IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

Real - Time Air Writing Recognition A Comprehensive Survey

Hemanth S, Goutham Hegde, Richith R Kumar, Vishwanath S, Prashanth J
Department Information Science and Engineering
Global Academy of Technology, Bengaluru, India
hemanth1ga22is062@gamil.com, gouthamashegde1ga23is405@gmail.com, richith1ga22is129@gamil.com
vishwanaths1ga22is188@gamil.com, prashanthj@gat.ac.in

Abstract: *Air-writing recognition has emerged as a promising touch-free interaction method with applications in human–computer interfaces, assistive technologies, authentication, emergency response systems, and virtual/augmented environments. Existing research spans diverse sensing modalities—including webcams, RGB cameras, depth sensors, radar, wearable motion sensors, and even EEG signals—to accurately capture and interpret characters written in free space. Vision-based systems leverage convolutional neural networks and spatio-temporal models for fingertip tracking, gesture segmentation, and unconstrained character-sequence recognition, achieving real-time performance and high accuracy. Sensor-based approaches utilize accelerometers, gyroscopes, and FMCW radar to extract motion trajectories, while recent studies address signal variability using interpolation and time-series modeling. Novel frameworks such as FMHash enable user identification through deep fuzzy hashing, and NeuroAiR introduces EEG-driven air-writing recognition for hands-free interfaces. Literature reviews highlight persistent challenges, including intraclass variation, background noise, inconsistent gesture patterns, and lack of standardized datasets. Overall, advancements across machine learning, deep learning, and multimodal sensing demonstrate rapid progress toward robust, lightweight, and adaptive air-writing systems suitable for next-generation touchless interaction.*

Keywords: *Air-writing*

I. INTRODUCTION

Air-writing recognition has emerged as a powerful touchless interaction technique, enabling users to write characters or gestures in free space without relying on physical input devices. The increasing demand for hygienic, contactless, and intuitive interfaces—especially in the post-COVID technological landscape—has accelerated research in this domain. According to a comprehensive literature review [5], air-writing systems are typically categorized into offline, online, and air-written modalities, each presenting challenges related to variation in writing style, fingertip tracking, lighting conditions, and the absence of standardized datasets. These challenges emphasize the need for more robust and adaptive recognition pipelines.

A variety of sensing technologies have been explored to capture air-writing motions. Vision-based approaches using standard webcams and RGB cameras focus on fingertip or hand-trajectory extraction, leveraging deep learning models such as CNNs and spatio-temporal networks. For instance, [13] proposes a lightweight CNN- based system that effectively solves hand tracking and push-to-write segmentation using only a single 2D camera. Similarly, an unconstrained text-recognition framework based on 3D spatio-temporal convolution was introduced in [11], establishing one of the largest benchmark datasets with over 209,000 video samples of finger-writing.

Sensor-based systems have also gained momentum due to their portability and noise resilience. Depth-camera approaches, such as the trajectory-based system in [10], extract 3D fingertip positions and achieve over 99% accuracy on large datasets. Radar-based solutions, including FMCW systems [15], offer robust alphanumeric gesture recognition under varying environmental conditions. Wearable technologies present another promising direction; a wristband equipped with ionic-hydrogel sensors [3] enables rapid adaptation through self-supervised contrastive learning, achieving high accuracy with minimal labeled data.

Beyond conventional modalities, novel explorations into alternative signal sources have expanded the design space for air-writing systems. EEG-based air-writing, as demonstrated in [7], investigates neural decoding for hands-free character recognition, although recognition rates remain limited due to noise and inter-subject variability. Deep hashing frameworks like FMHash [12] enable user identification by converting handwritten motion patterns into compact fuzzy hash codes, offering new possibilities in authentication and security.

Several practical applications further highlight the versatility of air-writing. An emergency-alert system developed in [6] uses air-written symbols captured through CCTV to automatically trigger SMS or voice alerts via cloud communication services, providing a low-cost safety mechanism. Meanwhile, studies such as [8] demonstrate how interpolation techniques can standardize variable-length motion signals, improving the performance of deep learning models across multiple datasets.

Overall, the collective research illustrates rapid advancement in air-writing recognition technologies. Yet, persistent challenges remain regarding signal variability, environmental interference, user-dependent behaviors, and computational efficiency. Addressing these limitations is essential for enabling reliable, real-time, and accessible air-writing systems for future human–computer interaction, authentication, assistive applications, and immersive AR/VR platforms.

II. BACKGROUND AND RELATED WORK

Air-writing recognition refers to the process of interpreting characters or gestures written in free space using hand or finger movements. It has gained significant attention due to the growing demand for touchless interaction systems, especially in the context of smart devices, virtual environments, and hygienic user interfaces. According to the survey in [5], air-writing is classified alongside offline and online handwritten recognition but introduces unique challenges such as inconsistent writing trajectories, fingertip tracking difficulties, lighting variations, and the absence of physical contact cues. Vision-based systems eliminate the need for touch surfaces, while sensor- based, radar-based, and EEG-based models aim to improve robustness and expand usability. As technology progresses toward immersive AR/VR and hands-free control, air-writing becomes a crucial input modality for next-generation human–computer interaction.

2.1 Evolution of Sensing Technologies

The field of air-writing recognition has progressed through multiple sensing technologies, ranging from simple webcams to sophisticated multimodal sensors. Vision-based methods using 2D RGB cameras have been widely employed due to their accessibility and low cost. For example, [13] introduces a robust webcam-based approach using CNNs combined with skin detection, motion cues, and CamShift tracking to solve both hand detection and writing-segmentation challenges. More advanced camera systems, such as those explored in [11], use spatio- temporal convolution to process large-scale datasets for unconstrained finger-motion recognition in both Korean and English. Depth sensors ([10]) further enhance the fidelity of trajectory capture by offering precise 3D fingertip localization. Complementary approaches using FMCW radar ([15]) and wearable motion sensors ([3]) demonstrate that air-writing can be effectively recognized even without visual input, opening pathways for versatile and environment-independent systems.

2.2 Advances in Machine Learning and Deep Learning Approaches

Machine learning plays a central role in enabling reliable air-writing recognition across different sensing environments. Classical methods, such as dynamic time warping and SVMs, have gradually been replaced by deep learning architectures capable of modeling spatial and temporal dependencies in writing trajectories. Convolutional neural networks (CNNs) have proven effective for both trajectory-based and image-based recognition tasks, as shown by the lightweight architectures in [13] that reduce computational cost without sacrificing accuracy. Spatio-temporal deep networks, such as those presented in [11], process sequential writing movements for real-time character-sequence

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

recognition. In wearable computing, contrastive-learning-based models ([3]) demonstrate rapid domain adaptation, allowing accurate recognition with minimal labeled data. Additionally, specialized frameworks like FMHash ([12]) convert motion patterns into fuzzy binary hash codes for secure user identification, showcasing the growing diversity of deep learning applications in the domain.

2.3 Current Research Trends and Remaining Challenges

Recent studies reveal significant progress but also highlight ongoing challenges that limit large-scale deployment of air-writing systems. Variability in writing behavior, environmental noise, and inconsistent signal lengths continue to obstruct universal model generalization. Interpolation-based solutions in [8] demonstrate how proper signal normalization can drastically improve deep learning performance, especially for sensor-based time-series data. Meanwhile, unconventional modalities such as EEG-based air-writing ([7]) introduce new possibilities for hands-free interfaces but struggle with lower accuracy due to neural noise and user variability. Practical applications also emerge, including emergency-alert systems that identify air-written distress signals via camera feeds ([6]). Despite these advancements, the literature consistently emphasizes the need for standardized datasets, cross-user adaptability, and robust multimodal fusion techniques. These gaps motivate continued research into more accurate,lightweight, and generalizable models capable of supporting real-world touchless interaction.

III. RESEARCH GAP AND MOTIVATION

3.1 Lack of Standardized and Comprehensive Air-Writing Datasets

Although several air-writing datasets exist across different sensing domains, a major research gap is the absence of standardized, cross-lingual, and large-scale datasets that can support robust benchmarking. Literature reviews highlight that datasets for English, Devanagari, and other scripts remain limited, inconsistent, or highly user-dependent [5]. Even large datasets such as WiTA [11] focus primarily on specific languages or camera- based setups, while radar- and sensor-based datasets remain small and fragmented. This lack of standardization restricts model generalization, comparative evaluation, and real-world applicability..

3.2 Sensitivity of Existing Systems to Environmental and User Variability

A persistent challenge in air-writing recognition is the high variability in user writing styles, lighting conditions, and background clutter, which significantly affect accuracy in vision-based systems. Webcam and RGB-camera approaches still struggle with fingertip occlusion, illumination variations, and cluttered scenes [13]. Likewise, EEG-based systems face noise interference and inconsistent neural responses [7], while radar- and wearable-based systems can suffer from inconsistent signal lengths and motion noise [3], [15]. These limitations hinder the deployment of robust air-writing systems in uncontrolled environments.

3.3 Limitations of Current Preprocessing and Feature Extraction Techniques

Another gap arises from the difficulty in handling variable-length trajectory signals, especially in sensor-based and radar-based systems. Many methods rely on padding, truncation, or heuristic thresholds, causing loss of temporal information and reduced accuracy. The interpolation study in [8] demonstrates that improper normalization results in degraded model performance across multiple datasets. Additionally, many deep learning frameworks require heavy computation, large memory, or complex spatio-temporal modeling [11], limiting their integration into low-power, real-time devices such as wearables or IoT systems.

3.4 Motivation for Developing Robust, Adaptive, and Lightweight Air-Writing Models

These gaps collectively motivate the development of lightweight, environment-independent, and adaptive air- writing recognition systems. There is growing need for models that operate reliably in real time using minimal hardware—such as a single webcam [13] or a compact wearable sensor [3]—while maintaining high accuracy across different users, lighting conditions, speeds, and writing styles. Moreover, practical applications such as emergency-alert systems [6], authentication technologies [12], and AR/VR interfaces demand flexible models capable of handling diverse input modalities and real-world variability. Addressing these gaps can significantly advance the usability, inclusivity, and deployment potential of air-writing systems in next-generation human– computer interaction.

IV. RESEARCH GAP AND MOTIVATION

4.1 Need for Standardized and Diverse Air-Writing Datasets

A major gap identified in recent literature is the lack of standardized, diverse, and large-scale datasets for air- writing recognition. While several datasets exist, they are often limited to specific languages, devices, or acquisition conditions, restricting the generalizability of models across users and environments. For example, the review in [5] notes that datasets for English, Devanagari, and other scripts remain inconsistent, while large-scale benchmarks such as WiTA focus only on vision-based modalities [11]. Sensor-based datasets are similarly fragmented, with wearable and radar datasets covering only small user groups [3], [15]. This fragmentation prevents fair comparison between models and limits progress toward universal air-writing solutions.

4.2 Sensitivity to Environmental and User-Dependent Variations

Despite advancements in sensing and machine learning, current systems remain highly sensitive to lighting changes, background clutter, occlusion, and variations in user writing behavior. Vision-based systems using webcams and RGB cameras still struggle with consistent fingertip tracking under real-world conditions, as noted in [13]. Wearable systems face noise and drift in inertial signals [3], while EEG-based methods suffer from low signal quality and inter-subject variability [7]. Radar-based solutions, though robust to lighting, are affected by complex motion noise and lack interpretability [15]. These dependencies highlight the need for more robust and adaptive recognition frameworks capable of performing reliably outside controlled laboratory environments.

4.3 Limitations in Preprocessing and Adaptive Modeling Approaches

Another significant gap lies in the handling of variable-length, noisy, and irregular motion trajectories, especially in sensor-driven systems. Many studies rely on padding or truncation, which can distort motion information and reduce model accuracy. The work in [8] demonstrates that improper normalization severely impacts recognition performance across multiple datasets. Additionally, many deep learning models are computationally heavy—such as spatio-temporal networks used in [11]—making them unsuitable for real-time or low-power applications. These limitations motivate the development of lightweight, adaptive, and hardware- efficient models capable of delivering high accuracy with minimal preprocessing and computation. Such improvements are crucial for real-world applications including emergency alerting [6], authentication [12], and AR/VR interaction systems.

V. CHALLENGES AND LIMITATIONS

5.1 Variability in User Writing Styles and Motion Patterns

A fundamental challenge in air-writing recognition arises from the high variability in individual writing behaviors, such as speed, stroke formation, hand orientation, and writing size. As highlighted in [5], users write characters differently in mid-air compared to on paper, causing inconsistencies that reduce classifier accuracy. Systems relying on free-space trajectories often struggle to generalize when signal shapes fluctuate significantly across individuals. This problem becomes even more prominent in large-scale datasets like WiTA [11], where writers exhibit different writing habits, hand dynamics, and stroke continuity, leading to greater intra-class variation and reduced robustness in recognition models.

5.2 Sensitivity to Environmental Factors in Vision-Based Systems

Vision-based air-writing systems suffer heavily from environmental dependencies such as illumination changes, shadows, background clutter, and fingertip occlusion. Webcam-based solutions like those in [13] are particularly vulnerable to inconsistent lighting, which affects skin detection and hand tracking. Similarly, RGB-based fingertip extraction faces difficulties in low light or highly reflective environments, causing frequent tracking interruptions.

DOI: 10.48175/IJARSCT-30732

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, January 2026 Impact Factor: 7.67

Background movement, clothing colors, and camera quality further complicate trajectory acquisition. These constraints limit the real-world deployment of camera-driven systems, especially in uncontrolled or outdoor environments.

5.3 Noise and Signal Distortion in Sensor- and Radar-Based Approaches

Sensor-based and radar-based air-writing recognition also face notable limitations due to motion noise, signal drift, and inconsistent sampling rates. Wearable wristband systems described in [3] must handle micro-vibration noise, arm-swing interference, and sensor drift, all of which distort raw motion signals. FMCW radar methods [15] struggle with multipath interference, hand-shape variations, and overlapping Doppler signatures. These distortions create noisy trajectories that degrade recognition accuracy unless extensive preprocessing or filtering techniques are applied.

5.3 Difficulty in Handling Variable-Length Time-Series Signals

A recurring limitation in time-series–driven air-writing systems is the difficulty in managing variable-length trajectory signals. Since users write at different speeds, signals vary in duration, complicating model input standardization. Many studies rely on padding or truncation, which causes loss of important temporal information. Research in [8] shows that improper normalization or resampling drastically reduces model performance across multiple datasets. Although interpolation mitigates this problem, selecting improper interpolation techniques may introduce unwanted distortion. This remains a substantial challenge for sensor-based and inertial systems that depend on sequential data.

5.4 High Computational Cost of Deep Learning Architectures

Many state-of-the-art air-writing recognition systems employ deep and computationally expensive models, limiting real-time performance and compatibility with low-power devices. Spatio-temporal convolutional models like those used in large-scale video datasets [11] require significant memory and processing power to capture both spatial and temporal dependencies. Similarly, 3D CNNs, RNN-LSTM hybrids, and multimodal fusion networks often exceed the computational capacity of embedded systems, wearables, and mobile devices. This constraint suppresses practical deployment in IoT, AR/VR, and portable consumer electronics.

5.5 Limited Cross-User Adaptability and Real-World Generalization

Despite progress in adaptive learning, cross-user and cross-environment generalization remains a major limitation. Many models trained on a specific group of users show degraded performance when tested on new users with different writing habits or motion patterns. EEG-based systems [7] demonstrate particularly strong inter- subject variability, resulting in low accuracy. Similarly, systems designed for controlled lab environments struggle when exposed to real-world settings that present unpredictable motion patterns, background interference, or device constraints. Although contrastive-learning-based adaptation in wearable systems [3] shows promise, achieving universal generalization across all sensing modalities remains an unsolved challenge.

VI. FUTURE RESEARCH AND DIRECTIONS

6.1 Development of Standardized, Multilingual, and Multimodal Datasets

A major direction for future research is the creation of large-scale, standardized, and multilingual air-writing datasets that support consistent benchmarking across different sensing technologies. Current datasets are fragmented by language, sensor type, and acquisition conditions, as highlighted in [5] and [11]. A unified dataset incorporating English, Devanagari, Chinese, and other scripts—collected under varying lighting, backgrounds, and user demographics—would significantly improve model generalization. Additionally, multimodal datasets combining vision, radar, inertial, and EEG signals would enable a comprehensive evaluation of hybrid recognition frameworks.

6.2 Robust Multimodal Fusion for Real-World Environments

Future air-writing systems must prioritize robustness and adaptability to unpredictable real-world environments. Vision-based systems often break down under illumination changes or occlusion [13], while radar and wearable systems face noise and signal irregularities [3], [15]. Combining multiple sensing modalities—such as fusing camera data with inertial or radar signals—can compensate for the limitations of single-sensor systems. Research should explore intelligent multimodal fusion models that dynamically select or integrate sensor streams to maintain accuracy in challenging conditions, particularly for outdoor, low-light, or cluttered environments.

6.3 Lightweight and Hardware-Efficient Deep Learning Models

As demand increases for mobile, wearable, and IoT-based air-writing interfaces, future work should focus on lightweight neural networks that reduce computational burden without sacrificing accuracy. Current spatio- temporal models and 3D CNNs remain too resource-intensive for real-time processing on edge devices [11]. Techniques such as model pruning, quantization, knowledge distillation, and efficient CNN design could make air-writing systems deployable on consumer devices like smartwatches, AR glasses, and smartphones. Models such as interpolation-aware CNNs [8] and contrastive-learning approaches [3] provide promising foundations for efficient, adaptive architectures.

6.4 Personalized, Adaptive, and Continuous-Learning Systems

Given significant variability in individual writing styles, future research must explore personalized and adaptive air-writing recognition. Few-shot learning techniques, self-supervised adaptation strategies, and online learning frameworks can help systems adjust to new users with minimal calibration effort. The rapid adaptation shown in wearable systems [3] and the potential of contrastive and meta-learning approaches highlight the importance of personalization. Additionally, continuous-learning mechanisms could allow systems to refine recognition over time based on user feedback. This direction is crucial for developing inclusive, user-friendly systems for authentication, rehabilitation, assistive technologies, and AR/VR environments.

VII. CONCLUSION

Air-writing recognition has progressed from a niche research topic into a significant enabler of next-generation touchless human–computer interaction, offering an intuitive and hygienic means of communication in environments where physical contact is impractical or undesirable. The collective body of work demonstrates substantial advancements across multiple sensing paradigms, including vision-based systems that rely on RGB cameras and webcams for fingertip tracking and trajectory acquisition [13], large-scale unconstrained datasets such as WiTA that support high-speed spatio-temporal deep learning architectures for multilingual text recognition [11], and depth-sensor and wearable technologies that provide enhanced motion capture precision in diverse real-world settings [10], [3]. Radar-based approaches have further expanded the robustness of air-writing recognition, enabling accurate gesture interpretation under variations in lighting, occlusion, and environmental noise [15], while EEG-based methods have opened pathways for hands-free, neural-driven writing interfaces that support accessibility applications despite their current limitations in accuracy and consistency [7]. However, across all modalities, persistent challenges remain a barrier to universal deployment, including pronounced user-dependent writing variations, instability caused by illumination changes and background clutter, complex noise patterns in inertial and radar signals, variable-length time-series trajectories, and the heavy computational requirements of many deep learning architectures. Additionally, the absence of standardized, multilingual, and multimodal datasets continues to hinder fair benchmarking and limits the generalization capability of existing models, as emphasized in recent reviews [5]. While some studies have proposed solutions—such as robust interpolation strategies for signal normalization [8] and adaptive contrastive learning for rapid personalization [3]—there is still a pressing need for integrated frameworks that combine efficiency, robustness, and cross-user adaptability. Looking ahead, future research must prioritize the creation of unified datasets, the development of lightweight yet powerful neural architectures suitable for deployment on mobile and wearable devices, and the design of multimodal fusion systems that leverage the complementary strengths of visual, inertial, radar, and neural signals to overcome the limitations of single-sensor approaches. Furthermore, continuous-learning and personalization techniques are essential for managing the natural variability in human writing behavior, ensuring that air-writing interfaces remain inclusive and accessible to all users. With focused efforts in these areas, air-writing recognition is poised to become a practical and widely adopted interaction paradigm for AR/VR systems, authentication and security mechanisms, assistive technologies, smart environments, and emergency communication scenarios, ultimately contributing to more seamless, adaptive, and intelligent touchless computing experiences.

REFERENCES

[1] Y. L. Tonape et al., "Result paper on air writing recognition using machine learning algorithms," Int. J. Adv. Comput. Theory Eng., vol. 14, no. 1, 2025.

[2] Y. L. Tonape et al., "Machine learning-based air writing: A literature review on techniques, progress, and challenges," ITSI Trans. Electr. Electron. Eng., vol. 13, no. 2, 2024.

[3] Y. Guo et al., "A rapid adaptation approach for dynamic air-writing recognition using wearable wristbands with self-supervised contrastive learning," Nano-Micro Lett., vol. 17, p. 41, 2025

[4] M. Wu et al., "Finger in camera speaks everything: Unconstrained air-writing for real-world," IEEE Trans. Circuits Syst. Video Technol., 2024.

[5] M. K. Jabde, C. H. Patil, A. D. Vibhute, and S. Mali, "A comprehensive literature review on air-written online handwritten recognition," Int. J. Comput. Digit. Syst., vol. 15, no. 1, pp. 307–322, Jan. 2024.

[6] K. Navya et al., "A Machine Learning Approach for Air Writing Recognition," Int. J. Res. Appl. Sci. Eng. Technol., vol. 11, no. 4, pp. 3528–3533, 2023.

[7] A. Tripathi et al., "NeuroAiR: Deep Learning Framework for Airwriting Recognition from Scalp-Recorded Neural Signals," arXiv preprint arXiv:2308.03555, 2023.

[8] F. A. Abir et al., "Deep Learning-Based Air-Writing Recognition with the Choice of Proper Interpolation Technique," Sensors, vol. 21, no. 8407, 2021.

[10] M. S. Alam et al., "Trajectory-Based Air-Writing Recognition Using Deep Neural Networks and Depth Sensor," Sensors, vol. 20, no. 376, 2020

[11] U.-H. Kim, Y.-W. Hwang, S.-K. Lee, and J.-H. Kim, "Writing in the Air: Unconstrained Text Recognition from Finger Movement Using Spatio-Temporal Convolution," arXiv preprint arXiv:2104.09021, 2021.

[12] D. Lu, D. Huang, and A. Rai, "FMHash: Deep Hashing of In-Air-Handwriting for User Identification," arXiv preprint arXiv:1806.03574, 2019.

[13] C.-H. Hsieh, Y.-S. Lo, J.-Y. Chen, and S.-K. Tang, "Air-Writing Recognition Based on Deep Convolutional Neural Networks," IEEE Access, vol. 9, pp. 142827–142838, 2021.

[14] [1] C.-H. Hsieh, Y.-S. Lo, J.-Y. Chen, and S.-K. Tang, "Air-Writing Recognition Based on Deep Convolutional Neural Networks," IEEE Access, vol. 9, pp. 142827–142836, Oct. 2021.

[15]W. Kim et al., "FMCW Radar Based In-Air Alphanumeric Gesture Recognition with Machine Learning," IEEE Trans. Instrum. Meas., 2025.