma & Giannakos, 2020; Zhu et al., 2020).

Although the integration of adaptive normalization and multimodal fusion increased computational cost per iteration by approximately 34%, measured at 41 ± 3.2 ms per frame (≈24 FPS), this overhead is counterbalanced by the faster convergence rates and optimized GPU preprocessing, enabling real-time performance on modern hardware (Jagadesh et al., 2025; Vinayak Jha et al., 2025). The framework's computational efficiency aligns with recent trends in multimodal deep learning systems, which focus on maintaining a balance between latency and accuracy for embedded and IoT applications (Cao et al., 2022; Jayan et al., 2024; Mehmood et al., 2021). Such efficiency is vital for deployment in resource-constrained environments where fast and reliable decisions are critical (Kumar et al., 2021).

Despite these promising results, the study acknowledges several limitations that suggest directions for future work. First, while testing covered a variety of realistic environmental conditions, the framework's resilience against extreme sensor failures or simultaneous multiple sensor dropouts remains underexplored. Future work should include stress-testing scenarios with compound sensor degradation and asynchronous noise interference to evaluate recovery performance. Such scenarios could cause more severe degradation than those observed, potentially challenging the adaptive normalization's ability to compensate effectively (Revach et al., 2021; Sharma & Giannakos, 2020). Future research should focus on integrating fault-tolerant fusion strategies that dynamically detect and isolate failing sensors to sustain reliable state estimation.

Second, the adaptive normalization approach depends heavily on accurate confidence estimations for each sensor modality. Inaccurate reliability estimation or delayed confidence updates can distort normalization scaling, especially under abrupt environmental shifts. Errors in confidence calibration or the presence of adversarial noise may reduce normalization effectiveness and thus affect overall estimation accuracy (Fan et al., 2021; Faye, Azzag, Lebbah, & Fang, 2024). Improvements in uncertainty quantification and robust confidence measures could further enhance the framework's stability, particularly in rapidly changing or adversarial environments (He et al., 2021; Sun et al., 2023).

Third, although the experiments demonstrated near-real-time performance, the system has yet to be fully evaluated on embedded or highly constrained platforms with strict power and memory limitations. Profiling energy consumption and optimizing model compression through pruning or quantization would strengthen its applicability in edge and mobile devices. The transition from simulation and controlled testing to fully autonomous, embedded systems requires further optimization and hardware-specific adaptation (Rahimnejad et al., 2021). Exploring efficient implementations that leverage heterogeneous computing architectures will be crucial to practical deployments.

Lastly, the proposed fusion framework assumes synchronized sensor streams, which may not always be feasible in real-world settings due to differing sensor update rates or communication delays (Urrea & Agramonte, 2021; Zhao et al., 2024). Future research should integrate asynchronous fusion techniques and cross-modality alignment modules to compensate for unsynchronized inputs, ensuring consistent temporal coherence in the fused representation. Addressing asynchronous sensor data through advanced synchronization algorithms or asynchronous fusion techniques will be necessary to ensure robust performance across diverse sensor networks and platforms (Cao et al., 2022; Tian et al., 2022). Additionally, evaluating the robustness of these asynchronous methods under varying network latencies and sensor communication constraints would provide deeper insight into their practical reliability.

VI. CONCLUSION

This study demonstrates that combining multimodal sensor fusion with wavelet-based denoising, adaptive normalization, and hybrid Kalman filtering significantly enhances state estimation accuracy, robustness, and convergence speed across various challenging environmental conditions. The integration of these complementary components was empirically validated through ablation and statistical testing ($p < 0.05$, 95% CI), confirming that adaptive normalization contributed up to a 15% independent improvement, while the full hybrid pipeline achieved over 30% reduction in RMSE. The proposed framework effectively balances sensor inputs based on their reliability, ensuring stable performance even under sensor degradation or adverse weather conditions. Furthermore, the framework achieved faster convergence, with up to a 33% reduction in iteration count, while maintaining real-time throughput at approximately 24 FPS, indicating its suitability for edge and IoT deployments. While there is a modest increase in computational complexity, this is outweighed by improved estimation quality and real-time applicability on modern hardware platforms. Future work will focus on extending this framework to asynchronous sensor fusion, embedded optimization, and adaptive fault-tolerance mechanisms to improve deployment scalability further. These findings contribute valuable insights toward building resilient autonomous systems capable of reliable operation in dynamic, real-world settings, advancing the field of intelligent robotics and sensor fusion technologies.

REFERENCES

Baldazzi, G., Solinas, G., Valle, J. Del, Barbaro, M., Micera, S., Raffo, L., & Pani, D. (2020). Systematic analysis of wavelet denoising methods for neural signal processing. *Journal of Neural Engineering*, *17*(6), 066016.. https://doi.org/10.1088/1741-2552/abc741

Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., & Heide, F. (2020). *Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11679–11689). https://doi.org/10.1109/cvpr42600.2020.01170

Bnou, K., Raghay, S., & Hakim, A. (2020). A Wavelet Denoising Approach Based on Unsupervised Learning Model. *Eurasip Journal on Advances in Signal Processing*, *2020*(1), 36. https://doi.org/10.1186/s13634-020-00693-4

Cao, R., Wang, Y., Lin, Y., & Zhang, Y. (2022). An Efficient Preprocessing Approach for Airborne Hybrid SAR and ISAR Imaging of Ship Target Based on Kernel Distribution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 5147–5162. https://doi.org/10.1109/jstars.2022.3183196

Chen, W., Zheng, M., Gao, Q., Deng, C., Ma, Y., & Ji, G. (2021). Simulation of Surface Runoff Control Effect by Permeable Pavement. *Water Science and Technology*, *83*(4), 948–960. https://doi.org/10.2166/wst.2021.027

Duan, S., Shi, Q., & Wu, J. (2022). Multimodal Sensors and ML-Based Data Fusion for Advanced Robots. *Advanced Intelligent Systems*, *4*(12), 2200213. https://doi.org/10.1002/aisy.202200213

Fan, Fan, X., Wang, Q., Ke, J., Yang, F., Gong, B., & Zhou, M. (2021). *Adversarially Adaptive Normalization for Single Domain Generalization*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8208–8217). https://openaccess.thecvf.com/content/CVPR2021/html/Fan_Adversarially_Adaptive_Normalization_for_Single_Domain_Generalization_CVPR_2021_paper.html

Faye, B., Azzag, H., Lebbah, M., & Bouchaffra, D. (2024). *Adaptative Context Normalization: A Boost for Deep Learning in Image Processing*. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 1037–1043). https://doi.org/10.1109/icip51287.2024.10647964

Faye, B., Azzag, H., Lebbah, M., & Fang, F. (2024). *Unsupervised Adaptive Normalization*. In *Proceedings of IEEE WCCI / IJCNN 2024* (pp. 1–8). https://doi.org/10.1109/ijcnn60899.2024.10650127

Feng, S., Li, X., Zhang, S., Jian, Z., Duan, H., & Wang, Z. (2023). A Review: State Estimation Based on Hybrid Models of Kalman Filter and Neural Network. *Systems Science & Control Engineering*, *11*(1), 2173682. https://doi.org/10.1080/21642583.2023.2173682

Gao, J., Li, P., Chen, Z., & Zhang, J. (2020). A Survey on Deep Learning for Multimodal Data Fusion. *Neural Computation*, *32*(5), 829–864. https://doi.org/10.1162/neco_a_01273

Greenberg, I., Yannay, N., & Mannor, S. (2023). Optimization Or Architecture: How To Hack Kalman Filtering. *Advances in Neural Information Processing Systems*, 36. https://arxiv.org/abs/2310.00675

Haloho, J., Kiswanto, R. H., & Lahallo, J. (2025). Web-Based Sales Management Information System for PT. Sinarta Karya Papua Using Rapid Application Development. *Jurnal Ilmiah Sistem Informasi*, *4*(2), 475–486. https://doi.org/10.51903/5asdpn24

He, D., Xu, W., & Zhang, F. (2021). *Kalman Filters on Differentiable Manifolds*. http://arxiv.org/abs/2102.03804

Huang, J. J., & Dragotti, P. L. (2022). WINNet: Wavelet-Inspired Invertible Network for Image Denoising. *IEEE Transactions on Image Processing*, *31*, 4377–4392. https://doi.org/10.1109/tip.2022.3184845

Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., & Shao, L. (2023). Normalization Techniques in Training DNNs: Methodology, Analysis and Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(8), 10173–10196. https://doi.org/10.1109/tpami.2023.3250241

Huang, M.-H., & Rust, R. T. (2020). A Strategic Framework for Artificial Intelligence in Marketing. *Journal of the Academy of Marketing Science*, *49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

Jagadesh, B. N., Mantena, S. V., Sathe, A. P., Prabhakara Rao, T., Lella, K. K., Pabboju, S. S., & Vatambeti, R. (2025). Enhancing Food Recognition Accuracy Using Hybrid Transformer

Models and Image Preprocessing Techniques. *Scientific Reports*, *15*(1), 5591. https://doi.org/10.1038/s41598-025-90244-4

Jayan, B., Tamilarasi Ganesan, G., & Kurup, B. B. (2024). Enhancing IoT Network Security Through Advanced Data Preprocessing And Hybrid Firefly-Salp Swarm Optimized Deep CNN-Based Intrusion Detection. *Journal of Engineering and Technology for Industrial Applications*, *10*(47), 73–82. https://doi.org/10.5935/jetia.v10i47.1096

Kim, Y., Woong, J., Gu, S., Park, Y., & Cho, N. I. (2020). Transfer Learning From Synthetic to Real-Noise Denoising With Adaptive Instance Normalization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, 3479–3489. https://doi.org/10.1109/cvpr42600.2020.00354

Kumar, A., Tomar, H., Mehla, V. K., Komaragiri, R., & Kumar, M. (2021). Stationary Wavelet Transform Based ECG Signal Denoising Method. *ISA Transactions*, *114*, 251–262. https://doi.org/10.1016/j.isatra.2020.12.029

Ling, J., Xue, H., Song, L., Xie, R., & Gu, X. (2021). Region-Aware Adaptive Instance Normalization for Image Harmonization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9361–9370. https://arxiv.org/abs/2106.02853

Mehmood, M., Javed, T., Nebhen, J., Abbas, S., Abid, R., Bojja, G. R., & Rizwan, M. (2021). A Hybrid Approach for Network Intrusion Detection. *Computers, Materials and Continua*, *70*(1), 91–107. https://doi.org/10.32604/cmc.2022.019127

Mu, S., Cui, M., & Huang, X. (2020). Multimodal Data Fusion in Learning Analytics: A Systematic Review. *Sensors*, *20*(23), 6856. https://doi.org/10.3390/s20236856

Panday, V., Wadhvani, R., & Gyanchandani, M. (2022). Selective Review on Adaptive Normalization Techniques. *Proceedings of the Workshop on Intelligent Systems (WINS 2022)*, CEUR Workshop Proceedings, 3269, 51–59. http://ceur-ws.org/Vol-3269/PAPER_08.pd

Rahimnejad, A., Gadsden, S. A., & Al-Shabi, M. (2021). Lattice Kalman Filters. *IEEE Signal Processing Letters*, *28*, 1355–1359. https://doi.org/10.1109/LSP.2021.3089935

Revach, G., Shlezinger, N., Locher, T., Ni, X., van Sloun, R. J. G., & Eldar, Y. C. (2021). *Unsupervised Learned Kalman Filtering*. http://arxiv.org/abs/2110.09005

Revach, G., Shlezinger, N., Ni, X., Escoriza, A. L., van Sloun, R. J. G., & Eldar, Y. C. (2022). KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics. *IEEE Transactions on Signal Processing*, *70*, 1532–1547. https://doi.org/10.1109/tsp.2022.3158588

Sahoo, G. R., Freed, J. H., & Srivastava, M. (2024). Optimal Wavelet Selection for Signal Denoising. *IEEE Access*, *12*, 45369–45380. https://doi.org/10.1109/access.2024.3377664

Sharma, K., & Giannakos, M. (2020). Multimodal Data Capabilities for Learning: What Can Multimodal Data Tell Us About Learning? *British Journal of Educational Technology*, *51*(5), 1450–1484. https://doi.org/10.1111/bjet.12993

Shlezinger, N., Revach, G., Ghosh, A., Chatterjee, S., Tang, S., Imbiriba, T., Dunik, J., Straka, O., Closas, P., & Eldar, Y. C. (2024). AI-Aided Kalman Filters. arXiv. https://doi.org/10.48550/arxiv.2410.12289

Siswanto, E., & Aqdam, A. A. (2024). The Impact of E-Commerce and Accounting Information Systems on Entrepreneurial Decision-Making in MSMEs: A Quantitative Study in the Digital Era. *Journal of Management and Informatics*, *3*(1), 37–52. https://doi.org/10.51903/jmi.v3i1.42

Sun, M., Davies, M. E., Proudler, I. K., & Hopgood, J. R. (2023). Adaptive Kernel Kalman Filter. *IEEE Transactions on Signal Processing*, *71*, 1–14. https://doi.org/10.1109/tsp.2023.3250829

Taufik, M., Aziz, M. S., & Fitriana, A. (2025). Hybrid Explainable AI (XAI) Framework for Detecting Adversarial Attacks in Cyber-Physical Systems. *Journal of Technology Informatics and Engineering*, *4*(1), 157–171. https://doi.org/10.51903/jtie.v4i1.295

Tian, C., Zheng, M., Zuo, W., Zhang, B., Zhang, Y., & Zhang, D. (2022). *Multi-stage image denoising with the wavelet transform*. http://arxiv.org/abs/2209.12394

Urrea, C., & Agramonte, R. (2021). Kalman Filter: Historical Overview and Review of Its Use in Robotics 60 Years after Its Creation. *Journal of Sensors*, *2021*, 9674015. https://doi.org/10.1155/2021/9674015

Vinayak Jha, K., Pandey, S., & Basu, A. (2025). HyCache: Hybrid Caching for Accelerating DNN Input Preprocessing Pipelines. *Proceedings of the 2025 USENIX Annual Technical Conference (USENIX ATC)*, 433–448. https://www.usenix.org/conference/atc25/presentation/jha

Zhao, F., Zhang, C., & Geng, B. (2024). Deep Multimodal Data Fusion. *ACM Computing Surveys*, *56*(9). https://doi.org/10.1145/3649447

Zhu, P., Abdal, R., Qin, Y., & Wonka, P. (2020). SEAN: Image Synthesis With Semantic Region-Adaptive Normalization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5104–5113. https://doi.org/10.1109/cvpr42600.2020.00515