Integrating computer vision and AI for interactive augmented reality experiences in new media

Ruonan Shi

University of Melbourne, Parkville VIC 3010, Australia

haleyssrn@gmail.com

Abstract. Augmented reality (AR) is a groundbreaking technology that fully immerse the user in a mixed 'reality' where the real and virtual coexist in a unique manner. Integrating more artificial intelligence (AI) and computer vision into AR devices can greatly improve user input and provide a whole novel interface. Through AI technology, such as gesture recognition, object tracking, and face recognition, AR systems can offer more intuitive and engaging interactions. An AR system featuring such improved AI technologies can process real-time data while having the contextual awareness to respond to users' input and the surrounding environment on the fly. It can also provide narrative integration, character development, and dynamic environments to users, thereby enabling them to have a more personalised and meaningful experience. This paper examines the evolution of new media by discussing the possibilities of using AI and computer vision in AR devices to create personalised experiences, all the while critically looking at technical challenges and the opportunities they present to the field of future research and development. It also looks at case studies across different sectors, such as education, training, tourism, gaming, retail, and aviation to justify the potential of future development of AI-enhanced AR.

Keywords: Augmented Reality, Artificial Intelligence, Computer Vision, User Interaction, Gesture Recognition.

1. Introduction

Augmented reality (AR) alters the conventional paradigm of user interaction with digital content by superimposing virtual objects onto the physical world, thereby enabling contextual immersion in multiple domains. While developers are accelerating the development of AR devices and applications, they are increasingly examining ways to improve the user experience from interaction perspectives. Recent advancements in artificial intelligence (AI) and computer vision technologies are emerging as important solutions for this purpose. By tracking, recognising and analysing user inputs as well as information gathered from the environment, these technologies enable AR applications to interact with users or the surrounding environment in a dynamic and adaptive manner. Gesture recognition enables users to manipulate virtual objects by natural hand-wave motions. Meanwhile, visual object tracking ensures that virtual content would be anchored to appropriate locations in the physical scene. Facial recognition has been frequently used across various AR applications in order to recognise users' facial expressions and features, thereby providing personalised AR content. The AI-driven content adaptation makes AR application content dynamically and contextually aware. Narrative technologies and AI could

be utilised to create dynamic and emotionally rich stories and characters for immersive AR experiences. This paper focuses on exploring the potential of AI and computer vision in enhancing human-to-computer interaction in AR. Based on comprehensive reviews and a set of case studies, we analysed three key technologies and representative use cases of these technologies in the AR domain, while demonstrating performance metrics to show that these AI and computer vision technologies are reshaping user interaction in AR [1].

2. Enhancing User Interaction with Computer Vision

2.1. Gesture Recognition

Computer vision-powered gesture-recognition technology enables users to interact with AR environments using natural hand movements. It opens up a new kind of interaction that's highly intuitive and immersive. Pattern-recognition systems can analyse your hand movements and gestures, and use this information to interpret commands and perform actions in your AR environment. This differs from earlier interaction models that rely on physical controllers or touchscreens. A gestural interaction model embeds a feeling of control more naturally into the experience, making your interaction seem more intuitive and lifelike. Imagine you're using an educational AR application where you're able to examine a 3D model of the human body. Shifting your gaze from the heart to the lungs, you point your finger at the neck and open your palm to expand the larynx. You also roll your hand outwards to turn the stomach. From an interaction perspective, it would be difficult to find a better way of engaging with this complex anatomy model. It just feels so natural. So how can gesture-recognition technology be built? [2] Algorithm-wise, it requires highly precise computer vision systems that are capable of tracking and recognising a vast number of hand movements under a variety of environmental conditions. Research in the field explores ways to make these systems faster and more accurate so that they're able to support more sophisticated and diverse interactions, opening the door to new use cases, and eventually broadening the scope for gesture-based AR applications.

2.2. Object Tracking

Object tracking is an essential part of interactive AR experiences, where digital content must be anchored to and accurately aligned with the physical world. Computer vision algorithms can track the position and orientation of objects in real time, ensuring that virtual content is stably rendered in place as a user moves and interacts with their environment. Without the ability to track an object, a simple AR experience, such as adding a floating label with some text to a real-world product on a shelf, would lose its integrity as soon as the user moves their head and breaks the line of sight with the physical object being augmented. In retail, AR can track products sitting on a shelf, and place information over the top of them, such as a price, user reviews or coupons. In manufacturing, object tracking may be used to guide workers along an assembly line, keeping them informed as to the correct placement and sequence of physical parts. In all of these cases, the goal is to enable users to seamlessly interact with smart, digitally enhanced reality [3]. Research in this area is heavily focused on improving the accuracy and reducing the latency of object-tracking systems-especially in dynamic cluttered environments containing many moving objects.

2.3. Facial Recognition

Facial recognition is an exploit that could leverage the pervasiveness of mobile phones for AR to create 'me-centric' experiences. Based on this input, the system could identify facial features and extract expressions, and use the extracted information to tailor the AR content for a user's context. The application of facial recognition to AR could increase user engagement by creating a personalised experience. Some of these experiences might include: customising an AR avatar based on a user's facial features; generating dynamic faces of avatars or game characters in accordance with the user's facial expressions, all of which should be synchronised in real time; and building emotions into AR content, where an avatar's expression could be mapped to the user's [4]. Some existing applications of facial

recognition to social media focus on creating personalised emojis or avatars that track a user's facial expressions in real time, to reflect the user's emotions. In the case of marketing, tools exploiting facial recognition might gauge a user's reactions to an advertisement. For instance, an advertisement could adapt to a user's feelings using tags to affect the user's mood in a positive way-like music that wouldn't normally be liked or an animated character that may appear to laugh when the user feels unhappy. The key to applying face recognition to AR is to develop algorithms that could operate on consumer-grade devices while also ensuring privacy and securely storing data. Current research towards improving the accuracy of facial recognition as well as its implementation time and energy consumption on resource-constrained devices could be applicable to this application [5].

Table 1 below depicts the performance metrics and user preferences for different application of AI and computer vision technologies in AR areas.

Table 1. AI and Computer Vision in AR Applications

Technology	Use Case	Application	Interaction Enhancement	Precision (%)	User Satisfaction (out of 10)	Response Time (ms)
Gesture Recognition	Education	3D Model Manipulation	Hands-on Learning	93.25273707	8.903972658	53.68544137
Gesture Recognition	Gaming	Complex Actions	Natural Interaction	93.06909213	9.272540131	67.86345438
Object Tracking	Retail	Product Information Overlay	Enhanced Shopping	97.22130853	9.012126616	51.77513944
Object Tracking	Manufacturing	Assembly Line Guidance	Real-time Instructions	96.68354491	9.102402275	29.46469686
Facial Recognition	Social Media	Personalized Emojis	Authentic Interactions	88.08231365	8.921661674	55.09513556
Facial Recognition	Marketing	Emotion-based Content	Adaptive Advertising	90.00067294	8.022366256	51.52857315

3. AI-Driven Content Adaptation

3.1. Contextual Awareness

This is a major advantage that AR applications can take advantage of. AI is able to contextualise data collected from sensors and cameras, and consequently adjust content, based on the context in which the user operates. A detailed understanding of the user's context allows for more meaningful and relevant interactions, from delivering location-based information to adapting to ambient light, detecting environmental features, and reconfiguring the displayed content accordingly. AR navigation apps can generate directions in real time, based on the user's location and orientation. In museums, a context-aware AR system can assist the visitor and personalise the tour, based on the exhibits around the user and his or her interests [6]. For enabling contextual awareness, we need to integrate diverse data sources and algorithms that can process and interpret these data in real time. Research in this area involves making it more accurate and reducing the computational overhead so that more powerful context-aware systems can be implemented on mobile and wearable machines.

3.2. Real-Time Data Processing

Real-time data processing is also greatly important for AR applications. In order to provide fast and responsive interactions, AI algorithms must be able to process and analyse enormous amounts of data quickly. Through advancements in computer technologies and AI algorithms, real-time data processing can collect, store, analyse and filter relevant information in a fraction of a second, which allows an AR application to respond to user inputs and environmental changes instantly. For example, an AR application in sports training can process enormous amounts of data to analyse athletes' motions and immediately provide feedback to train their technique, improving their performance. Similarly, in retail, AR mirrors can process customer data and instantly generate outfit recommendations that are tailored to fit the customer. In order to provide real-time data processing to improve the overall user experience,

AI researchers are experimenting with new techniques to reduce latency caused by data processing. For instance, edge computing processes data closer to where data are generated and interacted with, eliminating the need for data to be initially uploaded to the cloud before processing [7]. This technique offers a faster and more reliable data-processing time, which is crucial for maintaining the AR experience. In figure 1, it is shown the performance metrics of real-time data processing by different AR applications [8]. These metrics are based on virtual data.
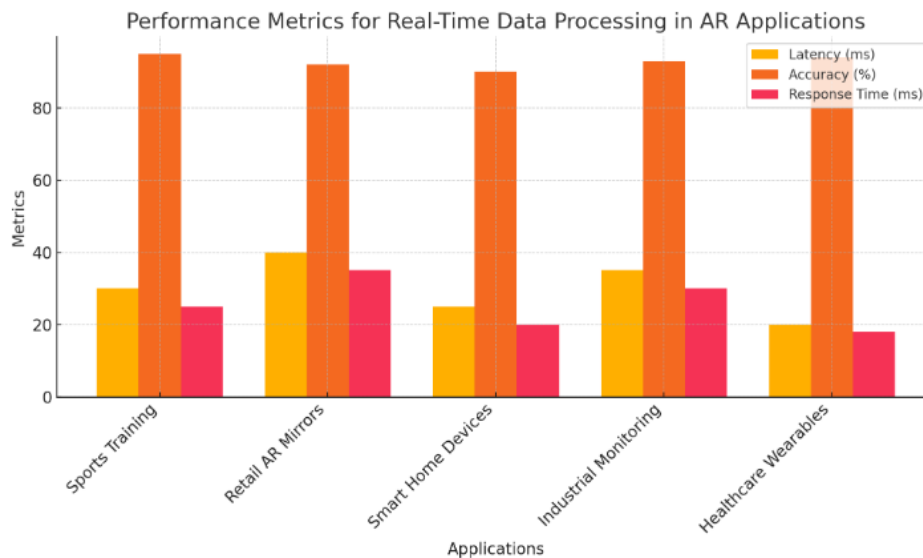

Figure 1. Performance Metrics for Real-Time Data Processing in AR Applications

3.3. Personalized Content Delivery

Alongside this, one of the most celebrated features of AI-augmented AR is easy delivery of personalised content. If an AR application analyses a user's preferences, behaviour and patterns of interaction, then the AI can make the content of the app more tailored to the user, thereby increasing user satisfaction and retention. Personalised content delivery can support more meaningful interactions between the user and the device, such as customised recommendations, adaptive learning paths, and timely marketing messages. For instance, AR education applications could modify their lessons based on each student's individual learning pace and style, by providing tailored resources, suggestions, and point of information. In marketing, aggregating previous data about consumers (purchases, preferences, demographics, geographical and contextual information, etc), AI algorithms can predict the right time and content of a promotion (graphic, sound, product displayed, etc) that has a higher chance to result in a conversion. This, in turn, leads to increased sales [9]. Retention is another aspect that can be enhanced thanks to personalised content delivery. Once a consumer is provided with a product or an experience that is tailored to their preferences, it is more likely that they will return for more [10].

4. Immersive Storytelling in AR

4.1. Narrative Integration

The integration of narrative elements can be useful in enhancing users' experience in AR experiences by creating meaningful interactions with narratives and thus compelling users' engagement. AI technologies can facilitate narrative integration in AR by creating immersive environments where users can interact with dynamic stories through AR toolkits that use narratives as contexts for human activities in our real world. The integration of AI-driven narrative can help to build interactive, adaptive and context-aware stories by communicating with accompanying data or triggers from additional sensors in AR environments to create engaging experiences. For instance, tourism can utilise an AI-driven

narratives in AR guides that tell historical stories and legends, and establish meaningful connections with informational contents related to historical tourists' location. Users can find their physical locations in augmented historical space by incorporating narrative contents when they visit ancient historical sites [11]. Table 2 shows various case studies of AI-driven narrative integration in AR applications [12]. These metrics demonstrate the effectiveness of narrative integration in enhancing user experiences across different sectors.

Table 2. AI-Driven Narrative Integration Case Studies

Case Study	Application	Narrative Type	Interactivity Level	User Engagement Score (out of 10)	Learning/Experience Enhancement (%)	Adaptability Index
Historical Tourism	AR Guides	Historical Stories	High	8.675284422	22.49504734	0.837686719
Education	Interactive Lessons	Educational Stories	Medium	8.842656573	19.62500221	0.862992146
Gaming	AR Adventures	Fantasy Narratives	Very High	9.183600026	26.77090394	0.960236165
Retail	Virtual Showrooms	Product Stories	Low	7.794634805	15.78545266	0.781035527
Healthcare	Patient Education	Medical Scenarios	Medium	8.256977185	20.09308756	0.826476124

4.2. Character Development
In this way, AI-driven character development is critical to developing AR stories and interactions that feel real and dynamic. Through natural language processing and machine learning algorithms, AI-driven characters can be developed that, for example, coyly respond to user inputs, can adjust their behaviour for different user inputs, and can exhibit realistic emotions. For instance, in gaming, they can further enhance the experience by developing conversations with the users, providing personalised advice and companionship throughout the game. Meanwhile, in educational AR apps, gameful learning can be augmented by learning from virtual tutors that can dynamically adjust their style and pace in response to users' feedback and progress, thereby personalising learning experiences. Realising AI-driven character development requires sophisticated algorithms that can understand and generate natural language, and, importantly, create believable emotional responses [13]. The challenge for researchers today in this area lies in improving the plausibility and interactivity of AI-driven characters so that they can become more effective in engaging users across different contexts and applications.

5. Conclusion
The marriage of AI and CV to AR can transform the potential for interactive and immersive real-time user experiences. The use of gesture recognition, object tracking and facial recognition engages users in a whole new way. By precision-tracking human movements, it could make AR UIs more intuitive and engaging. The AI's adaptive content and narrative could make AR more immersive and contextually aware. Convolutional Neural Networks (CNNs) can train a computer to recognise patterns in data, and Natural Language Processing (NLP) can read text as sounds, tones, intonations and voice inflections. Ultimately, it could awaken AR users to a more profound digital reality. While there are technical and ethical barriers to overcome, such as processing data in real time and respecting user privacy, ongoing RD will seize the potential for what's possible in AR. The case studies presented here have shown how AI-enhanced AR is already changing sectors in ways that weren't conceivable without computer vision and AI, and how these technologies might define the future of new media-a world where users interact with a technology that can learn and understand us more profoundly. As AI and CV technologies progress further, they will continue pushing AR innovation to the forefront, allowing users to interact with a more immersive real-life experience.

References

[1] Habil, Sandra Gamil Metry, Sara El-Deeb, and Noha El-Bassiouny. "The metaverse era: leveraging augmented reality in the creation of novel customer experience." Management & Sustainability: An Arab Review 3.1 (2024): 1-15.

[2] Buchner, Josef, and Michael Kerres. "Media comparison studies dominate comparative research on augmented reality in education." Computers & Education 195 (2023): 104711.

[3] Gong, Taeshik, and JungKun Park. "Effects of augmented reality technology characteristics on customer citizenship behavior." Journal of Retailing and Consumer Services 75 (2023): 103443.

[4] Kumar, Harish, Parul Gupta, and Sumedha Chauhan. "Meta-analysis of augmented reality marketing." Marketing Intelligence & Planning 41.1 (2023): 110-123.

[5] Dargan, Shaveta, et al. "Augmented reality: A comprehensive review." Archives of Computational Methods in Engineering 30.2 (2023): 1057-1080.

[6] Walk, Jannis, et al. "Artificial intelligence for sustainability: Facilitating sustainable smart product-service systems with computer vision." Journal of Cleaner Production 402 (2023): 136748.

[7] Khang, Alex, et al. "Application of Computer Vision (CV) in the Healthcare Ecosystem." Computer Vision and AI-Integrated IoT Technologies in the Medical Ecosystem. CRC Press, 2024. 1-16.

[8] Waelen, Rosalie A. "The ethics of computer vision: an overview in terms of power." AI and Ethics 4.2 (2024): 353-362.

[9] Khang, Alex, et al., eds. Computer vision and AI-integrated IoT technologies in the medical ecosystem. CRC Press, 2024.

[10] William, P., et al. "Crime analysis using computer vision approach with machine learning." Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022. Singapore: Springer Nature Singapore, 2023. 297-315.

[11] Górriz, Juan M., et al. "Computational approaches to explainable artificial intelligence: advances in theory, applications and trends." Information Fusion 100 (2023): 101945.

[12] Oliveira, Rosana Cavalcante de, and Rogério Diogne de Souza E. Silva. "Artificial intelligence in agriculture: benefits, challenges, and trends." Applied Sciences 13.13 (2023): 7405.

[13] Svanberg, Maja, et al. "Beyond AI Exposure: Which Tasks are Cost-Effective to Automate with Computer Vision?." Available at SSRN 4700751 (2024).