International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Jan-Feb-2026, ISSN (Online): 2395-566X

FlowBeats: Gesture-Based Control Technique for Intelligent Music Interaction System

Siddhi Pawar, Anuradha Raut, Tanuja Suryawanshi, Shravani Wadghare, Guide: Prof. Pradnya Satpute
Department of Computer Engineering, Zeal Polytechnic

Abstract- — Gesture recognition has become an important research area in Human–Computer Interaction (HCI). It enables users to control digital systems using natural hand movements instead of traditional input devices such as keyboards or touch screens. This paper presents FlowBeat, a gesture-controlled music interaction system that allows users to control music playback using simple hand gestures captured through a webcam. The system uses computer vision techniques with OpenCV and MediaPipe to detect hand landmarks and classify gestures in real time. The recognized gestures are mapped to commands such as play, pause, next track, and previous track. The proposed system provides a low-cost, touchless, and intuitive interface for music control. The paper discusses existing gesture recognition techniques, system architecture, algorithm design, and advantages of the proposed solution. The motivation behind developing the FlowBeat system is to create a more natural and convenient way for users to interact with multimedia applications. Traditional music control methods often require physical contact with devices, which may not always be practical in certain situations. Gesture-based interaction allows users to control music playback without touching the device, thereby improving accessibility and user comfort. The proposed system focuses on providing an efficient and user-friendly gesture recognition framework that can operate using commonly available hardware such as a standard webcam. By combining computer vision techniques with real-time gesture detection, the system aims to deliver smooth interaction and reliable performance. The study also highlights the potential of gesture-based interfaces in future multimedia systems and interactive technologies.

Keywords—Gesture Recognition, Human-Computer Interaction, Computer Vision, OpenCV, MediaPipe, Music Control

I. INTRODUCTION

Human-Computer Interaction has evolved rapidly with the development of modern technologies. Traditional computer interfaces rely on devices such as keyboards, mice, and touch screens. Although these devices provide accurate control, they require physical interaction with the system.

In many real-life situations, users may not be able to interact with devices using touch-based interfaces. For example, while cooking, exercising, or performing multiple tasks simultaneously, touching a device repeatedly may be inconvenient. Gesture recognition technology provides an effective solution by enabling touchless interaction.

Gesture recognition systems use cameras and computer vision algorithms to detect hand movements and interpret them as commands. With the help of frameworks such as OpenCV and MediaPipe, developers can detect hand landmarks and analyze finger positions with high accuracy.

The FlowBeat system uses this concept to develop a gesture-controlled music player. The system captures real-time video from a webcam, detects hand gestures, and converts them into music control commands.

II. PROBLEM DEFINITION

Existing gesture recognition systems often depend on specialized hardware devices such as infrared sensors, depth cameras, or motion controllers. These devices increase the cost of the system and limit accessibility for general users.

Vision-based gesture recognition using standard webcams provides a cost-effective alternative. However, such systems must handle challenges such as lighting variations, background clutter, and camera quality differences.

The main challenge is to design a system that can detect gestures accurately while maintaining real-time performance and simplicity for users.

Another important challenge in gesture recognition systems is ensuring reliable performance in different environmental conditions. Factors such as poor lighting, shadows, and complex backgrounds can affect the accuracy of hand

detection. These variations may lead to incorrect gesture recognition, which can reduce the overall usability of the system. Therefore, designing algorithms that can adapt to different environments is an essential requirement.

In addition, real-time processing is a critical aspect of gesture-based interaction systems. The system must quickly capture video frames, analyze hand movements, and execute the corresponding command without noticeable delay. If the system responds slowly, the user experience may be negatively affected. Achieving both high accuracy and fast processing speed is therefore a significant technical challenge.

Despite these challenges, the development of efficient computer vision frameworks and machine learning techniques has created new opportunities for gesture recognition systems. With proper algorithm design and optimized processing methods, it is possible to build reliable vision-based systems that provide smooth and intuitive user interaction. This motivates the development of solutions like FlowBeat that aim to deliver a practical and user-friendly gesture-controlled music system.

III. OBJECTIVES

The objectives of the FlowBeat system are:

- Study existing gesture recognition technologies.
- Implement real-time hand landmark detection using MediaPipe.
- Classify gestures and map them to music control commands.
- Develop a low-cost and user-friendly gesture-based music control system.

IV. LITERATURE REVIEW

Gesture recognition systems can generally be classified into two categories: sensor-based systems and vision-based systems.

Sensor-based systems use hardware components such as accelerometers, gyroscopes, and motion sensors to detect hand movements. Devices such as Microsoft Kinect and Leap Motion provide accurate gesture tracking but require dedicated hardware.

Vision-based systems rely on cameras and image processing algorithms. OpenCV is commonly used for capturing frames and performing image processing tasks. MediaPipe provides a powerful framework for real-time hand tracking by detecting 21 hand landmarks.

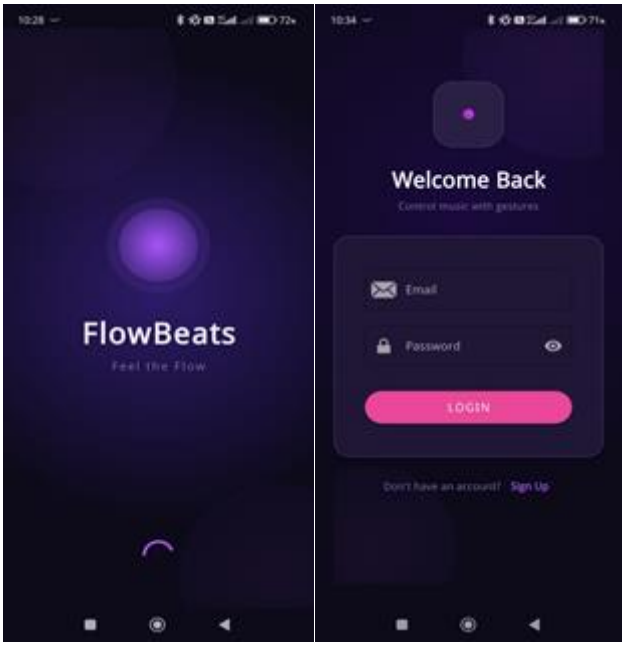
Machine learning techniques are often used to classify gestures based on landmark positions and finger movements.

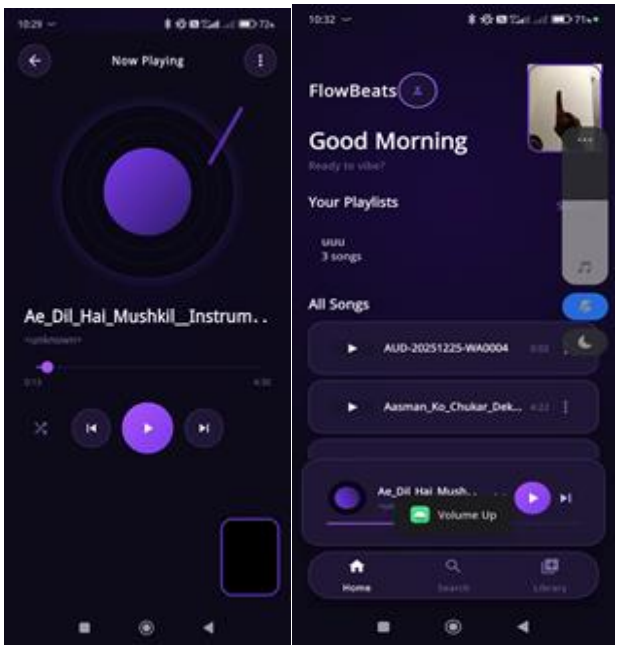
V. SYSTEM ARCHITECTURE

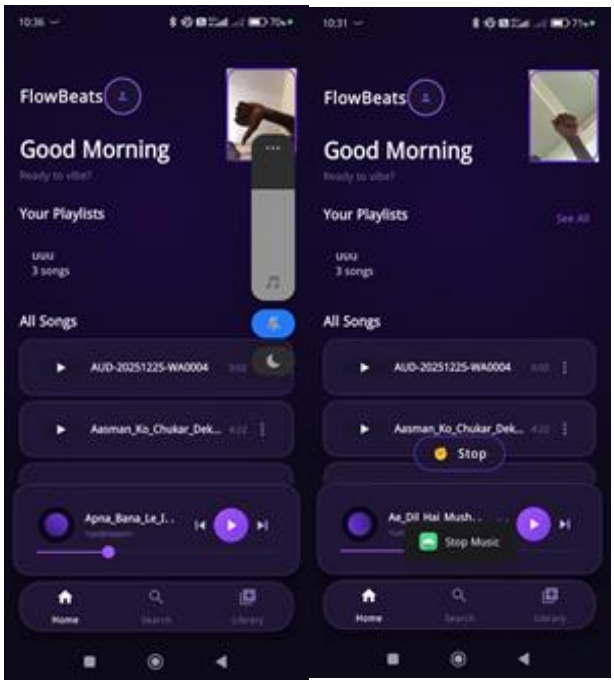
The FlowBeat system consists of several modules that work together to detect gestures and control music playback.

- Camera Module – Captures video input from the webcam.
- Frame Processing Module – Processes video frames using OpenCV.
- Hand Detection Module – Detects hand landmarks using MediaPipe.
- Gesture Classification Module – Identifies the performed gesture.
- Command Execution Module – Sends commands to the music player.

VI. IMPLEMENTATION







VIII. ALGORITHM

Step 1: Start the system and initialize the webcam.
Step 2: Capture continuous video frames.
Step 3: Process frames using OpenCV.
Step 4: Detect hand landmarks using MediaPipe.
Step 5: Analyze finger positions to recognize gestures.
Step 6: Map the gesture to a predefined command.
Step 7: Execute the command in the music player.
Step 8: Repeat the process for continuous control.

IX.ADVANTAGES

- Touchless and intuitive interaction.
- Low-cost implementation using a webcam.
- Easy to use and accessible.
- Enhances user convenience.
- Provides real-time music control.

X. FUTURE SCOPE

The system can be extended by integrating advanced machine learning models to improve gesture recognition accuracy. Additional gestures can be introduced to control volume, playlist navigation, and other multimedia features.

Another possible improvement is the integration of the system with mobile devices and smart applications. The gesture recognition module can be implemented in smartphone applications so that users can control music directly from their phones using the built-in camera. This would make the system more portable and accessible, allowing gesture-based interaction without requiring a dedicated computer setup.

Furthermore, the system can be enhanced by incorporating deep learning models and artificial intelligence techniques to improve gesture recognition under complex conditions such as low lighting or cluttered backgrounds. By training the model with larger gesture datasets, the system can achieve higher accuracy and support a wider variety of gestures, making the interaction more flexible and reliable.

The system may also be integrated with smart home devices, allowing users to control lights, appliances, and other systems using gestures.

XI. CONCLUSION

The FlowBeat gesture-controlled music system demonstrates how computer vision can enhance human-computer interaction. By using OpenCV and MediaPipe, the system can detect hand gestures and convert them into music control commands.
Overall, gesture recognition systems like FlowBeat represent an important step toward creating more natural and intuitive user interfaces. As computer vision and machine learning technologies continue to improve, gesture-based interaction is

expected to become more reliable and widely adopted in multimedia systems, smart devices, and interactive applications. This development will further enhance the way humans interact with digital technologies.

The implementation of the FlowBeat system also highlights the growing importance of vision-based interaction technologies in modern computing. By utilizing widely available tools such as OpenCV and MediaPipe, developers can build efficient gesture recognition systems without requiring expensive hardware. This makes gesture-based interfaces more accessible and practical for everyday applications.

The proposed approach provides a simple and cost-effective alternative to traditional touch-based interfaces and highlights the potential of gesture recognition technology in future interactive systems.

REFERENCES

1. Gesture Controlled Music Player – IJIREEICE
2. Gesture Based Multimedia Control – IJRP Reviews
3. MediaPipe Hand Tracking Documentation – Google Developers
4. OpenCV Documentation – OpenCV.org
5. Human-Computer Interaction Research – IEEE Xplore