IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Implementation and Performance Evaluation of Curriculum Learning and Behavior Cloning for Multi-Agent Freeze Tag Environment using Unity ML-Agents

[1]Sarang Shrihari Acharya, [2]Dr. Sushil V Kulkarni

[1]MTech Student, [2]Head Of Department,

[1]Computer Science and IT department,

MSBE's College Of Engineering, Ambajogai,India

Abstract: Artificial Intelligence (AI) and Computer Vision technologies have transformed the way users interact with digital systems by enabling contactless, intelligent, and immersive experiences. The integration of machine learning with browser-based applications has created new opportunities in retail, entertainment, education, and customer engagement. This paper presents the design and implementation of an AI-powered web platform that combines virtual clothing visualization, gesture-controlled interactive games, augmented reality hairstyle preview, avatar animation, automated customer greeting, and smart multimedia display management into a unified browser application.

The proposed system employs MediaPipe landmark detection models for face, hand, and body pose estimation, enabling real-time human-computer interaction without requiring specialized hardware. A virtual clothing try-on prototype utilizes image capture, cloth sample extraction, prompt-guided AI image generation, and cloud-based processing to visualize garments on users while preserving facial identity and body posture. Interactive modules including Bubble Pop, Ball Bounce, Clap Firework, and Mirror Avatar demonstrate the application of computer vision for entertainment using natural human gestures instead of conventional input devices.

Keywords: Artificial Intelligence, Computer Vision, Machine Learning, Virtual Try-On, MediaPipe, Gesture Recognition, Browser-Based AI, Human Computer Interaction, Three.js, Virtual Avatar.

1. Introduction

Artificial Intelligence has significantly transformed modern software engineering by enabling applications to perceive, interpret, and interact with users in ways that were previously impossible using conventional programming techniques. Rapid advancements in computer vision, deep learning, browser technologies, and cloud computing have accelerated the development of intelligent applications capable of performing real-time object detection, facial analysis, gesture recognition, virtual visualization, and automated decision making.

The retail industry has particularly benefited from these innovations through virtual fitting rooms, AI-assisted product recommendations, digital mirrors, and personalized shopping experiences. Simultaneously, gesture-controlled gaming and touchless interaction have gained increasing attention in education, healthcare, museums, exhibitions, and smart environments where physical interaction should be minimized while maintaining an engaging user experience.

Traditional virtual try-on systems frequently require specialized hardware, depth sensors, expensive cameras, or dedicated mobile applications. Similarly, gesture-based entertainment systems often depend on gaming consoles equipped with proprietary sensors, making them costly and less accessible for educational institutions and small organizations. Recent developments in browser-based computer vision have demonstrated that standard webcams, combined with lightweight AI models, can provide highly responsive interactive experiences without requiring dedicated hardware installations.

The proposed system addresses these challenges by developing a unified browser-based AI platform capable of integrating multiple intelligent applications within a single architecture. The implementation combines computer vision, augmented reality concepts, image generation, gesture recognition, and database-driven content management into one scalable framework. Instead of treating each AI application as an independent solution, the proposed architecture shares perception modules, interaction pipelines, and rendering mechanisms across multiple functional components.

The developed platform consists of several independent yet interconnected modules including an AI-powered clothing visualization prototype, real-time avatar animation, gesture-controlled gaming applications, augmented hairstyle preview, automated visitor greeting system, and a database-supported smart slideshow management system. Each module demonstrates a different practical application of Artificial Intelligence while sharing a common computer vision infrastructure.

MediaPipe-based landmark detection serves as the perception layer of the system, providing facial landmarks, body pose estimation, and hand tracking. These landmarks are processed through smoothing algorithms and rule-based interaction logic before being translated into visual responses including avatar animation, game interaction, image generation requests, or multimedia playback. The modular architecture enables individual components to operate independently while maintaining consistency across the overall platform.

From an engineering perspective, the project demonstrates how AI technologies can be integrated using widely available web development tools including HTML5, CSS3, JavaScript, PHP, MariaDB, Three.js, Canvas API, and cloud-based AI services. The resulting platform remains hardware independent, browser compatible, scalable, and suitable for deployment across educational institutions, retail stores, exhibitions, and customer engagement environments.

The implementation further illustrates that combining deterministic rule-based algorithms with AI perception produces more stable and explainable interactive systems than relying exclusively on deep learning. This hybrid approach improves responsiveness while reducing computational complexity, making the solution practical for ordinary computing environments.

2. Problem Statement

Although Artificial Intelligence has introduced numerous advancements in virtual shopping and interactive media, most existing systems remain limited to a single functionality such as virtual fitting, gesture recognition, or augmented reality visualization. Commercial virtual try-on applications often require proprietary hardware, expensive mobile platforms, or cloud infrastructures that are not affordable for educational institutions, small retailers, and prototype developers.

Similarly, existing gesture-controlled gaming applications typically operate as isolated systems without integration into broader customer engagement platforms. Most browser-based demonstrations focus only on showcasing computer vision algorithms rather than providing complete deployable solutions capable of supporting real-world business scenarios.

Furthermore, conventional retail environments still rely heavily on physical product trials, printed promotional materials, and manual customer interaction. These approaches increase operational costs while limiting customer engagement and personalization.

Therefore, there exists a need for an integrated browser-based AI platform capable of combining virtual clothing visualization, gesture-based entertainment, augmented reality interaction, intelligent multimedia management, and automated customer engagement using affordable hardware and widely available web technologies.

The proposed implementation addresses this research gap by developing a unified intelligent platform that demonstrates multiple practical applications of Artificial Intelligence within a single modular architecture while maintaining scalability, usability, and deployment simplicity.

3. Proposed System

The proposed system adopts a modular, browser-based architecture that integrates Artificial Intelligence, Computer Vision, Machine Learning, Augmented Reality, cloud-based image generation, and database-driven content management into a unified interactive platform. The architecture has been designed to maximize modularity, scalability, and real-time responsiveness while allowing each intelligent module to operate independently using a shared perception pipeline.

Unlike conventional AI applications that focus on a single feature, the proposed platform combines multiple intelligent services, including AI-powered clothing visualization, virtual avatar animation, gesture-controlled games, hairstyle preview, automated greeting, and multimedia slideshow management. Each module is implemented as an independent component but shares common computer vision resources, reducing computational redundancy and simplifying system maintenance.

The overall architecture consists of six major layers:

3.1 User Interaction Layer

The User Interaction Layer represents the entry point of the system where users interact using a standard webcam through a web browser. This layer captures live video streams, still photographs, gesture inputs, and administrative commands without requiring additional sensors or specialized hardware. The browser interface provides responsive controls for selecting AI modules, capturing clothing samples, choosing avatars, previewing hairstyles, and managing multimedia content. Since all interaction occurs through ordinary webcams and web technologies, the proposed platform remains highly portable across desktop computers and laptops.

3.2 Computer Vision Processing Layer

The Computer Vision Layer performs real-time landmark detection using Google's MediaPipe framework. Three independent landmark detection models operate simultaneously depending on the selected module:

Face Landmark Detector

Hand Landmark Detector

Pose Landmark Detector

These models continuously estimate facial geometry, body posture, and hand positions from live video frames. The detected landmarks form the fundamental perception layer that drives all subsequent AI modules.

To improve tracking stability, the system applies smoothing algorithms and interpolation techniques that minimize landmark jitter caused by camera noise or sudden user movement. Mirror compensation logic further aligns the coordinate system with natural human perception, ensuring intuitive interaction between users and the virtual environment.

3.3 AI Processing Layer

The AI Processing Layer converts detected landmarks into meaningful interaction events. Instead of directly mapping landmark coordinates to output animations, this layer applies rule-based interpretation algorithms combined with AI-assisted image generation.

Different processing pipelines are implemented according to the selected application:

Facial landmarks drive avatar expressions and hairstyle alignment.

Hand landmarks control gesture-based games through collision detection.

Pose landmarks animate virtual characters.

Presence detection activates automated greeting functionality.

Captured images are transmitted to cloud-based AI models for clothing generation.

This hybrid AI approach combines deterministic algorithms with machine learning perception, producing stable and explainable system behavior while maintaining computational efficiency.

3.4 Application Layer

The Application Layer contains all intelligent modules developed within the proposed platform. Each module utilizes the common perception pipeline while implementing its own interaction logic.

The major application modules include:

AI Clothing Try-On Prototype

Mirror Avatar System

Bubble Pop Interactive Game

Clap Firework Game

Ball Bounce Physics Game

Haircut Preview Module

Namaste Greeting System

Smart Slideshow Management System

Because every application operates independently, future intelligent modules can be added without modifying the underlying computer vision infrastructure.

3.5 Backend Service Layer

Although most interaction occurs locally inside the browser, several modules require backend processing.

The backend is implemented using PHP and performs:

Image upload processing

AI clothing request management

Prompt generation

API communication

User authentication

Device validation

Media management

Database communication

The clothing visualization module communicates with cloud-based AI image generation services through REST-based APIs, while slideshow management interacts directly with the relational database.

3.6 Database Layer

The Database Layer is responsible for persistent storage of system configuration, media assets, user information, slideshow settings, TV registration, playlists, and display configuration.

MariaDB has been selected because of its lightweight deployment, relational integrity, and compatibility with PHP applications.

Separate relational tables maintain:

User information

TV registration

Device authentication

Uploaded images

Uploaded videos

Slideshow playlists

Video playlists

Display settings

Scrolling ticker configuration

Social media configuration

The normalized database design minimizes redundancy while improving maintainability and scalability.

4. Implementation

The proposed intelligent platform has been designed as a collection of independent yet interconnected modules that utilize a common computer vision pipeline. Each module implements a unique interaction mechanism while sharing the same perception layer consisting of face, hand, and body landmark detection. This modular design improves maintainability, scalability, and code reusability by separating perception, decision making, rendering, and backend communication.

The implementation follows a browser-first approach where most AI computations execute locally using MediaPipe and JavaScript, while computationally intensive image generation tasks are delegated to cloud-based AI services. Such architectural separation minimizes latency for real-time interactions while maintaining high-quality visual generation for asynchronous modules.

Implementation Algorithm

The implementation follows the algorithm shown below.

Algorithm 1: AI Clothing Generation

Step 1: Start webcam.
Step 2: Capture user photograph.
Step 3: Select cloth sample region.
Step 4: Crop selected fabric image.
Step 5: Generate AI prompt.
Step 6: Upload photograph and fabric image.
Step 7: Invoke cloud image generation API.
Step 8: Receive generated output.
Step 9: Display virtual try-on image.

Implementation Process

The webcam continuously captures live video frames which are processed by MediaPipe Face and Pose Land marker models.

Detected landmarks include

Head orientation
Eye movement
Shoulder position
Neck rotation
Upper-arm movement
Elbow coordinates

Raw landmark values often fluctuate because of camera noise and lighting variations. Therefore, each coordinate passes through an interpolation
filter before being mapped onto the VRM skeleton.

Three.js renders the animated avatar inside a white virtual environment. The white background intentionally removes visual distractions and allows observers to focus exclusively on avatar movement during demonstrations.

The browser continuously updates the avatar skeleton approximately every rendering frame, producing smooth motion synchronization between the user and the virtual character.

5. Experimental Results

The developed platform was evaluated through a series of functional and performance-oriented experiments to verify the correctness, responsiveness, and usability of each implemented module. Testing was conducted in a controlled indoor environment using a standard laptop equipped with an integrated webcam. The objective was to evaluate whether the proposed browser-based architecture could deliver real-time AI interaction without specialized hardware.

The experimental environment consisted of a web browser running the developed application, a local Apache server with PHP support, and a MariaDB database for multimedia management. AI-powered clothing generation was executed through cloud-based image generation services,
whereas all gesture-recognition modules operated locally within the browser.
The evaluation focused on four primary performance indicators:
Functional correctness
Interaction responsiveness
User experience
Computational efficiency
Each intelligent module was executed multiple times under different user positions, lighting conditions, and interaction patterns to evaluate its operational stability.

6. Comparative Analysis

Existing AI-based virtual try-on and gesture-recognition systems generally focus on solving a single problem. Commercial retail applications emphasize clothing visualization, while educational computer vision demonstrations concentrate on gesture recognition. Few systems integrate multiple AI capabilities within one unified architecture.
The proposed platform differs from existing approaches by combining virtual clothing generation, avatar animation, gesture-controlled games, augmented reality visualization, automated visitor greeting, and multimedia management into a single browser-based application.
Unlike proprietary commercial systems, the proposed implementation operates entirely using ordinary consumer hardware without requiring depth cameras or dedicated sensors.

7. Conclusion

The rapid advancement of Artificial Intelligence (AI), Machine Learning (ML), and Computer Vision has enabled the development of intelligent systems capable of transforming traditional human-computer interaction into immersive, touchless, and personalized digital experiences. This research presented the design and implementation of a browser-based intelligent platform integrating AI-powered virtual clothing visualization, gesture-controlled interactive games, real-time avatar animation, augmented reality hairstyle preview, automated visitor greeting, and a database-driven multimedia slideshow management system within a unified software architecture.
Unlike conventional solutions that primarily address a single application domain, the proposed platform combines multiple AI-driven modules using a shared perception pipeline based on MediaPipe landmark detection. The modular software architecture allows independent operation of each component while maintaining seamless integration through common frontend, backend, and database services. This approach significantly improves software maintainability, scalability, and code reusability.

8. Future Work

Although the proposed implementation successfully achieves its primary objectives, numerous opportunities remain for extending the platform with more advanced Artificial Intelligence capabilities. Future research can focus on improving realism, expanding application domains, and enhancing system scalability.

8.1 Advanced AI Virtual Try-On

Future versions of the platform can incorporate diffusion-based image generation models capable of producing photorealistic virtual clothing with improved garment fitting, wrinkle simulation, fabric texture preservation, and body-aware rendering. Integration of three-dimensional garment reconstruction would further improve realism for commercial retail applications.

8.2 AI Jewelry and Accessory Try-On

The existing clothing visualization framework can be extended to support intelligent visualization of jewelry, watches, handbags, footwear, sunglasses, and cosmetic products. This enhancement would transform the platform into a comprehensive virtual shopping assistant suitable for fashion retail and e-commerce applications.

9. REFERENCES

1. Google AI Edge, *Face Landmarker for Web Documentation*, Google Developers, 2025.
2. Google AI Edge, *Hand Landmarker for Web Documentation*, Google Developers, 2025.
3. Google AI Edge, *Pose Landmarker Documentation*, Google Developers, 2025.
4. Google AI Edge, *MediaPipe Solutions Guide*, Google Developers, 2025.
5. Three.js Documentation, "Three.js JavaScript 3D Library."
6. Pixiv, *three-vrm Documentation*, GitHub Repository.
7. Google AI for Developers, *Gemini API Documentation*.
8. Google AI for Developers, *Image Generation API Documentation*.
9. MDN Web Docs, *MediaDevices getUserMedia() API*.
10. MariaDB Documentation, *CREATE TABLE and Foreign Key Documentation*.
11. Apache Software Foundation, *Apache HTTP Server Documentation*.
12. Mozilla Developer Network, *HTML5 Canvas API Documentation*.
13. Mozilla Developer Network, *JavaScript Web APIs Documentation*.
14. W3C, *HTML5 Specification*.
15. W3C, *CSS3 Specification*.