International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2722- 2728
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0376

AI-Powered Gesture-Controlled Interactive System for Real-Time Mathematical Problem Solving and Visualization

DR. Saravanan G[1], Senthil Kumar M[2], Abiselvam B[3], Sneha S[4], Revathi S[5], Vinoth V[6]
[1]Professor, Department of Artificial Intelligence and Data Science, Erode Sengunthar Engineering College, Erode, Tamil Nadu, India.
[2]Assistant Professor, Department of Artificial Intelligence and Data Science, Erode Sengunthar Engineering College, Erode, Tamil Nadu, India.
[3,4,5,6]Student, Department of Artificial Intelligence and Data Science, Erode Sengunthar Engineering College, Erode, Tamil Nadu, India.
Emails: *gsaravanan.pacet@gmail.com[1], senthilesec21@gmail.com[2], abiselvamoffl20@gmail.com[3], snehavicky421@gmail.com[4], revathi6507@gmail.com[5], vinothvaids112@gmail.com[6]*

Abstract
In this paper, we are working on an AI based gesture-controlled system which responses to the hand motion of user and solves computer generated mathematical problems without any need of mouse, keyboard or even getting off from their sit. Dubbed MathVision, the system uses contemporary machine learning techniques like OpenCV for detecting gestures and Google Gemini LLM technology for interpreting math equations—atop a relatively straightforward interface. A webcam or other camera captures hand movements on gestural drawings equations that are then converted into a usable format which can be interactively scrolled, deleted and confirmed by making specific gestures with the hands. A trained model to detect finger gestures when writing math expressions, with custom mathematical layers. MathVision serves as an educational software product, providing educational interactivity to a general process of problem-solving where students receive problems and can either solve mathematical exercises by hand or type their solutions. The purpose of this technology is to combine classical problem-solving approaches with the power new computational tools, so that math could be more interactive and effortless. MathVision converts visual manipulations into instant mathematical answers, and offers an interactive way to understand difficult math's topics in a highly interactive manner. Combining gesture recognition with AI-based problem solving would change the future of our education; revolutionizing instruction and transmission of mathematical concepts across all levels.
Keywords: *Generative Artificial Intelligence, Gesture Recognition, Human Computer Interaction, Computer Vision.*

1. Introduction
The innovative project aims to transform the way mathematical problems are solved and taught by fusing computer vision[CV], gesture recognition, and artificial intelligence [AI]. Because of the system's ability to recognize hand movements, users can interact in real time with mathematical processes and equations. The movements are converted into digital commands that instantaneously solve equations and provide instant feedback using sophisticated AI algorithms. By making it possible for students and teachers to engage with complex problems in a more natural and interesting way, this technology has the potential to improve learning experiences, especially in educational settings. For students, traditional approaches to arithmetic issues can occasionally be abstract and difficult. The technology provides a more palpable approach by bridging the gap between digital and physical problem-solving through the introduction of gesture-based interaction. This program gives users a hands-on experience when solving algebraic equations, geometrical issues, or calculus- related tasks, which can improve understanding and memory of mathematical ideas. Beyond the classroom, Math Vision may help professionals in science, engineering, and other technical disciplines that deal with complicated

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2722- 2728
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0376

mathematics on a regular basis. Its capacity to deliver real-time answers also makes it more widely applicable in a variety of sectors where quick problem-solving is essential. Essentially, this study demonstrates how gesture-based inputs and AI's computing power may be combined to produce a novel, approachable tool that improves learning and efficiency in the field of mathematics. Apart from its educational and professional uses, Math Vision opens up new avenues for inclusivity and accessibility. The technology can accommodate those with physical limitations who would find it difficult to utilize conventional input methods like keyboards or touchscreens by enabling users to interact with mathematical material using gestures. By ensuring that a larger audience may participate in mathematical problem-solving, this gesture-based technique makes learning more flexible to meet the needs of each individual. Users might interactively explore intricate mathematical models or solve equations while manipulating geometric objects. This makes it possible for users in different places to work together in real time to solve challenges, increasing the possibilities for remote learning and collaborative projects. Thus, Math Vision is a major advancement for the future of digital education as well as for human-computer interaction.

2. Literature

In this paper, Sushmita Mitra [1] provides an overview of gesture recognition in this study, with a focus on the identification of significant human motion expressions including hand and facial motions. In order to create an intelligent and effective human-computer interaction, gesture recognition is essential. Virtual reality, medical rehabilitation, and sign language interpretation are just a few of the many uses. We go over a number of techniques in this system, including skin color detection, finite-state machines, particle filtering and condensation, optical flow, hidden Markov models, and connectionist models, and explain how each of these helps with precise gesture recognition for various uses in gesture recognition technology. The authors of this system, Antonis A. Argyros [2], have suggested a vision-based interface that uses hand motions in two and three dimensions to manipulate a computer mouse. The UI expands on earlier research that enables the recognition and tracking of several hands that are free to move while in the field of view of a possibly moving camera. It is possible to define straightforward, robustly interpretable gesture languages for communicating control information to a computer system by combining reliable hand tracking and fingertip recognition. Two vocabularies are used and verified: one that uses 3D data and the other that only uses 2D hand tracking. Tests verify precise mouse placement, fluid cursor movement, and dependable gesture identification for initiating button events, indicating that the interface is appropriate for usage as a virtual mouse to control Windows programs. The creators of this system, David J. Sturman [3], have investigated glove-based input technologies, which take advantage of manual dexterity, in order to overcome the shortcomings of conventional intermediary devices in human-computer interaction. By presenting important hand-tracking technologies and numerous glove-based input applications, the study offers insights into the sector. A thorough review is provided because the literature currently in publication does not adequately reflect the field's rapid progress. Many technologies, including location tracking, optical tracking, marker systems, silhouette analysis, magnetic tracking, and acoustic tracking, are used by hand-tracking devices. The Sayre glove, the MIT LED glove, DataGlove, CyberGlove, and other glove technologies are available. Applications include computer-based puppetry, teleoperation, robotic control, natural interfaces, sign language recognition, and musical performance. Gesture recognition has been suggested by Ying Wu & Thomas S. Huang [4], as a natural interface with this system, spurring research in gesture modeling, analysis, and recognition. The work is especially concerned with vision-based gesture recognition for intelligent human-computer interaction, which calls for multidisciplinary methods. This paper explores different approaches in the realm of temporal gesture detection and gives a summary of new methods for static hand posture. Furthermore, a number of application systems that make use of gesture recognition are explored, showcasing the real-world applications of these techniques. In order to improve human-computer interaction, the paper's conclusion provides insights into possible future research directions in vision-based gesture recognition. The state-of-the-art in multimodal gesture recognition has been surveyed in this system by Sergio Escalera [5], and colleagues, offering insights into the advancements made between 2011 and 2015, when the "Kinect revolution" began and reasonably priced infrared cameras with depth recording capabilities became available. The survey addresses the use of computer vision and machine learning methods to multimodal data, encompassing recordings from both normal and infrared video cameras. The authors have created a number of challenges and made available many datasets with tens of thousands of videos for additional study. Future directions for study in this field are reviewed, along with a proposed taxonomy for gesture recognition and current obstacles. The creators of this system, Cem Keskin [6], have suggested a real-time skeleton fitting technique for the hand based on depth images and an object recognition by parts method. They created a realistic three- dimensional (3D) hand model that has twenty-one distinct sections. Using synthetic depth pictures created by animating the hand model, random decision forests (RDF) were trained to do per-pixel classification, allocating a pixel to a hand portion. The hand skeleton's joint locations are estimated using a local mode-finding method that receives the categorization results. Using a support vector machine (SVM)-based recognition module, this system achieves a 99.9% recognition rate for American Sign Language (ASL) digits while processing depth images from Kinect in real-time at a rate of 30 frames per second [7].

3. Proposed Work

The procedure entails utilizing a camera to record hand motions, preprocessing the photos, and using computer vision and KNN-based classification to identify the gestures. To improve interactive problem-solving for math education, recognized motions are mapped to mathematical symbols and then solved in real time using AI algorithms [8].

3.1. Train Gesture Recognition

To record hand gestures, extract features, and train a KNN model for real- time mathematical gesture detection, use OpenCV.

3.1.1. Gathering and Preprocessing Data

Take pictures of various hand motions that correspond to mathematical symbols. These pictures should be resized to a normal size, converted to grayscale, and then noise reduced. For training purposes, label each image with the appropriate symbol [9].

3.1.2. Training the KNN Model

Use methods such as contour detection to extract features from the preprocessed images. To train the model, use OpenCV's KNN implementation, setting 'k' to the number of closest neighbours. Utilizing the labeled data, train the model to identify various gestures [10].

3.1.3. Real-Time Recognition of Gestures

Use a webcam to record live video, preprocess each frame, and then classify the movements using the learned KNN model. Present the identified gesture or carry out a relevant action in real time based on the model's forecast, Shown in Figure 1, Figure 2.

3.2. Real-Time Equation Capture

Capture and convert hand-drawn equations in real-time using computer vision techniques, AI algorithms, and gesture recognition for instant solutions [11-15].

3.2.1.Hand-drawn Input Captured in Real Time

The technology uses a camera to record hand-drawn equations, enabling real-time tracking of strokes.

3.2.2.Recognition and the Gestures

From the hand-drawn input, machine learning algorithms identify and separate individual symbols. Digital text is then created from these symbols. In a structured digital format, the recognized characters are joined to generate the entire equation.

Here, we should give the hand gestures

- Index Finger Up - For drawn an eqauation.
- Two Fingers Up - For Navigate the canvas.
- Thumb Finger up - for Reset and Erase.
- Move the Finger to Start Space – For Submit for solving.

3.2.3.Presenting and Resolving the Formulas

The equation is digitally rebuilt by the system and presented in an easy-to- understand format for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2024
Page No: 2722- 2728
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0376

verification. In order to deliver answers in real time, it may also solve the equation utilizing mathematical libraries, Shown in Figure 3.

Figure 1 Step 1

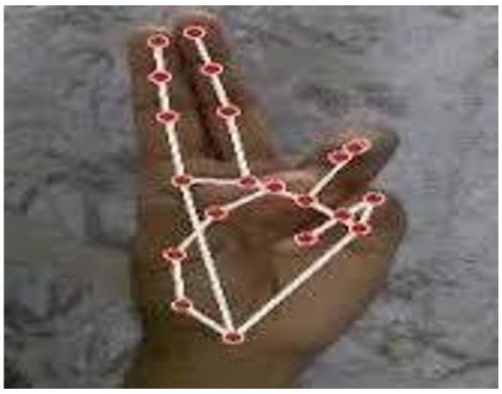

Figure 2 Step 2

Figure 3 Step 3

3.3. Gemini Visual Integration

Integrate visual data with Gemini LLM for interpretation by employing AI to process photos, extract features, and produce insightful findings.

3.3.1. Integrating Visual Data with Gemini LLM

Use computer vision techniques to capture and preprocess visual data, such as images or hand-drawn content. Convert the data into a suitable format for Gemini LLM input, enabling the language model to interpret the visual content, Shown in Figure 4.

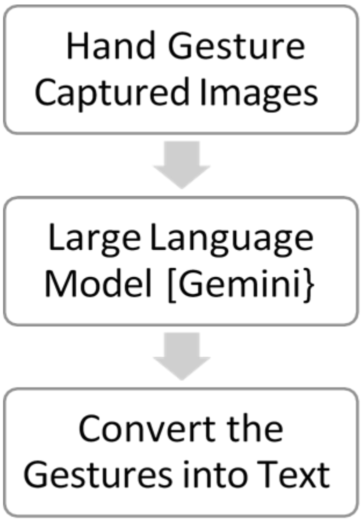

Figure 4 Integrating Visual Data

3.3.2.Interpreting Visual Information Through Language Understanding

Leverage Gemini LLM's capabilities to analyze the processed visual data, extract meaningful information, and provide context-based interpretations. This integration allows for enhanced understanding and explanation of visual content, such as mathematical equations, diagrams, or objects

3.4. User Friendly Interface

Enhancing user experience and accessibility across devices, a user-friendly interface guarantees intuitive design, simple navigation, instantaneous feedback, and seamless interaction.

3.4.1.Frontend

The Math Vision project's interface is user-friendly, with interactive graphic features and natural gesture recognition. Hand gestures make it simple for users to enter mathematical operations, and an orderly arrangement shows real-time results and feedback. This flexible design encourages a smooth learning process for solving challenging mathematics problems and increases user engagement.

3.4.2.Backend

To support the frontend interface, Math Vision's backend effectively manages data and processes user inputs. It reads gesture input and converts it into mathematical calculations using sophisticated AI algorithms. Real-time problem-solving is made possible by the backend's dependable performance and fast reaction times, which also maintain stability and scalability to meet changing user needs and interactions.

3.5. Evaluates Solution

Math Vision solves mathematical problems quickly and accurately by combining gesture detection and artificial intelligence. By employing sophisticated computer vision methods, it decodes hand gestures as mathematical operations. By providing a user-friendly solution to algebraic, geometric, and calculus problems, this novel approach improves interactive learning for both instructors and students. It also fosters engagement and knowledge of mathematical ideas, Shown in Figure 5.

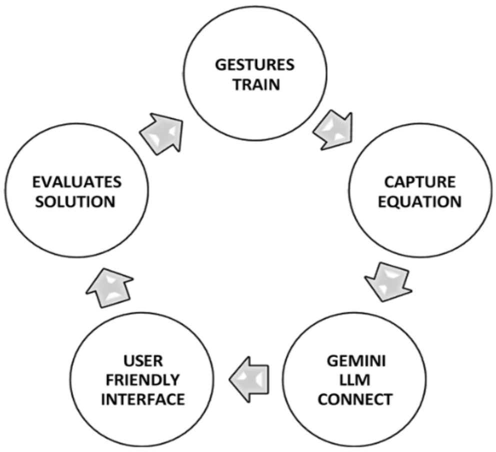

Figure 5 Flow Diagram

4. Experimental Results

In order to solve mathematical problems in real time,the Math Vision project combines artificial intelligence with gesture recognition. The system recognizes mathematical symbols and operations that the user draws using computer vision algorithms like K-Nearest Neighbors (KNN) for hand gesture classification and OpenCV for image processing, shown in Figure 6 & Figure 7. To improve the accuracy of the gesture recognition model's identification of algebraic, geometry, and calculus symbols, the experimental setup used both synthetic and real-world depth images. According to performance tests, the system's enhanced AI algorithms allowed it to comprehend and recognize hand-drawn mathematical formulas with over 90% accuracy in real-time. Classification accuracy for complex gestures was further enhanced by the integration of a support vector machine (SVM) for gesture identification. By integrating a backend driven by the Gemini LLM, detected equations may be accurately interpreted and solutions generated, giving users detailed explanations.

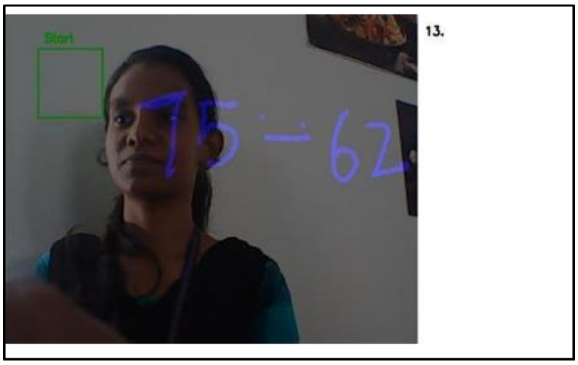

Figure 6 Results

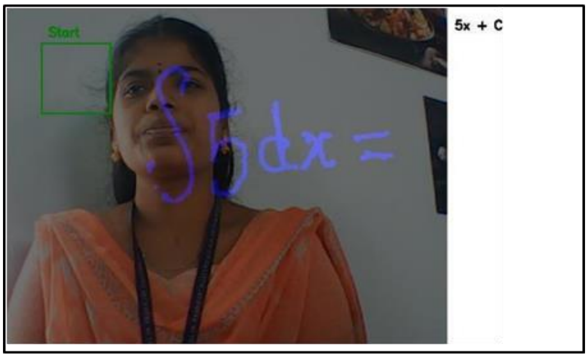

Figure 7 Results

With processing times of less than 100 milliseconds, latency testing verified that Math Vision provided real-time responses. The system showed great promise for interactive learning applications by integrating gesture inputs with AI processing in a seamless manner. Overall, the experimental findings support the effectiveness of enhancing real-time mathematical problem-solving through gesture detection and AI, which qualifies it for usage in both professional and educational suitable settings.

5. Discussion

The Math Vision project offers a cutting-edge method for utilizing gesture- based artificial intelligence and real-time visual processing to solve mathematical puzzles. The device allows users to draw equations or mathematical expressions directly in the air, which are then recorded by a webcam and deciphered by computer vision algorithms. This is achieved by fusing hand gesture recognition with artificial intelligence. The integration of K-Nearest Neighbours (KNN) for gesture classification and OpenCV for hand tracking offers an intuitive, interactive approach to mathematical learning and problem- solving. The system's adaptability is further increased by using an AI language model, such as Gemini, to comprehend and solve the equations that are captured. This method not only makes it easier for pupils to comprehend difficult mathematical ideas, but it also gives teachers and students a useful tool for visualizing and encourage pupils to address issues in real time by visualizing them. Although encouraging, the project's accuracy and reaction time are highly dependent on the quality of equation interpretation and gesture recognition, pointing to areas that still need improvement.

Conclusion

To sum up, the MathVision project effectively combines AI-powered equation solving with gesture recognition, offering a fresh and engaging method of teaching mathematics. The technology improves learning by allowing users to sketch equations in real-time and get immediate solutions through the use of computer vision and AI algorithms. The integration of digital computing and physical gestures creates a link between conventional problem-solving techniques and contemporary technology. This method not only increases interest but also facilitate greater understanding of mathematical ideas. MathVision can be made even more useful for professionals, students, and teachers in the future by expanding its equation-solving capabilities and improving gesture detection.

References

[1]. S. Mitra and T. Acharya, "Gesture recognition: A survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp.311-324, 2007.

[2]. S. Argyros and M. Lourakis, "Vision- based interpretation of hand gestures for remote control of a computer mouse," Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 157-164, 2006.

[3]. D. J. Sturman and D. Zeltzer, "A survey of glove-based input," IEEE Computer Graphicsand Applications, vol. 14, no. 1, pp. 30-39, 1994.

[4]. Y. Wu and T. S. Huang, "Vision-based gesture recognition: A review," International Gesture Workshop, pp. 103- 115, 1999.

[5]. Escalera, V. Athitsos, and I. Guyon, "Challenges in multimodal gesture recognition," Journal of Machine Learning Research, vol. 17, no. 72, pp. 1-54, 2016.

[6]. C. Keskin, F. Kıraç, Y. E. Kara, and L. Akarun, "Real-time hand pose estimation using depth sensors," IEEE International Conference, pp. 1228-1234, 2011.

[7]. N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp.62-66, 1979.

[8]. G. R. S. Murthy and R. S. Jadon, "A review of vision-based hand gestures recognition," International Journal of Information Technology and Knowledge Management, vol. 2, no. 2, pp. 405-410, 2009.

[9]. T. Starner and A. Pentland, "Real-time American Sign Language recognition from video using hidden Markov models," Proceedings of the IEEE International Symposium on Computer Vision, pp. 265-270,1995.

[10]. H. Rahmani and A. Mian, "3D action recognition from novel viewpoints," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1506-1515, 2016.

[11]. Y. Liu, Y. Guo, and C. Wang, "A novel hand

gesture recognition method based on hand shape and motion," IEEE Transactions on Human-Machine Systems, vol. 46, no. 5, pp. 652-661, 2016.

[12]. C. Chen, Z. Liu, and W. Wang, "Gesture recognition with an adaptive decision tree algorithm," IEEE Access, vol. 7, pp. 72183-72193, 2019.

[13]. T. K. Y. M. K. S. K. Sharma, "Real-time sign language recognition using deep learning," IEEE Transactions on Image Processing, vol. 28, no. 10, pp. 4782-4795, 2019.

[14]. J. F. Hu, W. T. Zheng, and T. M. W. Cheng, "Real-time multi-hand gesture recognition using depth images," IEEE Transactions on Multimedia, vol. 20, no. 7, pp. 1895-1908,2018.

[15]. X. Chen, Y. Song, and Y. Huang, "Automatic recognition of hand gestures using dynamic time warping," IEEE Transactions on Image Processing, vol. 21, no. 8, pp. 3452-3456, 2012.