iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.58855

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Image Processing Techniques in Virtual and Augmented Reality: A Comprehensive Overview

Rohan Dutta
Student, RCCIIT, India

Abstract: ***Virtual Reality (VR) and Augmented Reality (AR) have become increasingly popular technologies with diverse applications ranging from entertainment and gaming to healthcare and education. Central to the immersive experiences offered by VR and AR are advanced image processing techniques. This report provides a comprehensive overview of the image processing techniques commonly employed in VR and AR systems, including image recognition, depth sensing, stereo vision, and more. Additionally, it discusses their applications, challenges, and future directions in the field.***
Keywords: Image Processing, VR, AR

I. INTRODUCTION

VR can be described as a 4D simulation of the real world, inc1uding the 3D geometry space, 1D time and the immersive or semi-immersive interaction interface. Generally, VR can be c1assified as hardware-based VR and computer-based VR. A hardware-based VR system depends on special VR hardware such as a head-mounted display, VR-glove, etc. A PC-based VR system is implemented using software on personal computers (PCs). It uses standard PC peripherals as input and output tools. Currently, a hardware-based VR system can be considered an immersive virtual scene, whereas a PC-based VR system is semi-immersive. Dedicated VR peripherals are usually too costly for many applications. As PC-based Internet technologies are developing rapidly, they present a promising alternative to hardware-based VR. Whereas AR is a new form of human-machine interaction that overlays computer-generated information on the real world environment (Reinhart and Patron 2003). AR en-hances the existing environment rather than replaces it, as in the case of VR. AR can potentially apply to all human senses, such as hearing, touching and even smelling (Azuma 1997). In addition to creating virtual objects, AR could also re-move real objects from a perceived environment. The information display and image overlay are context sensitive, which means that they depend on the observed objects. This novel technique can be combined with human abilities to benefit manufacturing and maintenance tasks greatly. AR technologies are both hardware and software intensive. Special equipment, such as head-mounted devices, wearable computing gears, global positioning systems, etc., are needed. Real-time tracking and computation is a must, since synchronization between the real and the virtual worlds must be achieved in the shortest possible time interval. [1]

II. IMPORTANCE OF IMAGE PROCESSING TECHNIQUES IN ENHANCING IMMERSION AND INTERACTION

Image processing is an important technology that can be used to enhance the intelligence and realism of virtual worlds. By enabling virtual environments to perceive and react to the world around them, image processing can create more immersive and engaging experiences for users. Some of the features of image processing used in VR technologies are: Object recognition, facial recognition, motion tracking and avatar creation.

Deep learning is one of the methods by which AR/VR applications can be developed.

Deep learning algorithms have greatly improved the accuracy of object recognition in AR, enabling real-time tracking of objects. This has opened up a wide range of applications, from retail to education to gaming.

One notable example of deep learning in AR is the "Google Lens" application. Google Lens uses a combination of deep learning and computer vision techniques to identify objects, landmarks, and text through the user's camera. By analyzing the scene, Google Lens provides relevant information, such as the name of a product, the address of a restaurant, or the definition of a word.

Another example of deep learning in VR is the "GANverse3D" system developed by researchers at NVIDIA. The system uses generative adversarial networks (GANs) to create highly detailed 3D models of objects from 2D images. By training the system on a large dataset of real-world objects, GANverse3D can generate highly realistic 3D models with remarkable accuracy.

©IJRASET: All Rights are Reserved

Nowadays it's easier to find VR related contents on media platforms like YouTube. The 360-degree recordings allow the user to explore the virtual world just by moving the mouse on a desktop PC. Mobile access to these videos is also possible: the best device for this is a smartphone with a VR viewer, such as "Cardboard"; or Google's current "Daydream" version(discontinued 2019). The image from this videos is divided vertically and slightly offset which gives our eyes a 3D effect.

With "Augmented Reality" (AR), virtual content can be used in the real world. The best example for this being the game "Pokemon Go" by Niantic.

III. IMAGE RECOGNITION

Image recognition, a subset of computer vision, plays a pivotal role in enhancing user experiences in virtual reality (VR) and augmented reality (AR) environments. Here's how image recognition is utilized in VR/AR:

1) Object Recognition and Tracking
2) Gesture Recognition
3) Facial Recognition and Expression analysis
4) Object interaction and manipulation
5) Environmental awareness

The performance constraints for virtual reality call into two classes: visual display constraint the constraint on the display frame rate of the environment required to provide the effects of immersion and presence; and the interactivity constraint, the constraint on the latency time from when the user provides an input to when the system provides a response (visual or otherwise) required for the user to have useful control over objects in the environment. I wish to stress that, though these constraints are related, they are distinct. The visual display constraint refers to the frame rate of the system, while the interactivity constraint refers to the lags in the system. One constraint may be satisfied while the other may fail. Failing to satisfy the visual display constraint will lead to the failure of the illusion of immersion and presence. Failing to satisfy the interactivity constraint will lead to the inability of the user to accurately control objects in the environment through direct manipulation. There are two components to each of these constraints: a bottom-line component which will apply to all virtual environments, based on human factors studies; and a stricter component which will apply to objects with fast-moving objects, based on the theory of sampling.[2]

(i) Virtual reality headsets attempt to help a user enjoy an immersive 3D environment by putting a screen in front of the user's eyes to eliminate their connection with the real world.

(ii) An autofocus lens is placed between each eye and the screen. The lenses are adjusted based on the movement and positioning of the eyes. This allows tracking of the user movement vis-a-vis the display.

(iii) On the other end is a device such as a computer or mobile device that generates and renders the visuals to the eye through the lenses on the headset.

(iv) The computer is connected to the headset via an HDMI cable to deliver visuals to the eye through the lenses. When using a dedicated mobile device to deliver the visuals, the phone may be mounted directly on the headset such that the lenses of the headset simply lie over the mobile device's display to magnify the images or sense the movement of the eyes with respect to the mobile device's image and to finally create the visuals.



Fig: VR Games

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

IV. DEPTH SENSING

Depth sensing is a fundamental aspect of virtual reality (VR) and augmented reality (AR) systems that enables accurate spatial perception and interaction with virtual objects in three-dimensional (3D) space.

Depth sensing refers to the ability of a system to accurately measure distances to objects or surfaces in the surrounding environment. The principle behind depth sensing involves capturing depth information using sensors or cameras, which can be based on various technologies such as time-of-flight (ToF), structured light, stereo vision, or depth from defocus. Depth cameras take millions of measurements to create a moving model or point cloud—of a real person, object, or place. Currently, most digital models in virtual reality must be designed in a computer and then added to the virtual world. Depth cameras offer new ways to capture real physical forms and project them into virtual reality environments.



Fig: Woman using a VR headset to see visualize objects

Like two eyes, two sensors on a depth camera record different views. Using trigonometry to compare them, it determines the 3D location of each point.

An example of such a dept sensor is the Xbox Kinect sensor. The Kinect contains three vital pieces that work together to detect your motion and create your physical image on the screen: an RGB color VGA video camera, a depth sensor, and a multi-array microphone.

The camera detects the red, green, and blue color components as well as body-type and facial features. It has a pixel resolution of 640x480 and a frame rate of 30 fps. This helps in facial recognition and body recognition.

The depth sensor contains a monochrome CMOS sensor and infrared projector that help create the 3D imagery throughout the room. It also measures the distance of each point of the player's body by transmitting invisible near-infrared light and measuring its "time of flight" after it reflects off the objects.

The microphone is actually an array of four microphones that can isolate the voices of the player from other background noises allowing players to use their voices as an added control feature.

These components come together to detect and track 48 different points on each player's body and repeats 30 times every second.[3]

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

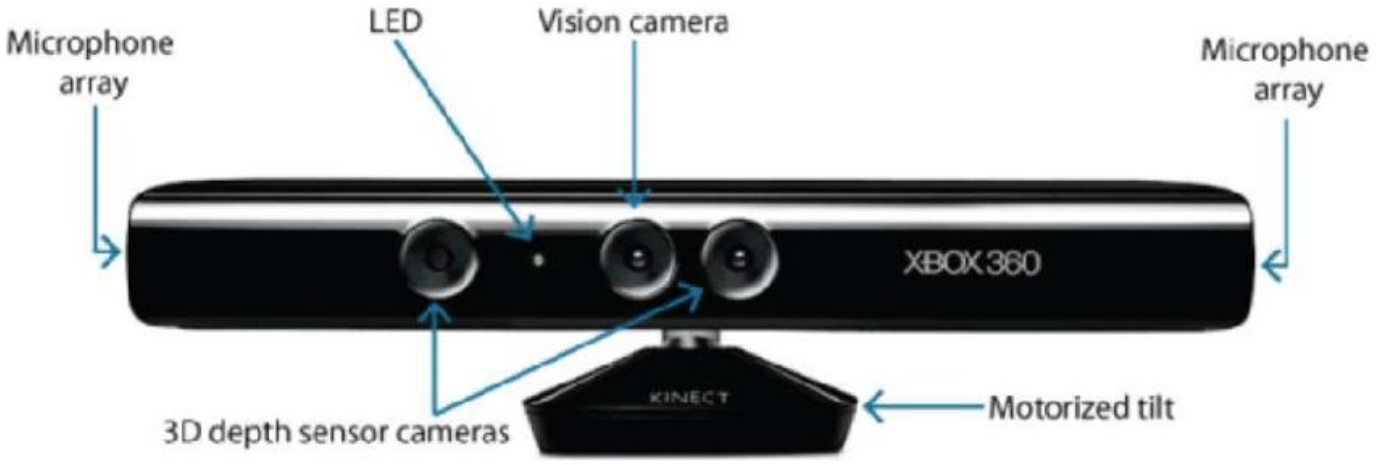


Fig: Kinect

V. FACTORS AFFECTING VISION IN VR

1) *Field of view (FOV:* FOV or the viewable area, is the extent to which the display will support the movement of the eye and head. It is the degree to which the device will contain the virtual world in front of our eyes.

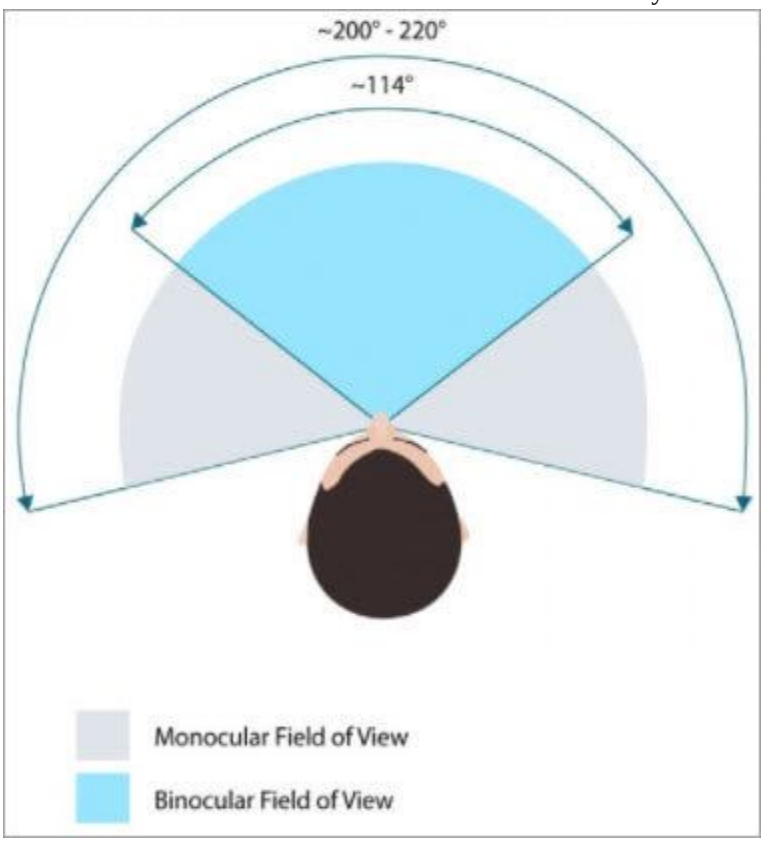


Fig: Binocular FOV and Monocular FOV

2) *Frame Rate:* The frame rate or the rate at which the GPU can process the visual images per second.
3) *The Screen Refresh Rate:* the pace of displaying the visual images.

Latency is a very important aspect related to the refreshing rate. For the brain to accept that the visual image generated on the screen is related to the head movement, the latency must be low to deliver the visual almost immediately. For Example, a lag of 7- 15 milliseconds is considered to be the ideal.

©IJRASET: All Rights are Reserved

VI. CHALLENGES AND FUTURE WORKS

Overall developing a robust VR system architecture implementing all the image processing techniques is not only challenging but also comes up with various issues like technical issues (These include tracking accuracy, latency, processing power, device battery life, field of view, and visual quality. For example, in VR, high-performance processing is crucial, and VR systems must capture user movements precisely), hardware issues (These include the cost and complexity of VR hardware and software, such as headsets, controllers, sensors, and computers. VR devices are often expensive, bulky, and require technical skills to set up and maintain.) or user adoption challenges (These include user comfort, motion sickness, eye strain, disorientation, and anxiety. Extended use of AR/VR can cause health and safety concerns such as headaches, motion sickness, and eye strain.) Also the problem of affordability stays as currently buying VR headset and other equipments is very costly.

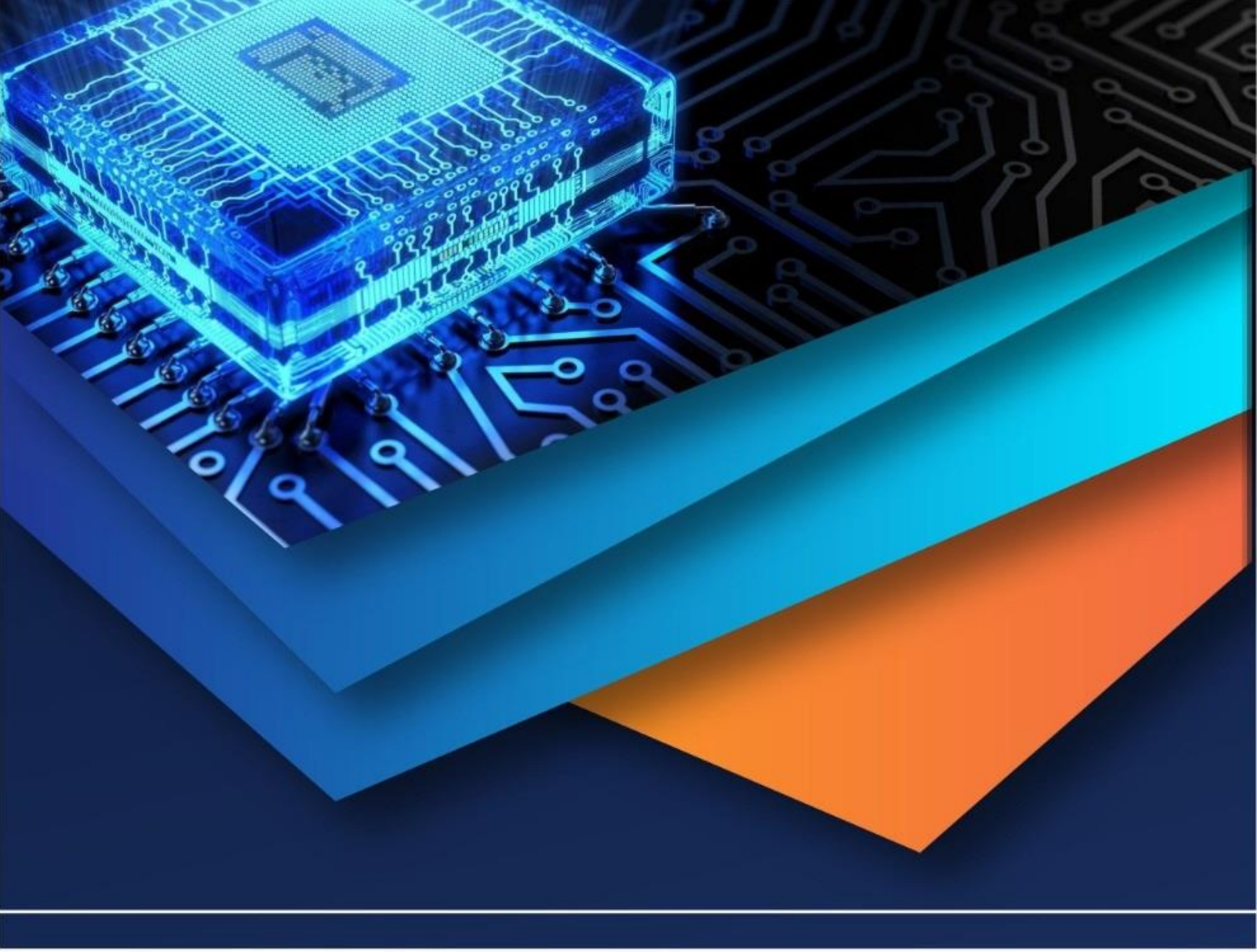
In the future I plan to make an optimized, less costly and more improved VR system.

VII. CONCLUSION

This report aims to provide a comprehensive understanding of the image processing techniques utilized in VR and AR systems, highlighting their significance in creating immersive and interactive experiences. By exploring the current state-of-the-art and future directions, it contributes to the ongoing advancements in these rapidly evolving technologies.

REFERENCES

[1] A Brief Introduction of VR and AR Applications in Manufacturing by SK Ong and AYC Nee DOI:10.1007/978-1-4471-3873-0_1
[2] Approaches to the Successful Design and Implementation of VR Applications by Steve Bryson Computer Science Corporation/NASA Ames Research Center Moffett Field, Ca.
[3] https://www.jameco.com/Jameco/workshop/Howitworks/xboxkinect.html



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)