Air Scribe: Seamless Air Gesture Text Input and Virtual Mouse Functionality Web Interfaces

Author Prof. Ashwini Pawar

Darshika Bhagvandas Salve CSE (AIML) -Lokmanya Tilak College of Engineering
Rutuja Pradeep Pisal CSE (AIML) -Lokmanya Tilak College of Engineering
Sakshi Dattatray Shende CSE (AIML) -Lokmanya Tilak College of Engineering
Ruoesh Mukesh Sharma CSE (AIML) -Lokmanya Tilak College of Engineering

---***---

Abstract - Air Scribe is a web-based application that uses advanced computer vision and machine learning techniques to convert hand-written gestures performed in the air into digital text. This application detects and interprets user movements, translating them into text on a screen. In addition to text transcription, Air Scribe incorporates a virtual mouse feature, enabling users to interact with digital interfaces using air gestures. This functionality aims to enhance user experience by providing an intuitive and innovative method for text input and digital navigation. The application offers a new approach to user interaction with digital systems, allowing for seamless integration of gesture-based controls. Air Scribe has the potential to revolutionize how individuals engage with technology, particularly in environments where traditional input methods may not be feasible or practical. This paper presents the underlying technology of Air Scribe, its development process, and the significance of its application in various fields.

Key Words: computer vision, machine learning, air gestures, digital text, virtual mouse, user interaction.

1.INTRODUCTION

Air Scribe is a cutting-edge web application designed to bring a new level of interaction with digital systems. Imagine being able to write in the air and see your words appear on the screen—this is exactly what Air Scribe allows you to do. Using sophisticated computer vision and machine learning algorithms, it captures hand-written gestures and translates them into digital text in real-time. But it doesn't stop there. Air Scribe also offers a unique virtual mouse functionality, allowing users to navigate and interact with digital interfaces using their gestures. This innovation makes it possible to interact with technology in a way that feels natural, intuitive, and hands-free. Whether you're looking to input text or control digital systems without touching a keyboard or mouse, Air Scribe opens up exciting possibilities for how we engage with technology. By blending cutting-edge technology with human-like interaction, Air Scribe is set

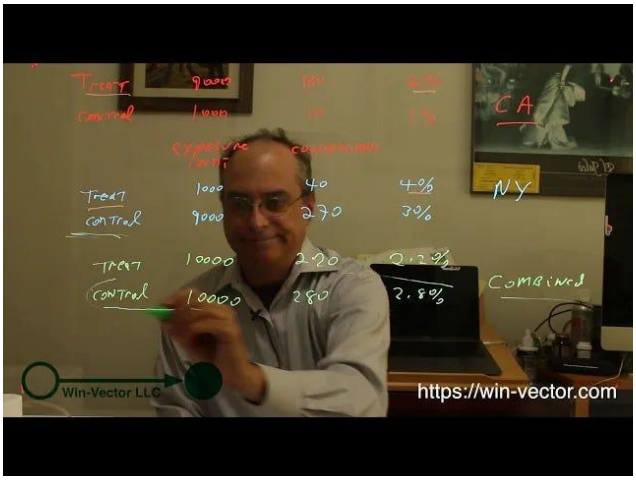

Fig. 1 Air Writing

to change the way we communicate with and control digital environments.

2. Body of Paper

Fig. 1 Air Writing

Fig.2 Gesture Recognitio

3. Literature survey

Sr. No	Details
1	**Authors**: G. R. S. Murthy & R. S. Jadon (2018) **Title**: A Survey of Gesture Recognition for Interactive Systems **Contributions**: This paper surveys gesture recognition techniques used to control devices or convey information, focusing on the two main approaches in human-computer interaction (HCI). **Gaps**: The study highlights the lack of standardized benchmarks and the limitations of cross-environment gesture recognition.
2	**Authors**: Jessie Y.C. Chen, Ellen C. Hass, Michael J. Barnes (2019) **Title**: Enhancing User Interfaces with Air Gesture

Sr. No	Details
	Control: Challenges and Solutions **Contributions**: Reviews over 150 studies on human performance issues in teleoperation, discussing solutions like stereoscopic displays and multimodal interfaces. **Gaps**: Identifies the need for more intuitive gesture controls and seamless integration with existing systems.
3	**Authors**: Prof. Ashvini Bamanikar, Vijay Kamble, Tanvi Ghule, Anuja Kotkar, Vinayak Kulkarni (2019) **Title**: Air Writing and Gesture Recognition Using Machine Learning Techniques **Contributions**: Proposes a novel method of computer interaction by using finger gestures in the air, employing MediaPipe for fingertip detection and an RNN model for pattern recognition to convert gestures into text. **Gaps**: The paper mentions challenges such as scalability of the models, performance across diverse users, and the need for better integration of depth sensing technology.
4	**Authors**: Moniruzzaman Bhuiyan, Rich Picking (2020) **Title**: Gesture Recognition for Human-Computer Interaction: A Review **Contributions**: Introduces Open Gesture, a prototype for controlling devices like phones and TVs with simple hand gestures. Evaluations suggest potential benefits for improving independence and quality of life, especially for older and disabled users. **Gaps**: Emphasizes the need for improved accuracy, better real-time solutions, and performance consistency across different environments.
5	**Authors**: Jaya Prakash Sahoo, Allam Jaya Prakash, Saunak Samantray (2021) **Title**: Real-Time Hand Gesture Recognition with Convolutional Neural Networks **Contributions**: Utilizes fine-tuning of pre-trained CNNs with score-level fusion for hand gesture recognition, validated through cross-validation on benchmark datasets, demonstrating its real-time application in an ASL system. **Gaps**: Highlights limitations of CNNs in terms of computational resources and the need for better adaptation to varying lighting conditions.

4. Summarized Findings or Research gaps

❖ Lack of Standardized Benchmarks and Cross-Environment Consistency:

Gesture recognition systems lack universally accepted benchmarks, making it difficult to compare performance. Additionally, these systems struggle to maintain consistent accuracy across different environments and conditions.

❖ Need for Intuitive Gesture Controls:

Existing gesture control methods are not always user-friendly or intuitive, which limits their widespread adoption. There is a need for more natural and easily understandable gestures to improve user experience.

❖ Challenges in System Integration:

Integrating gesture-based controls with current systems and platforms is a significant challenge. Ensuring seamless functionality alongside traditional input methods (like keyboards and touchscreens) remains an area that requires further research.

❖ Scalability and User Diversity:

Gesture recognition models often face difficulties when scaling across different user demographics, including age, hand size, and motion variations. Ensuring reliable performance across a diverse user base is crucial.

❖ Accuracy and Real-Time Performance:

Achieving high accuracy in gesture recognition while maintaining real-time responsiveness is a persistent challenge. Improvements in both areas are necessary to make gesture recognition systems practical for everyday use.

Fig.2 Gesture Recognition

5. Problem Statement and Objectives

Air Scribe replaces traditional input methods with air gestures for text and control. The
challenge is to accurately capture hand movements in 3D space, ensuring reliable output despite
variations in gesture speed and style. This requires integrating computer vision and machine
learning to recognize gestures in real-time.

Objectives :

1. Accurately capture hand movements in 3D space for input.
2. Develop real-time recognition of air gestures using computer vision.
3. Ensure high accuracy despite variations in gesture speed and style.
4. Integrate machine learning models for precise gesture interpretation.

6.Proposed System

4.1 Analysis/Framework/Algorithm

1. Start
2. Input and Sensing Module:
- Capture and send input data.
3. Machine Learning and Interpretation Module:
- Process data using machine learning.
- If successful, proceed; otherwise, loop to Interface Module for adjustments.
4. Interface Module:
- Handle user/device interaction and return data for reprocessing if needed.
5. Gesture Sensing Module:
- Capture and send gesture data.
6. Backend and Data Processing:
- Process all data and check completeness.
- If complete, proceed; otherwise, loop back to Interface Module.
7. Decision Point:
- If ready, proceed to visualization; otherwise, return to Interface Module.
8. Visualization and Final Interpretation:
- Visualize data and generate final output.
9. End

7. Methodology

Methodology for Console-Based Air Scribe (Shortened)

1. Data Collection & Preprocessing:

❖ Capture real-time hand gestures via webcam using OpenCV.

❖ Preprocess the input (background subtraction, thresholding) to isolate hand
movements.

2. Model Training:

❖ Train a deep learning model (CNN or LSTM) to map gestures to characters.

❖ Use labeled datasets of air gestures for training.

3. Real-Time Gesture Recognition:

❖ Process live video feed, detect gestures, and send data to the model for interpretation.

4. Console Output:

❖ Display recognized gestures as text in real-time on the console.

❖ Handle errors with feedback like "Invalid gesture, try again."

5. System Feedback:

❖ Refine transcription accuracy with feedback loops and adjust gestures as needed.

Tools: Python, OpenCV, TensorFlow/PyTorch.
Hardware: Webcam.

8.Experimental Setup

5.1 Details of Database/Dataset

❖ Data Collection Methodology

Air gestures recorded using a camera in varied lighting/backgrounds for robustness.
Multiple users perform predefined gestures (letters, numbers, symbols), captured as
frames or video streams.
MediaPipe can be used for real-time hand and finger movement detection.

❖ Types of Data:

Video Data: Continuous sequences of hand movements.
Skeleton Key Points: Hand landmarks extracted using MediaPipe.
Time-Series Data: Sequential movements of fingers for detecting writing gestures.
Text Labels: Each gesture labeled with corresponding letters, numbers, or symbols.

❖ Dataset Structure:

Training Data: 70-80% for diverse gesture training.
Validation Data: 10-15% for tuning and preventing overfitting.
Test Data: 10-15% for evaluating model performance.

❖ Size and Variability:

Number of Users: 20-30 individuals of varied demographics.
Number of Gestures: At least 26 alphabets, 10 digits, and basic symbols.
Environmental Variability: Capture in varied lighting and angles for adaptability.

❖ Preprocessing:

Noise Removal: Filtering irrelevant data.
Normalization: Standardizing key points for consistent performance.
Augmentation: Data variations (rotation, scaling) for robustness.

❖ Sources:

Use public datasets like NTU RGB+D or LeapGestRecog, or create a custom dataset tailored to Air Scribe.

Link : https://www.kaggle.com/datasets/gti-upm/leapgestrecog

9. Software and Hardware setup

➢ Processor: Intel Core i3/i5/i7 or AMD Ryzen.
➢ RAM: 4GB to 32GB DDR4.
➢ Storage: 128GB to 1TB SSD, or 500GB/1TB HDD.
➢ Graphics: Intel Integrated (UHD/Iris Xe) or dedicated NVIDIA/AMD GPUs.
➢ Display: 13-15.6 inches, HD or Full HD, with some touchscreen options.
➢ Ports: USB-A, USB-C, HDMI, 3.5mm jack, and SD card reader.
➢ Battery: 3-cell or 4-cell Li-ion, 4-8 hours.
➢ OS: Windows 10 or 11.

10. CONCLUSIONS

The Air Scribe project enhances human-computer interaction by enabling seamless transcription of hand-drawn gestures into digital text. Integrating letter recognition improves intuitive text input using computer vision and machine learning for accurate real-time interpretation. This innovative approach simplifies digital interactions, empowering users and setting the foundation for future enhancements in engagement with technology.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our Head of Department (HOD),Dr Chaitrali Chaudhari , for their continuous support and encouragement throughout the course of this project. We also wish to thank our families for their patience, understanding, and unwavering support.

References

(Books, journals and other online references)
• General research & study on Google & YouTube.
https://www.youtube.com/watch?v=ufm6tfgo-OA&t=6s
•
Real-Time Hand Gesture Recognition with Convolutional Neural Networks:
https://r.search.yahoo.com/_ylt=AwrPo.zX.71mcwQATx.7HAx.;_ylu=Y29sbwNzZzMEcG9zA
zEEdnRpZA-
MEc2VjA3Ny/RV=2/RE=1724936408/RO=10/RU=https%3a%2f%2fieeexplore.ie
ee.org%2fdocument%2f8843453/RK=2/RS=T9Ap66ff1IQgOp.ScNXZ
LilU_oU-
• Gesture Recognition for Human Computer Interaction: A Review
https://r.search.yahoo.com/_ylt=AwrKGQmw_L1m4wQAP_K7HAx.;_ylu=Y29sbwNzZzMEc
G9zAzEEdnRpZA-
MEc2VjA3Ny/RV=2/RE=1724936625/RO=10/RU=https%3a%2f%2fwww.
sciencedirect.com%2fscience%2farticle%2fpii%2fS1877050917319130/RK=2/RS=6t7vTAwT
uHVumFLHImKZLwxkH7o-
• Radar-Based Air-Writing Gesture Recognition Using a Novel Multistream CNN
Approach:
https://r.search.yahoo.com/_ylt=AwrKFi05_b1mwgQAXZy7HAx.;_ylu=Y29sbwNzZzMEcG9z
AzIEdnRpZA-
MEc2VjA3Ny/RV=2/RE=1724936762/RO=10/RU=https%3a%2f%2fieeexplore.i
eee.org%2fdocument%2f9820766/RK=2/RS=Lue5xFLsvfMANzd-
KiVBCrTc3g7w-

BIOGRAPHIES (Optional not mandatory)

Author Prof Ashwini Sangolkar

Co-Author Darshika Bhagvandas Salve

Co-Author Rutuja Pradeep Pisal

Co-Author Sakshi Dattatray Shende

Co-Author Rupesh Mukesh Sharma