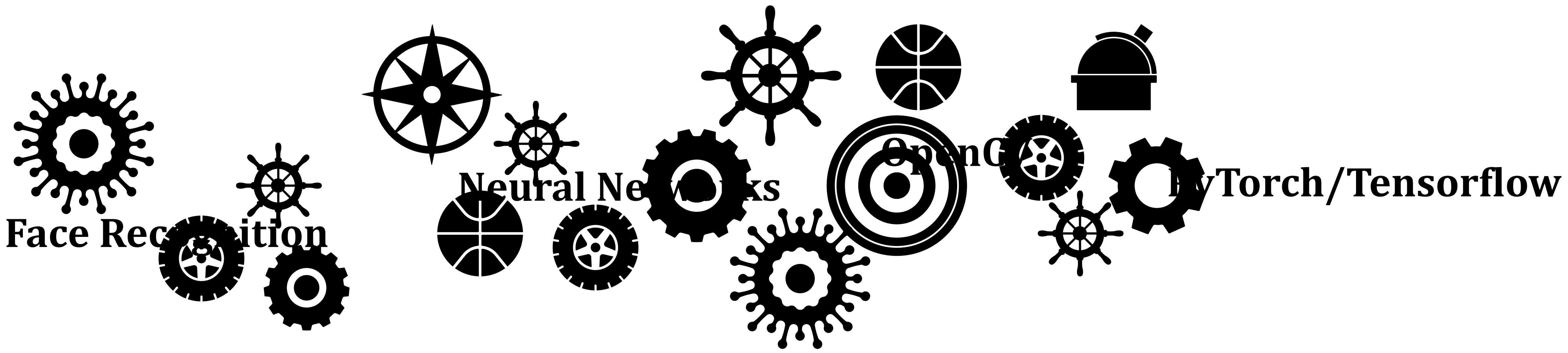
MPSSDEGB
कौशल से रोजगार, समृद्धि अपार

FOUNDATION FOR INNOVATION AND TECHNOLOGY TRANSFER
भारतीय प्रौद्योगिकी संस्थान दिल्ली
Indian Institute of Technology Delhi

Computer Vision



Deep Learning

Object Detection

Segmentation

Image Classification

3D Vision

The content is added from different sources

What can you determine about

1. the sizes of objects
2. the distances of objects from the camera?

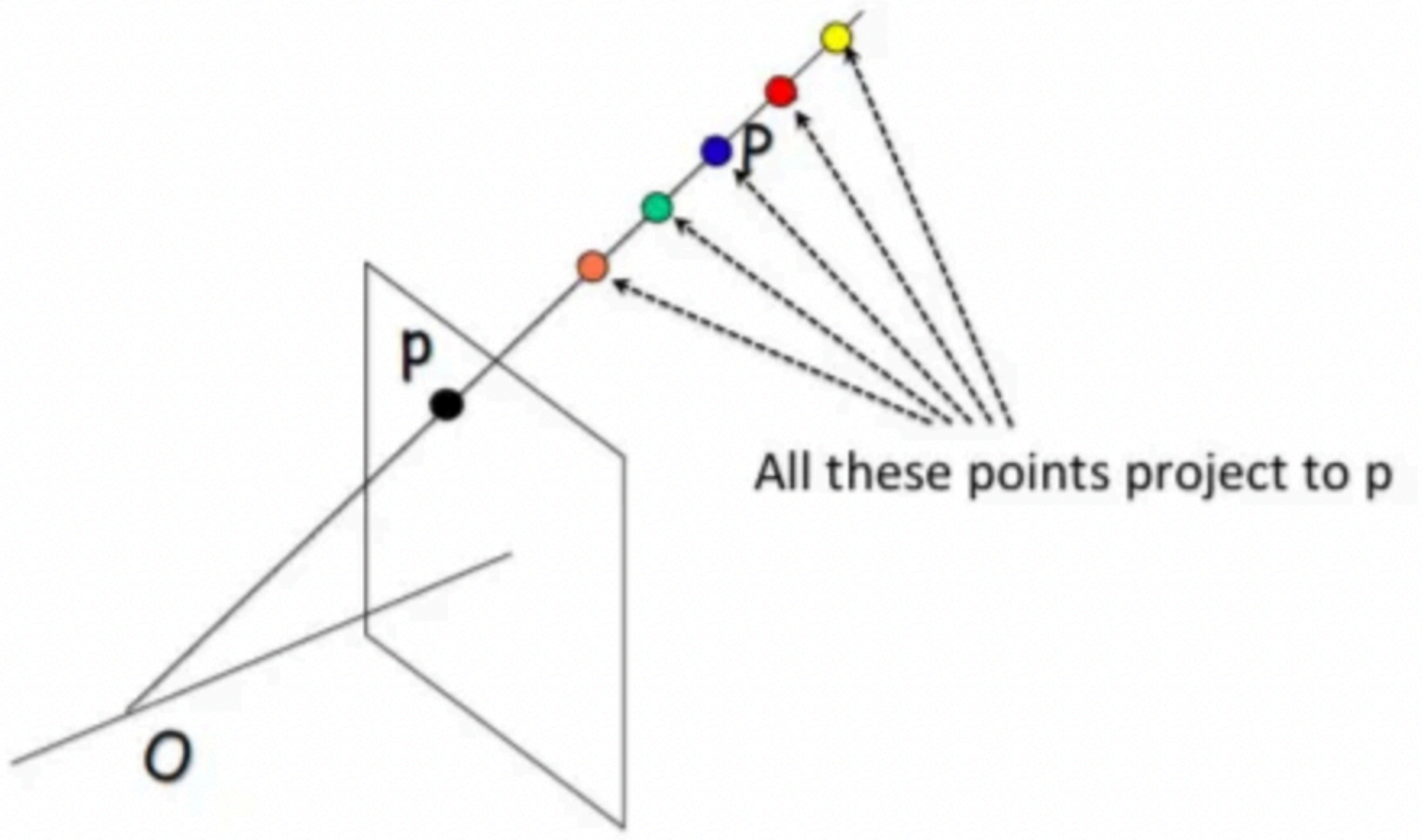
What knowledge do you use to analyze this image?

3D Vision

What objects are shown in this image?
How can you estimate distance from the camera?
What feature changes with distance?

Stereo Vision

It is not possible to estimate the distance (depth) of a point object 'P' from the camera using a single camera 'O'. This is because however close or far 'P' is on the projective line, it will map to the same point 'p' in the image.



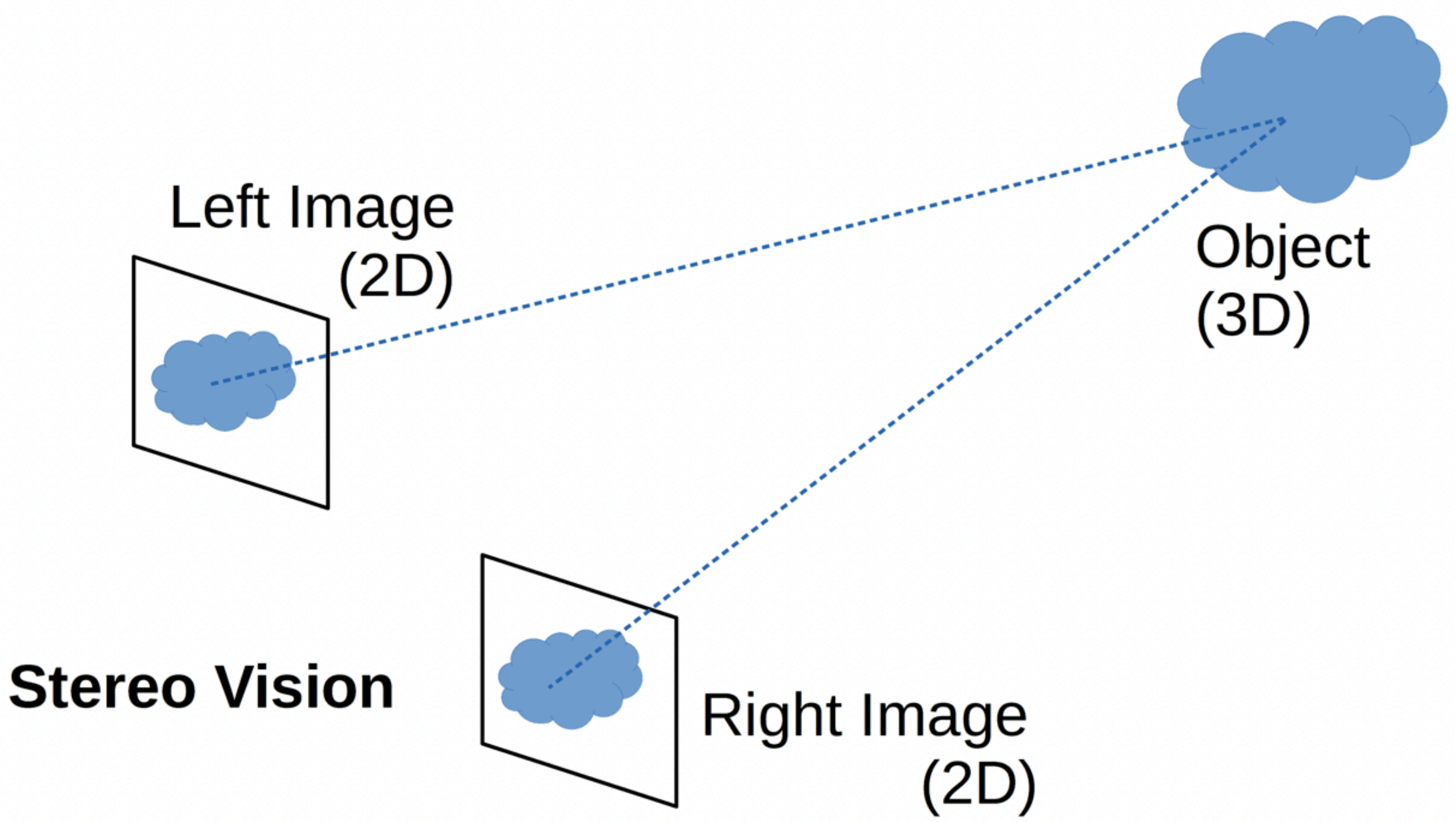
Source: [1]

https://medium.com/analytics-vidhya/distance-estimation-cf2f2fd709d8

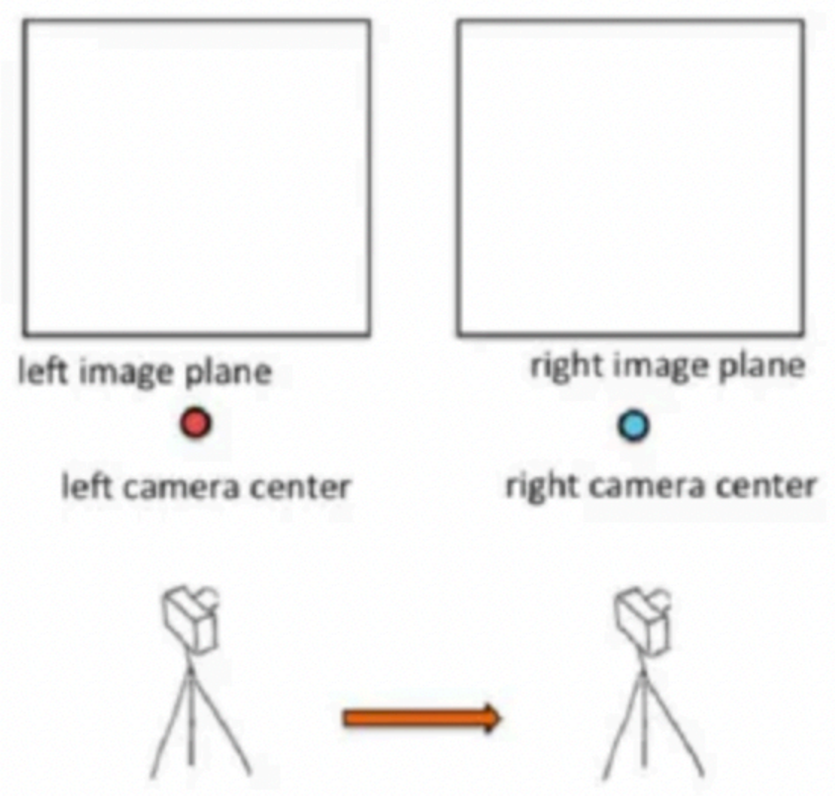
Stereo vision is a technique used to estimate the depth of a point object 'P' from the camera using two cameras. The foundation of stereo vision is similar to 3D perception in human vision and is based on the triangulation of rays from multiple viewpoints.

Computer stereo vision is the extraction of 3D information from 2D images, such as those produced by a CCD camera. It compares data from multiple perspectives and combines the relative positions of things in each view. As such, we use stereo vision in applications like advanced driver assistance systems and robot navigation.

It's similar to how human vision works. Our brains' simultaneous integration of the images from both of our eyes results in 3D vision:

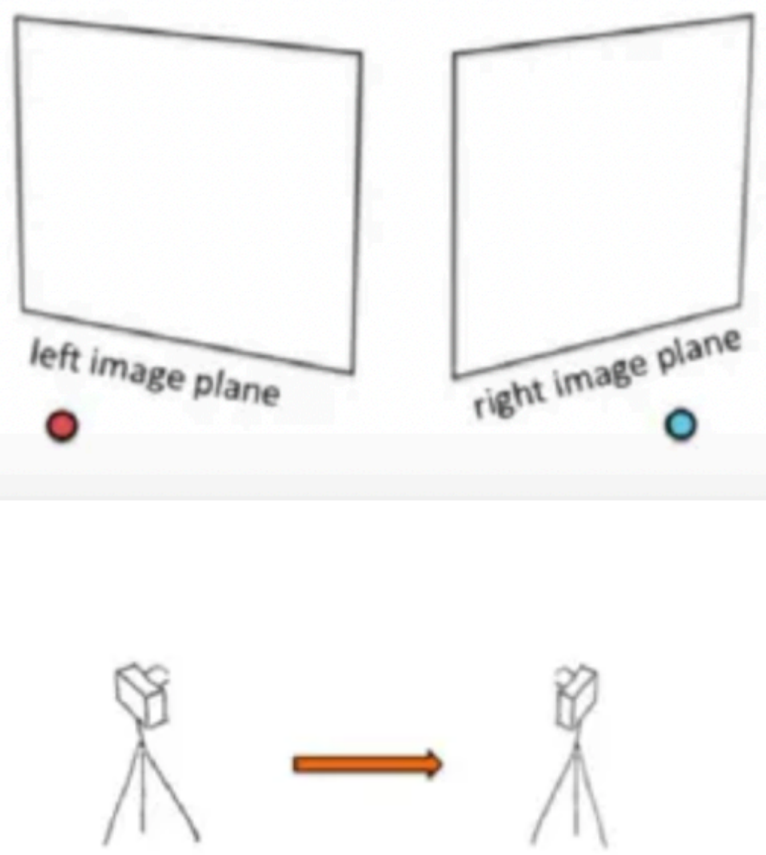


1. Parallel system



Source: [1]

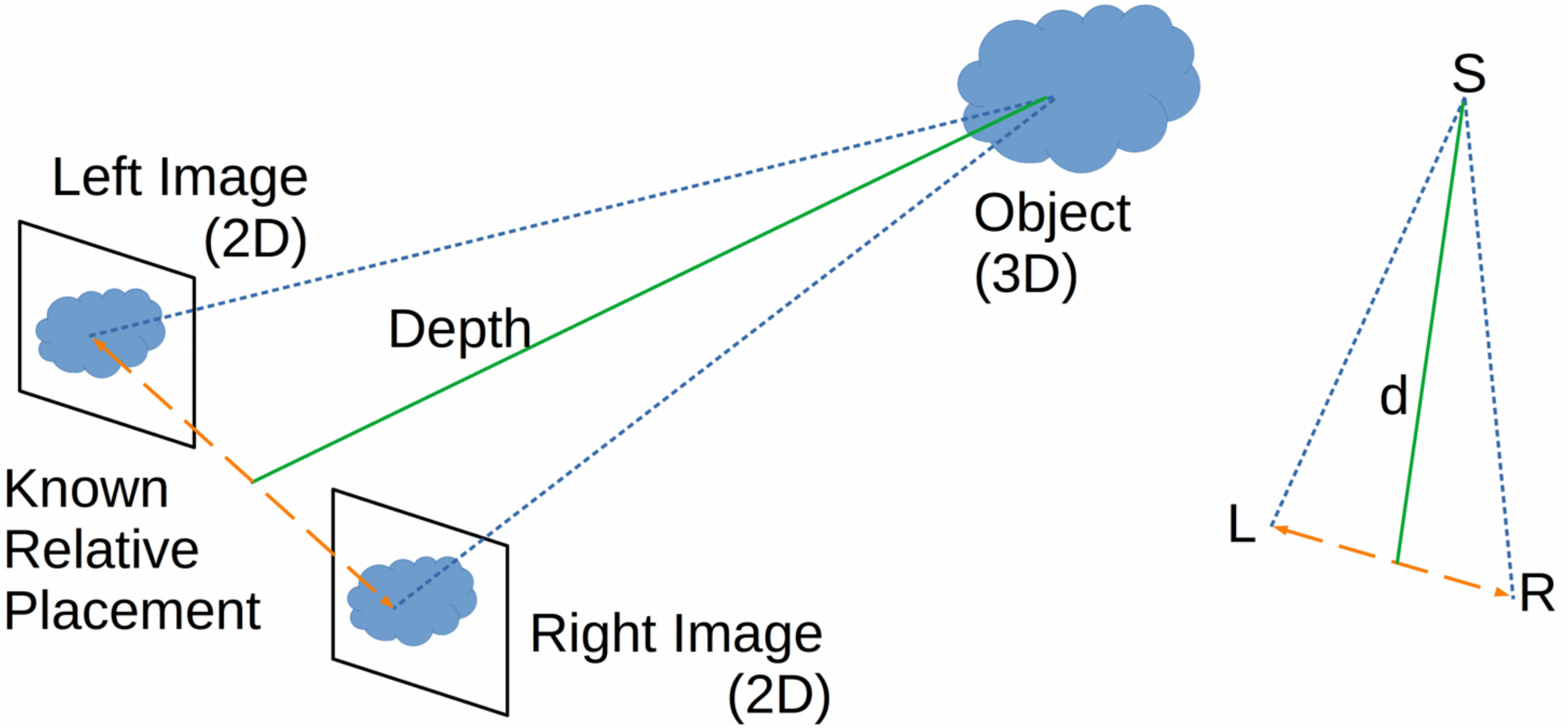
2. General system



https://www.baeldung.com/cs/stereo-vision-3d

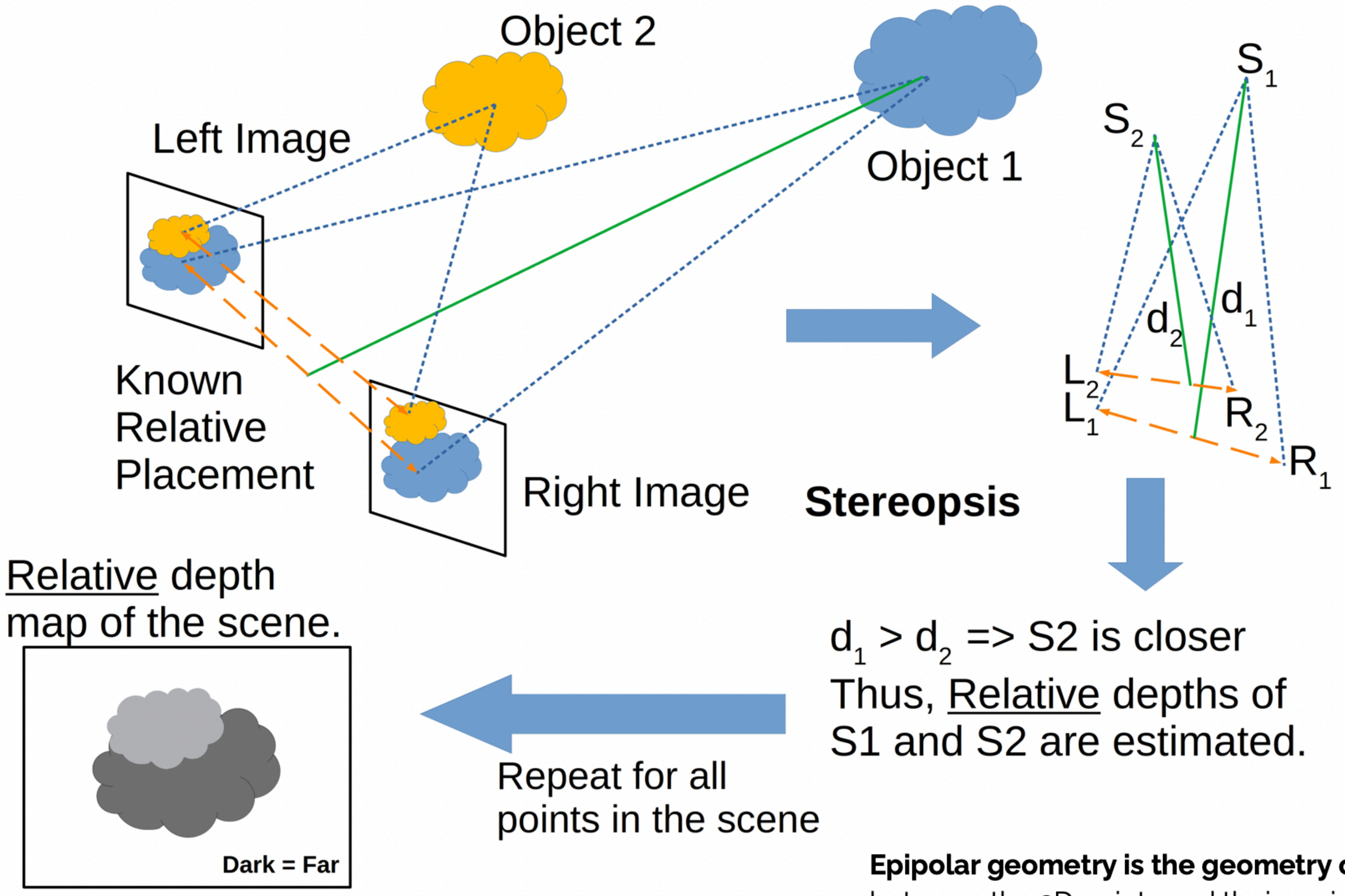
Stereo Vision

The content is added from different sources



https://www.baeldung.com/cs/stereo-vision-3d

Stereo Vision



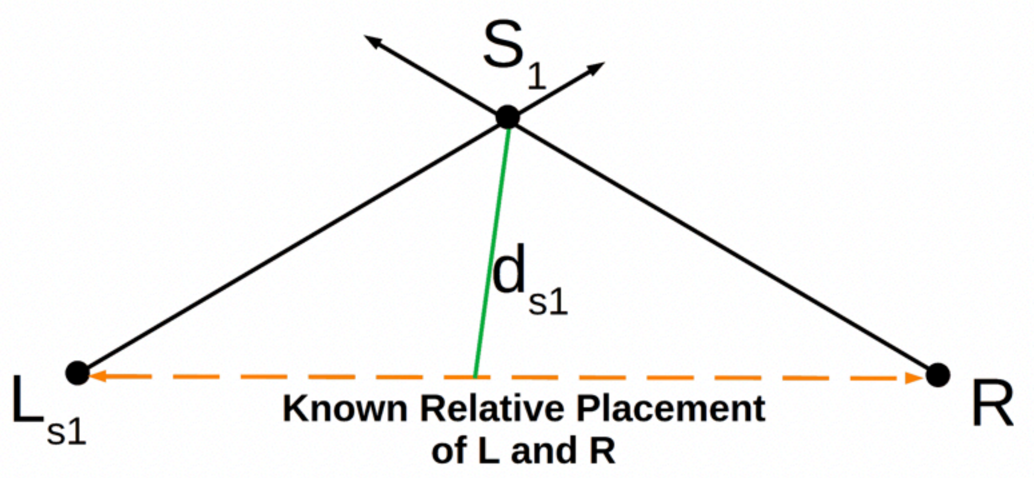
Epipolar geometry is the geometry of stereo vision. There are a variety of geometric relationships between the 3D points and their projections onto the 2D images. These relationships have been developed for the pinhole camera model. We assume that we can represent normal using these relationships.

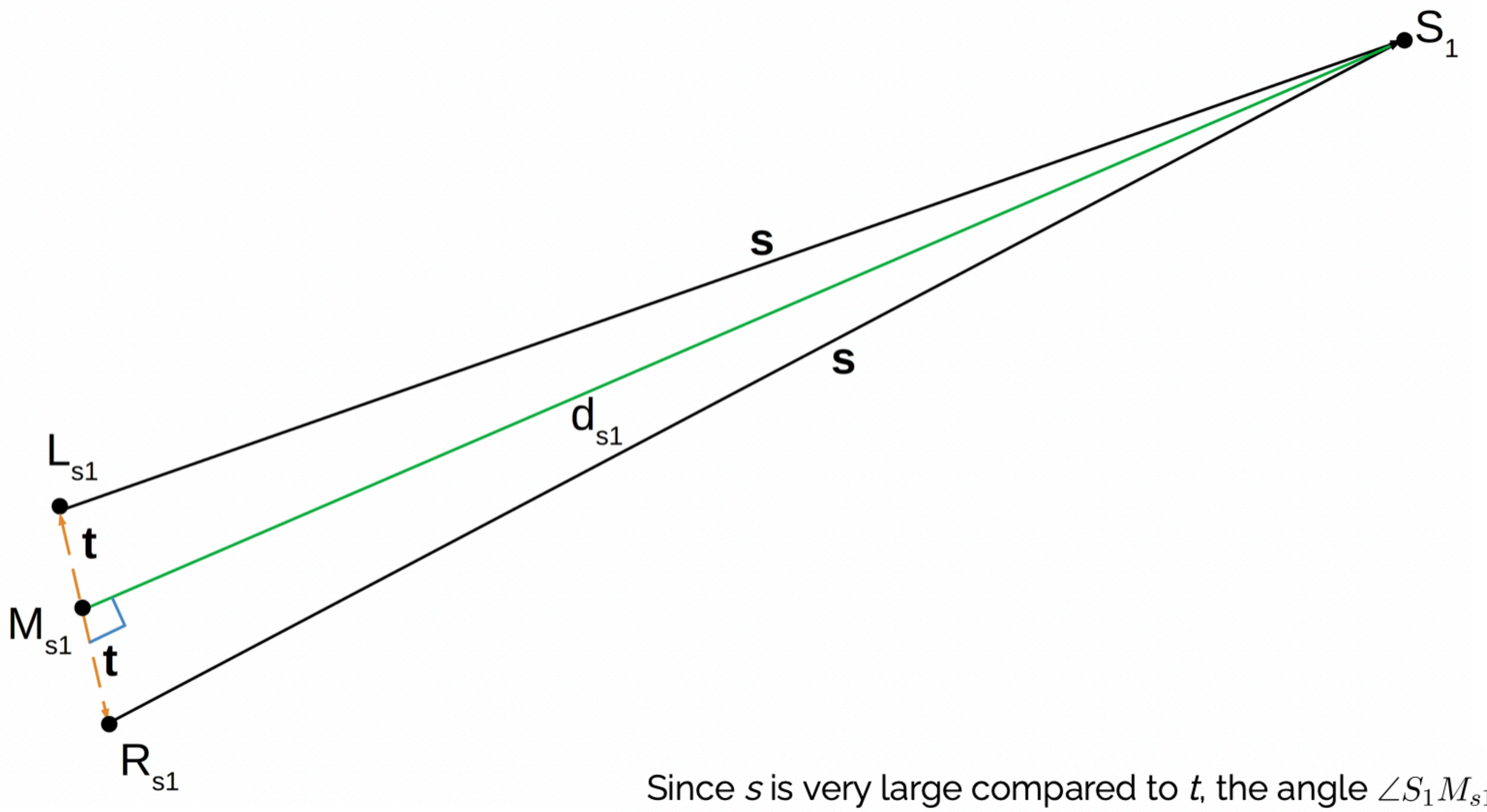
A 3D item is projected into a 2D (planar) projective space when captured (projected) in an image. The issue with this so-called "planar projection" is that it causes the loss of depth.

The disparity between the two stereo pictures is the apparent motion of things. If we close one eye and open it quickly while keeping the other closed, we'll observe that objects near us move quite a bit, whereas those farther away move barely at all. We refer to this phenomenon as "discrepancy."

https://www.baeldung.com/cs/stereo-vision-3d

Stereo Vision





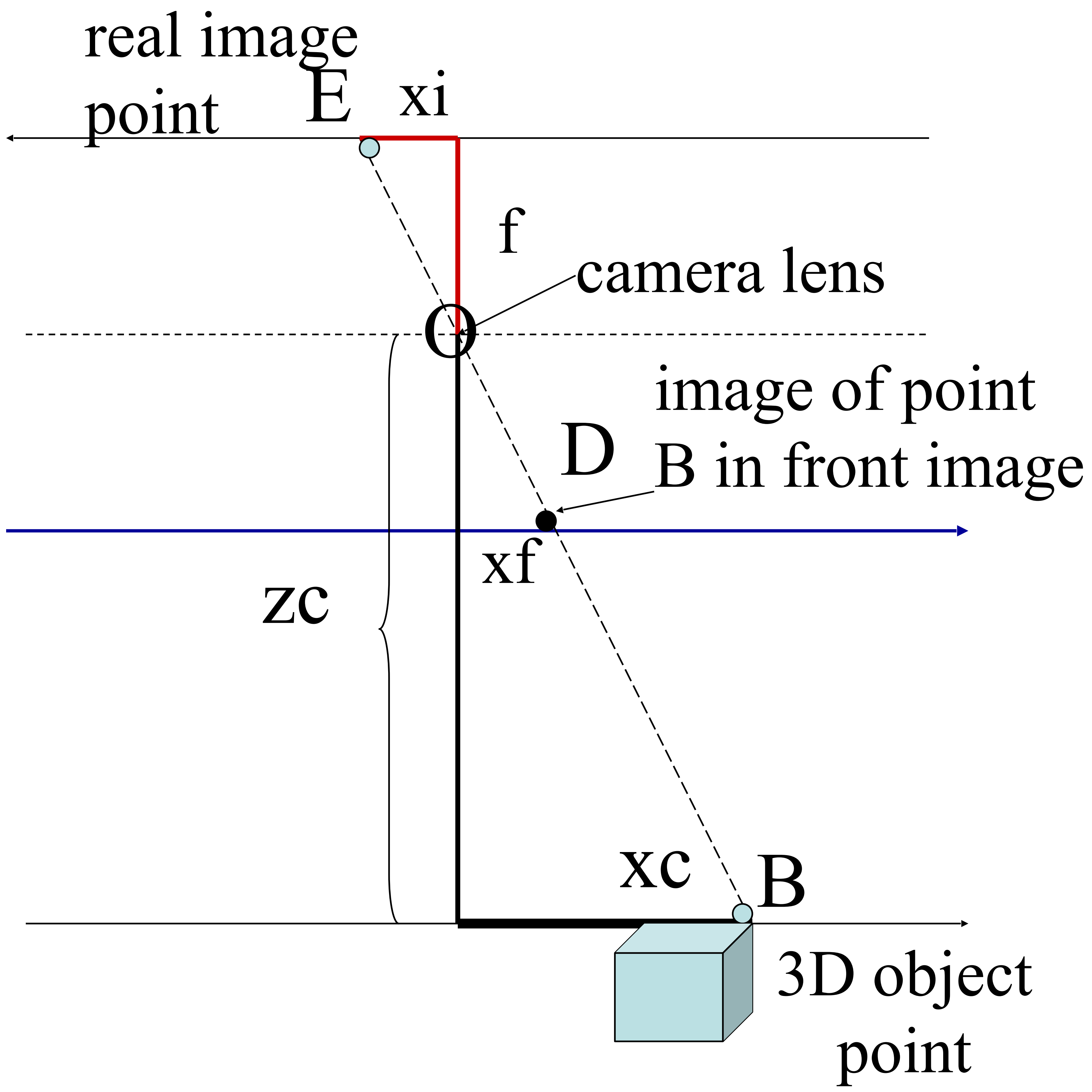
Since *s* is very large compared to *t*, the angle $\angle S_1 M_{s1} R_{s1}$ approaches $90°$. Lengths $L_{s1}M_{s1}$ and $M_{s1}R_{s1}$ are almost the same (denoted by *t*). Also, lengths $L_{s1}S_1$ and $R_{s1}S_1$ are almost the same (denoted by *s*). Applying the Pythagorean theorem, we get $s^2 = {d_{s1}}^2 + t^2$. Solving for the depth of point S_1 we get:

$$d_{s1} = \sqrt{s^2 - t^2}$$

Since *s* is very large compared to *t*, the depth d_{s1} is close to s.

https://www.baeldung.com/cs/stereo-vision-3d

Perspective Imaging Model: 1D



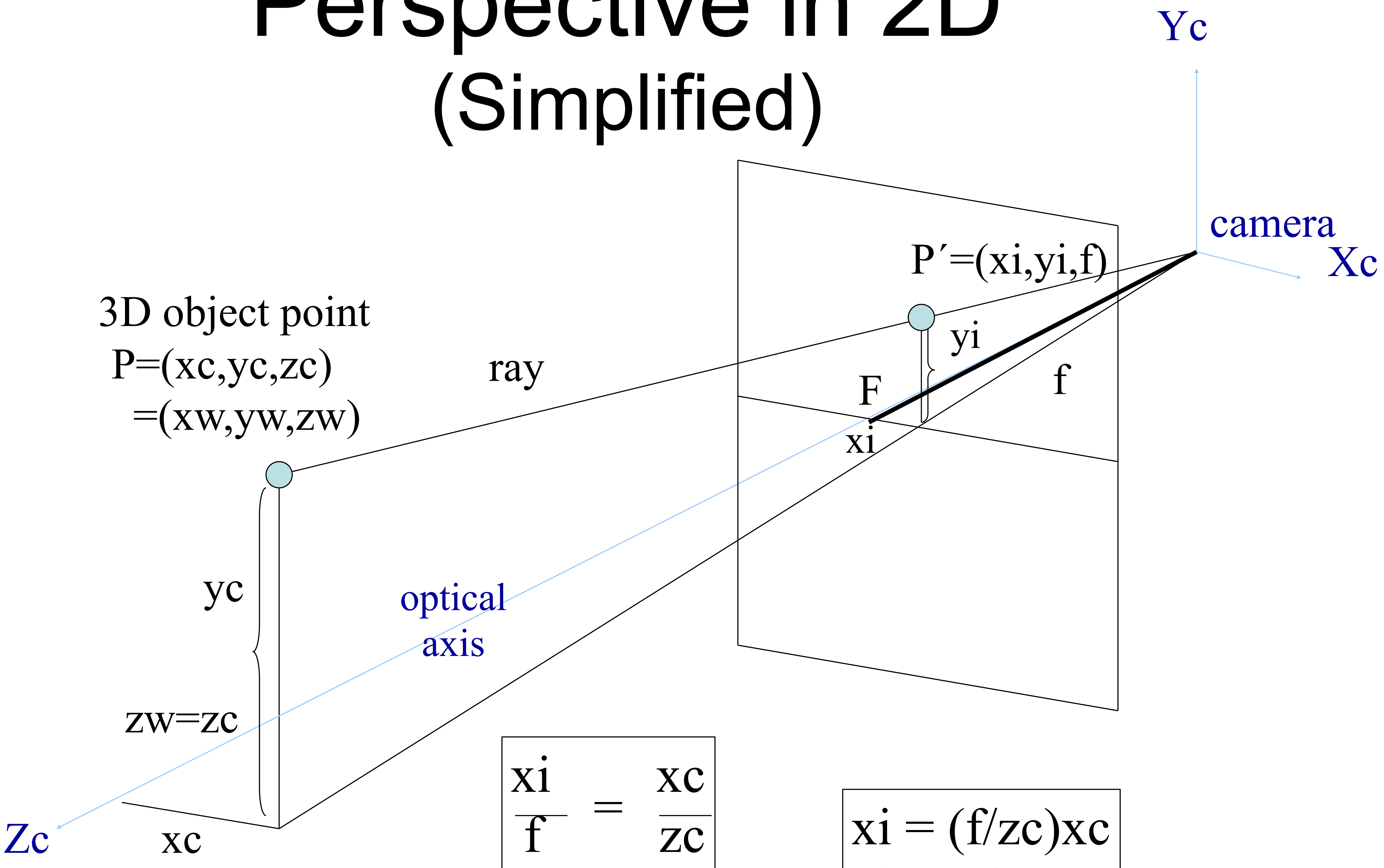
This is the axis of the real image plane.

O is the center of projection.

This is the axis of the front image plane, which we use.

$$\frac{xi}{f} = \frac{xc}{zc}$$

Perspective in 2D
(Simplified)



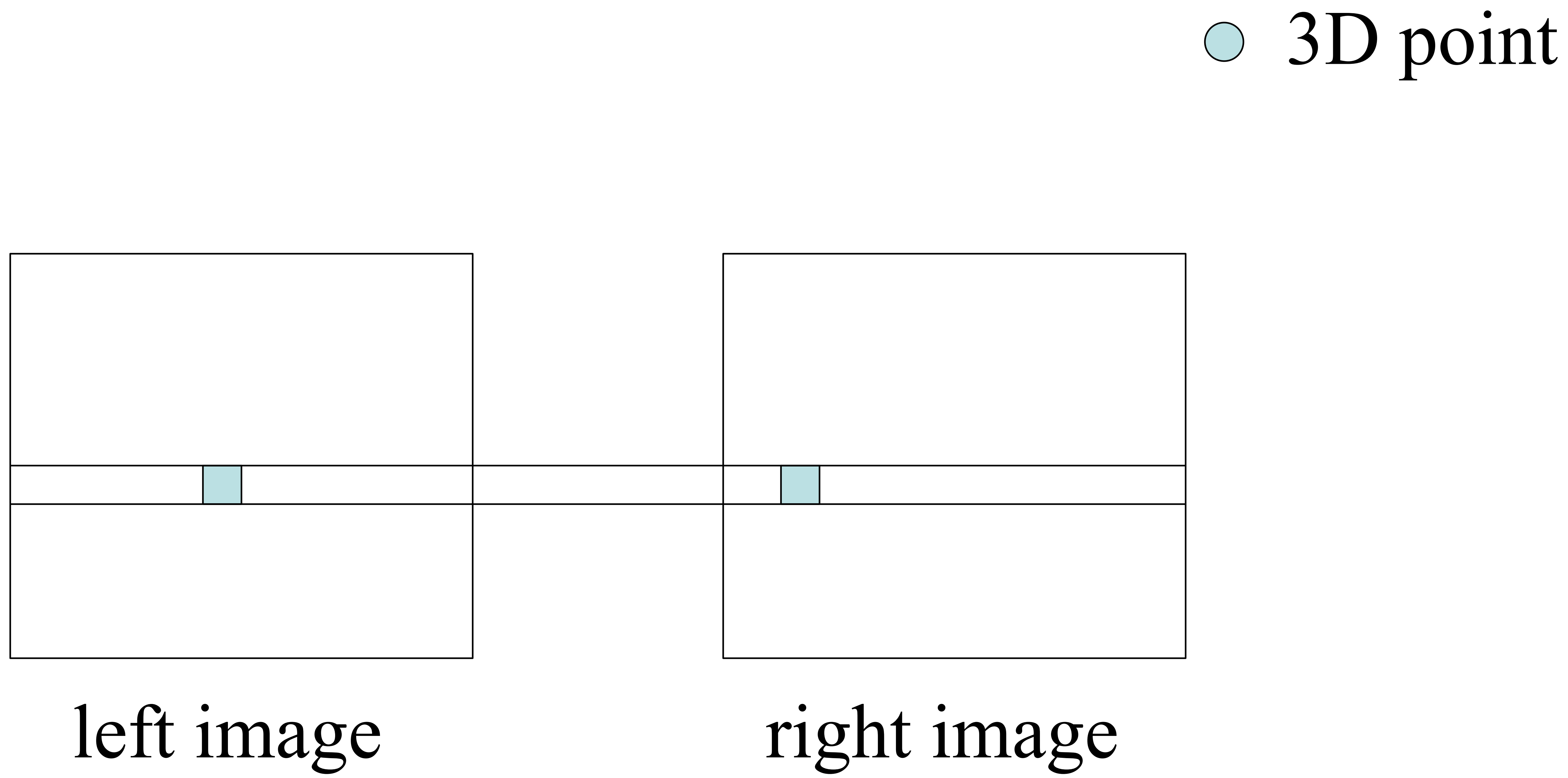
$$\frac{xi}{f} = \frac{xc}{zc}$$

$$\frac{yi}{f} = \frac{yc}{zc}$$

$$xi = (f/zc)xc$$
$$yi = (f/zc)yc$$

Here camera coordinates equal world coordinates.

3D from Stereo

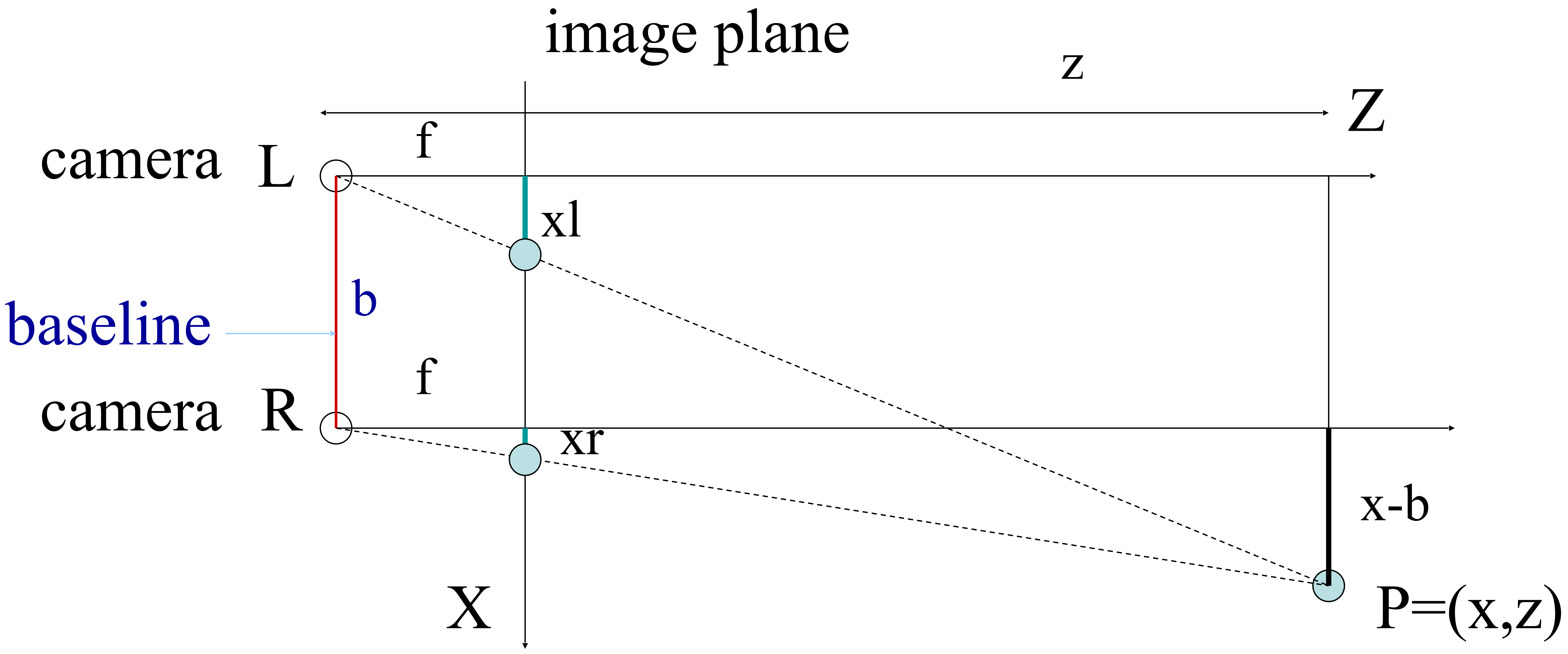


disparity: the difference in image location of the same 3D point when projected under perspective to two different cameras.

$$\boxed{d = xleft - xright}$$

Depth Perception from Stereo

Simple Model: Parallel Optic Axes



$$\frac{z}{f} = \frac{x}{xl}$$

$$\frac{z}{f} = \frac{x-b}{xr}$$

$$\frac{z}{f} = \frac{y}{yl} = \frac{y}{yr}$$

y-axis is perpendicular to the page.

Resultant Depth Calculation

For stereo cameras with parallel optical axes, focal length f, baseline b, corresponding image points (xl,yl) and (xr,yr) with disparity d:

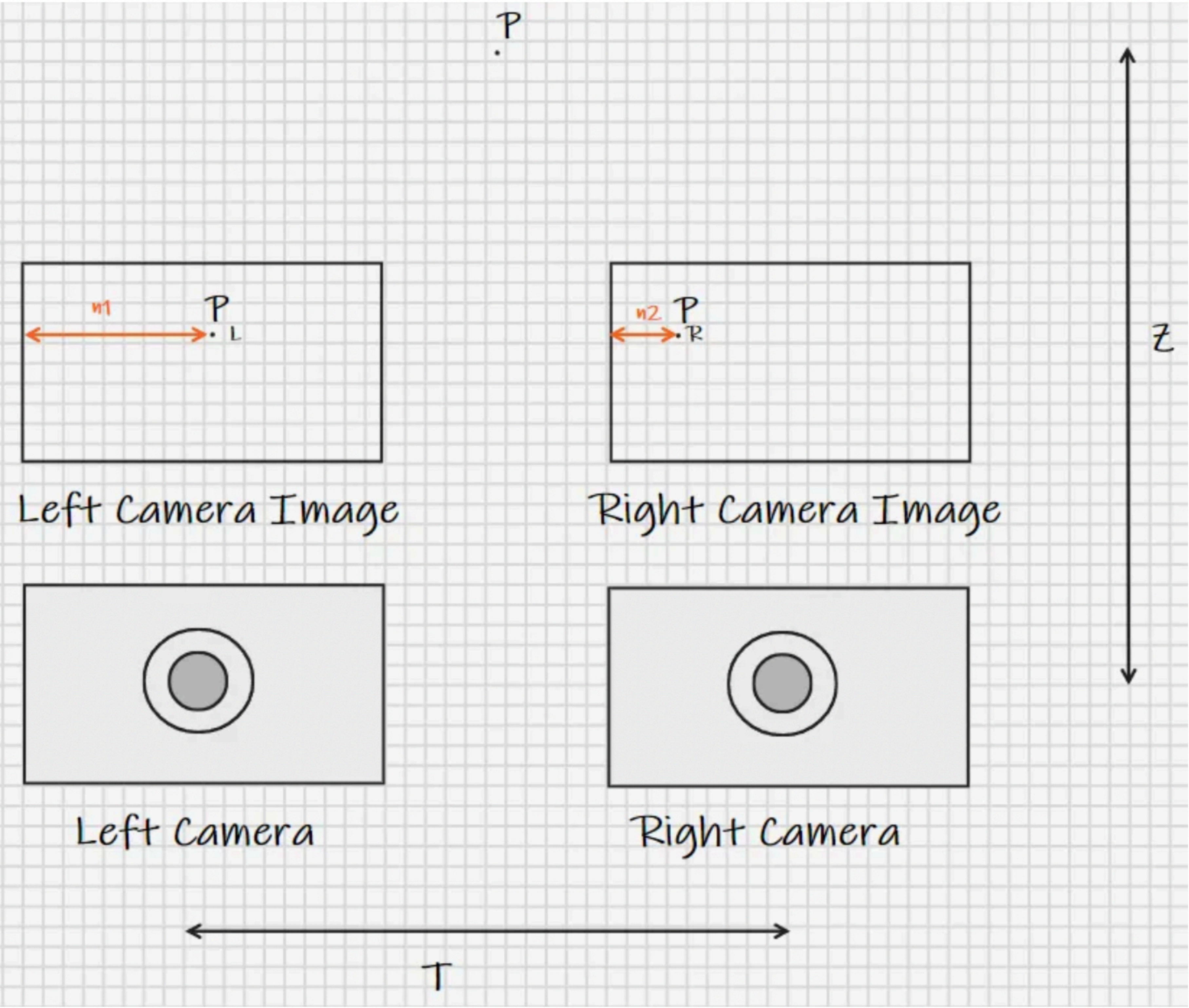
$$z = f*b / (xl - xr) = f*b/d$$

$$x = xl*z/f \quad \text{or} \quad b + xr*z/f$$

$$y = yl*z/f \quad \text{or} \quad yr*z/f$$

This method of determining depth from disparity is called **triangulation.**

Stereo Vision

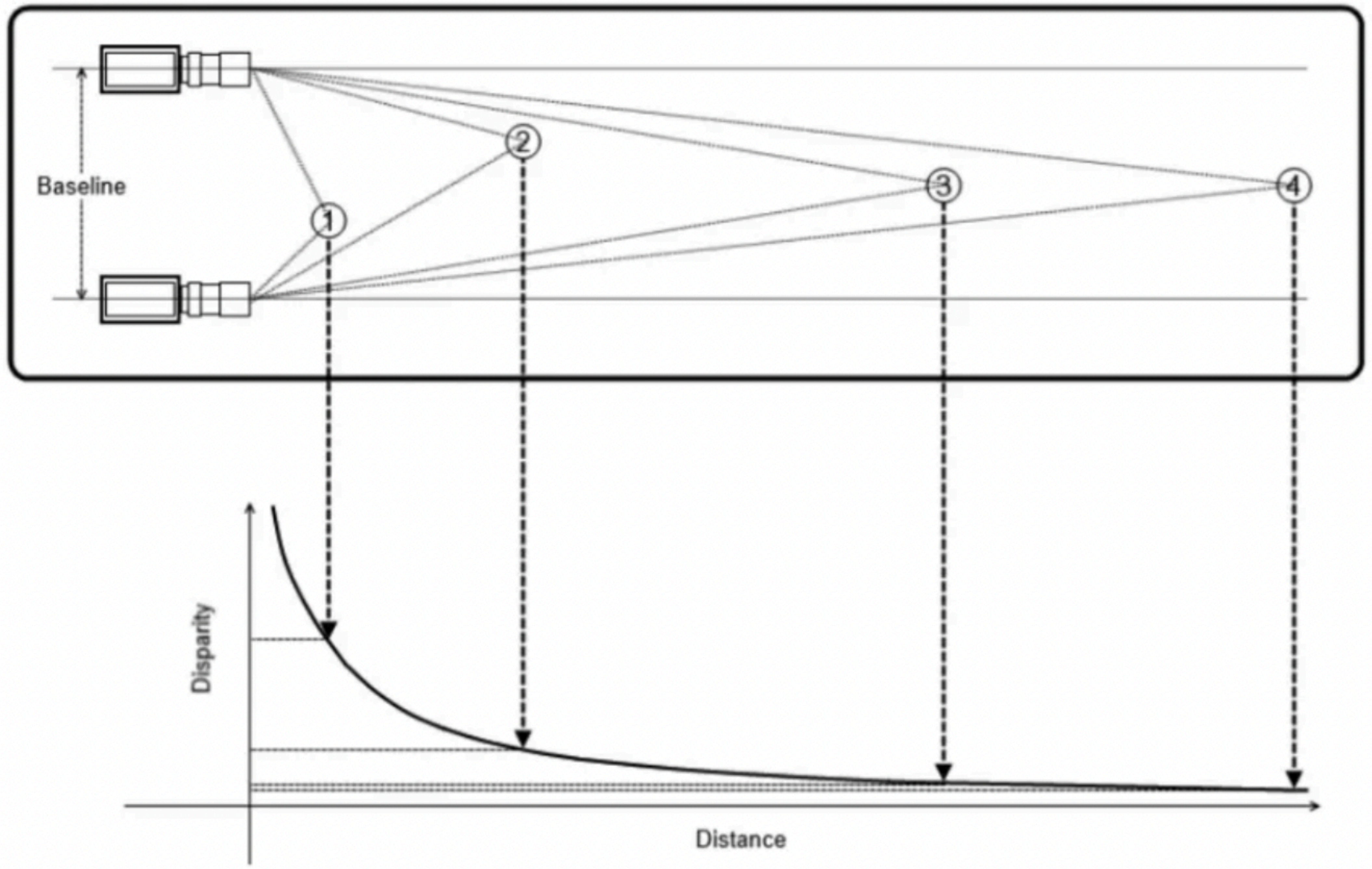


- P = Target point in the physical world (scene point)
- PL = (xl,yl) = Point P in left camera image
- PR = (xr,yr)= Point P in right camera image
- n1 = Horizontal pixel distance of point PL in left camera image
- n2 = Horizontal pixel distance of point PR in right camera image
- T = Baseline distance between center of left and right cameras
- f = Focal length of the camera
- d = Physical size of a pixel in camera sensor CMOS/CCD
- Z = Distance between point P and camera center

$$Z = \frac{f}{d} \times \frac{T}{D}$$

Stereo Vision

- **Depth is inversely proportional to disparity**. The more the disparity is, the closer the object is to the baseline of the camera. The less the disparity is, the farther the object is to the baseline.



Source: [2]

- **Disparity is proportional to baseline**. This is easy to visualize. If we have a small baseline distance between the two cameras, then the difference/disparity between the two images is going to be small. As we increase the baseline, the disparity is going to scale up.

https://medium.com/analytics-vidhya/distance-estimation-cf2f2fd709d8

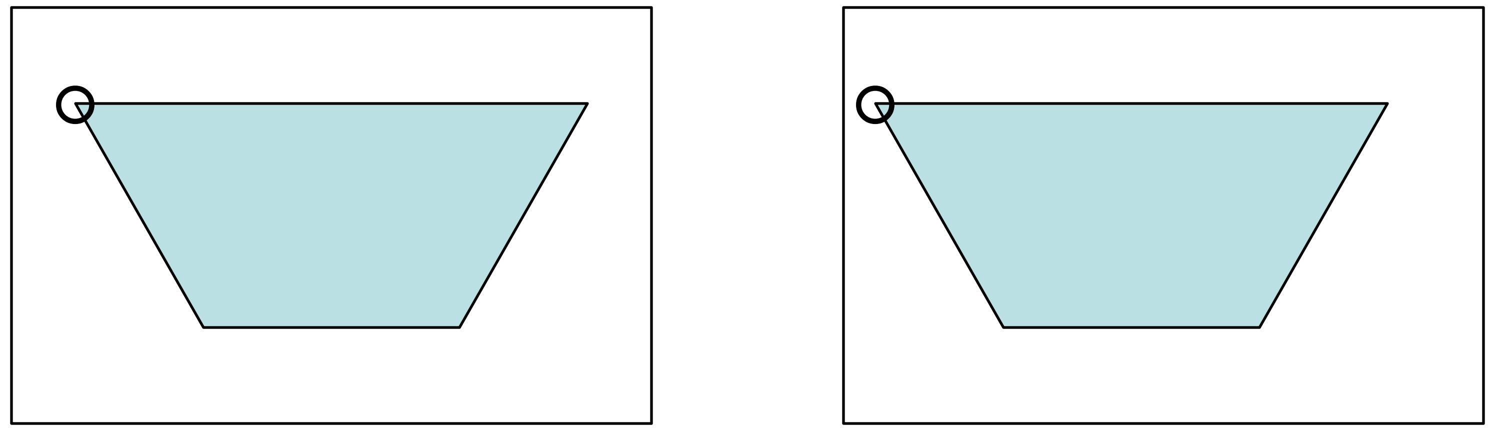
Stereo Vision

Finding Correspondences

- If the correspondence is correct, triangulation works **VERY** well.

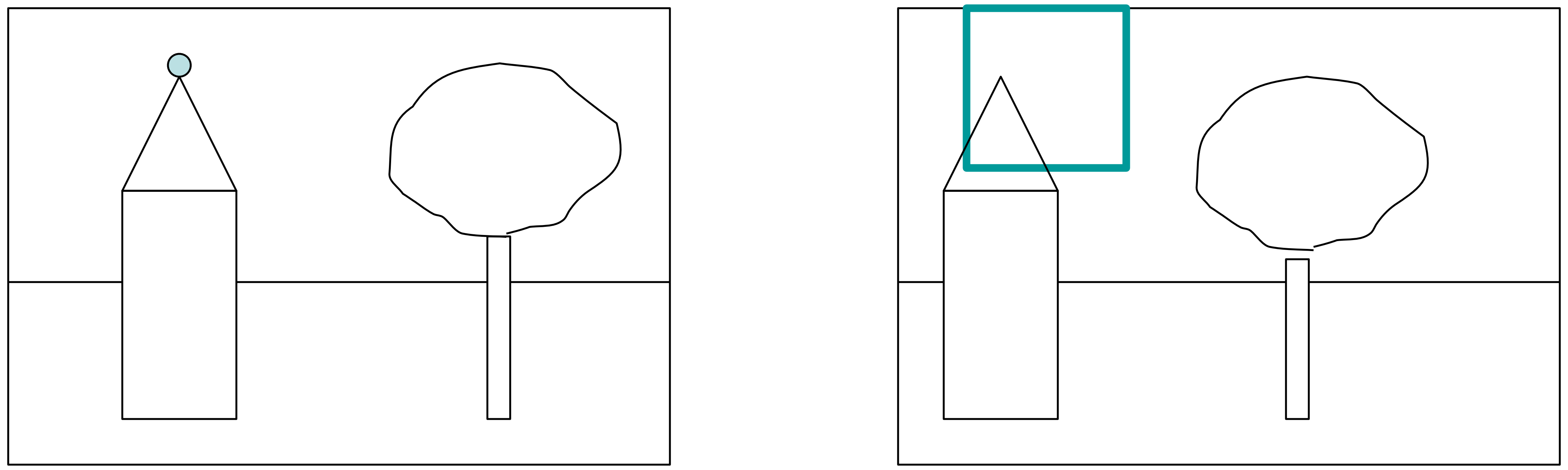
- But correspondence finding is not perfectly solved. for the general stereo problem.

- For some very specific applications, it can be solved for those specific kind of images, e.g. windshield of a car.



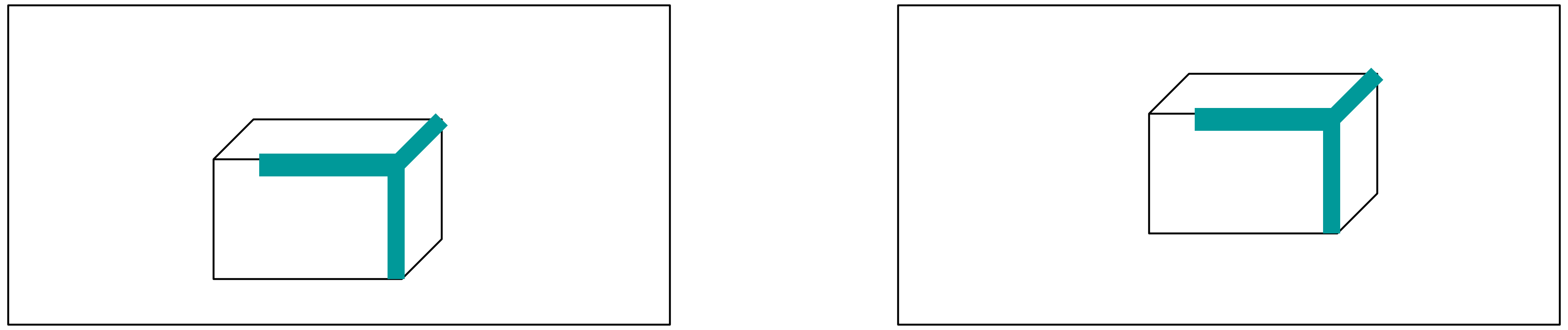
3 Main Matching Methods

1. Cross correlation using small windows.



dense

2. Symbolic feature matching, usually using segments/corners.



sparse

3. Use the newer interest operators, ie. SIFT.

sparse

The content is added from different sources

Stereo Vision

Illustration for sliding window along scanline for template matching

1. Sum of Squared Differences (SSD)

$$SSD(win_L, win_R) = \sum_x \sum_y \left(I_{win_L}(x,y) - I_{win_R}(x,y) \right)^2$$

Equation to calculate SSD

Window with Minimum SSD = Most Similar/Matching Window.

2. Sum of Absolute Differences (SAD)

$$SAD(win_L, win_R) = \sum_x \sum_y \left| I_{win_L}(x,y) - I_{win_R}(x,y) \right|$$

Equation to calculate SAD

https://medium.com/analytics-vidhya/distance-estimation-cf2f2fd709d8

Contents

AUGMENTED REALITY

Augmented reality (AR) is an interactive experience that combines the real world with computer-generated content. This content can be visual, auditory, haptic, somatosensory, and olfactory. AR experiences are typically accessed through a smartphone or tablet, but they can also be accessed through head-mounted displays (HMDs), smart glasses, and other wearable devices.

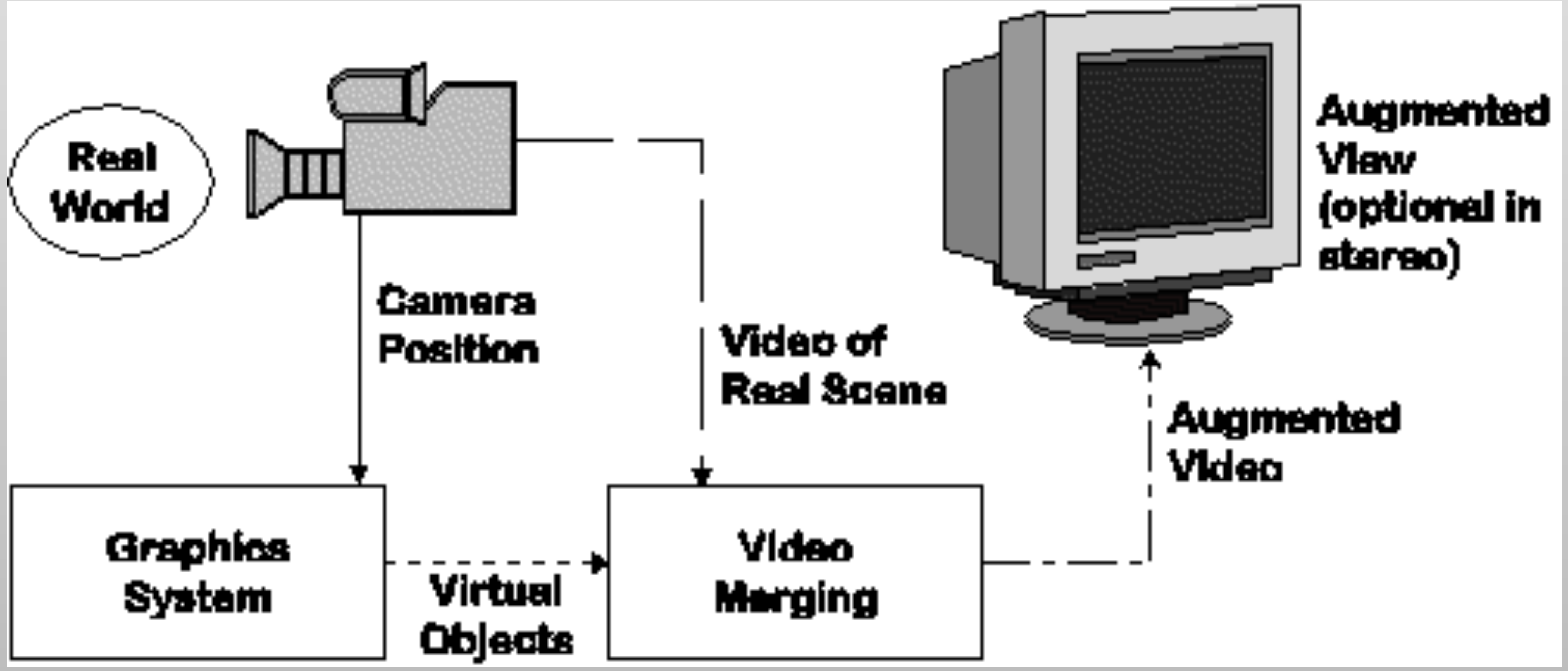
Augmented Reality

Artificial information about the environment and the objects in it can be stored and retrieved as an information layer on top of the real-world view. A combination of a real scene viewed by a user and a virtual scene generated by a computer that augments the scene with additional information.

iPhone 3GS Wikitude AR application uses GPS and compass technology to retrieve information about the user's surroundings

Augmented Reality

AR is different from virtual reality (VR) in that VR creates a completely artificial environment, while AR overlays digital content onto the real world. This means that AR can be used to enhance the real world, such as by providing information about objects or places, or by creating interactive games and experiences.



Applications of Augmented Reality

Education: AR can be used to create interactive learning experiences that are more engaging and memorable than traditional methods.

Retail: AR can be used to help customers visualize products in their own homes before they buy them.

Manufacturing: AR can be used to improve the efficiency and accuracy of manufacturing processes.

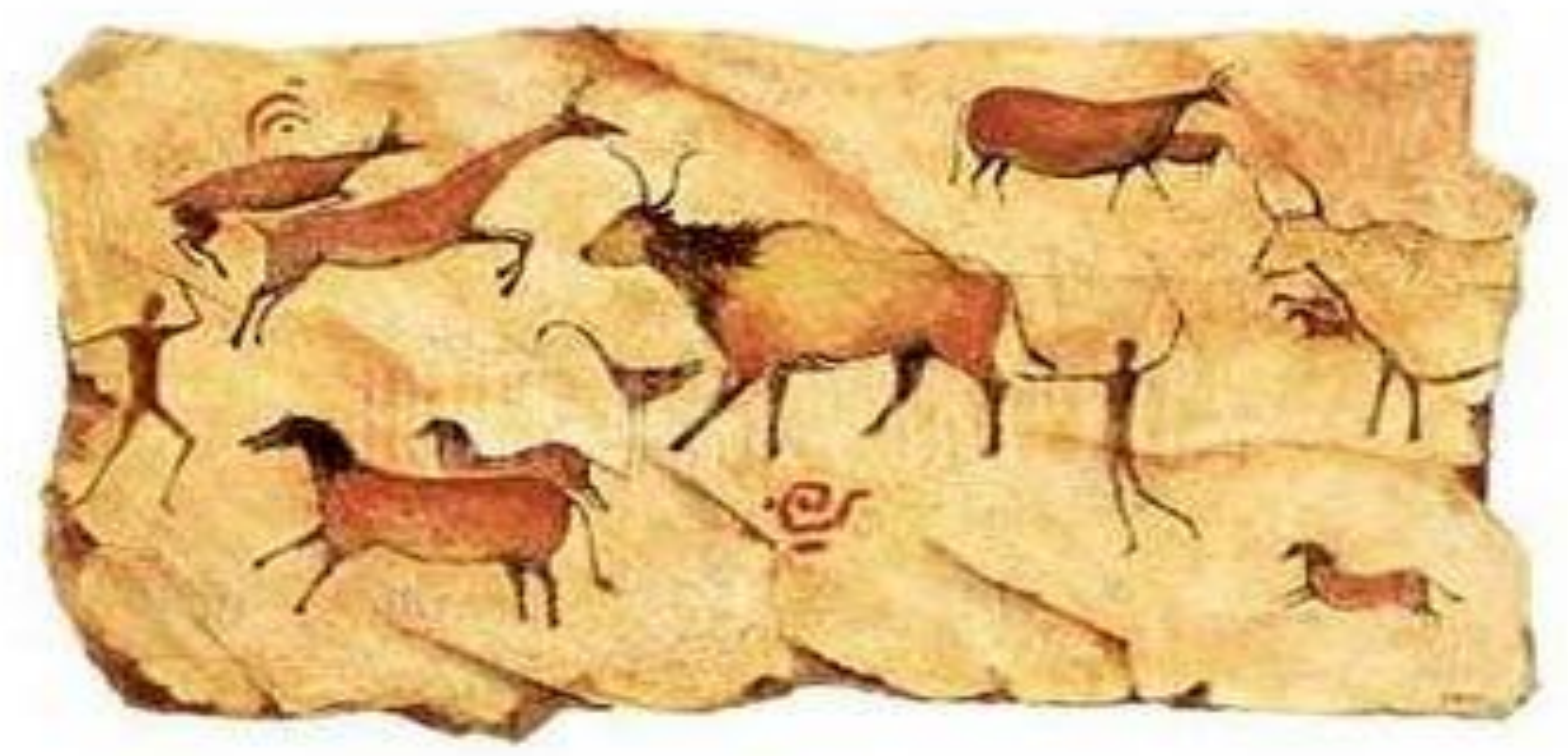
Healthcare: AR can be used to improve the quality of patient care.

Gaming: AR can be used to create immersive gaming experiences that are more engaging and realistic than traditional games.

VIRTUAL REALITY

Virtual reality (VR) is a simulated 3D environment that can be explored and interacted with by a person as if they were actually present. It achieves this by using a variety of technologies and devices.

First VR



Virtual Reality Devices

Headsets: These are worn on the head and project images directly onto the user's eyes, blocking out their view of the real world.

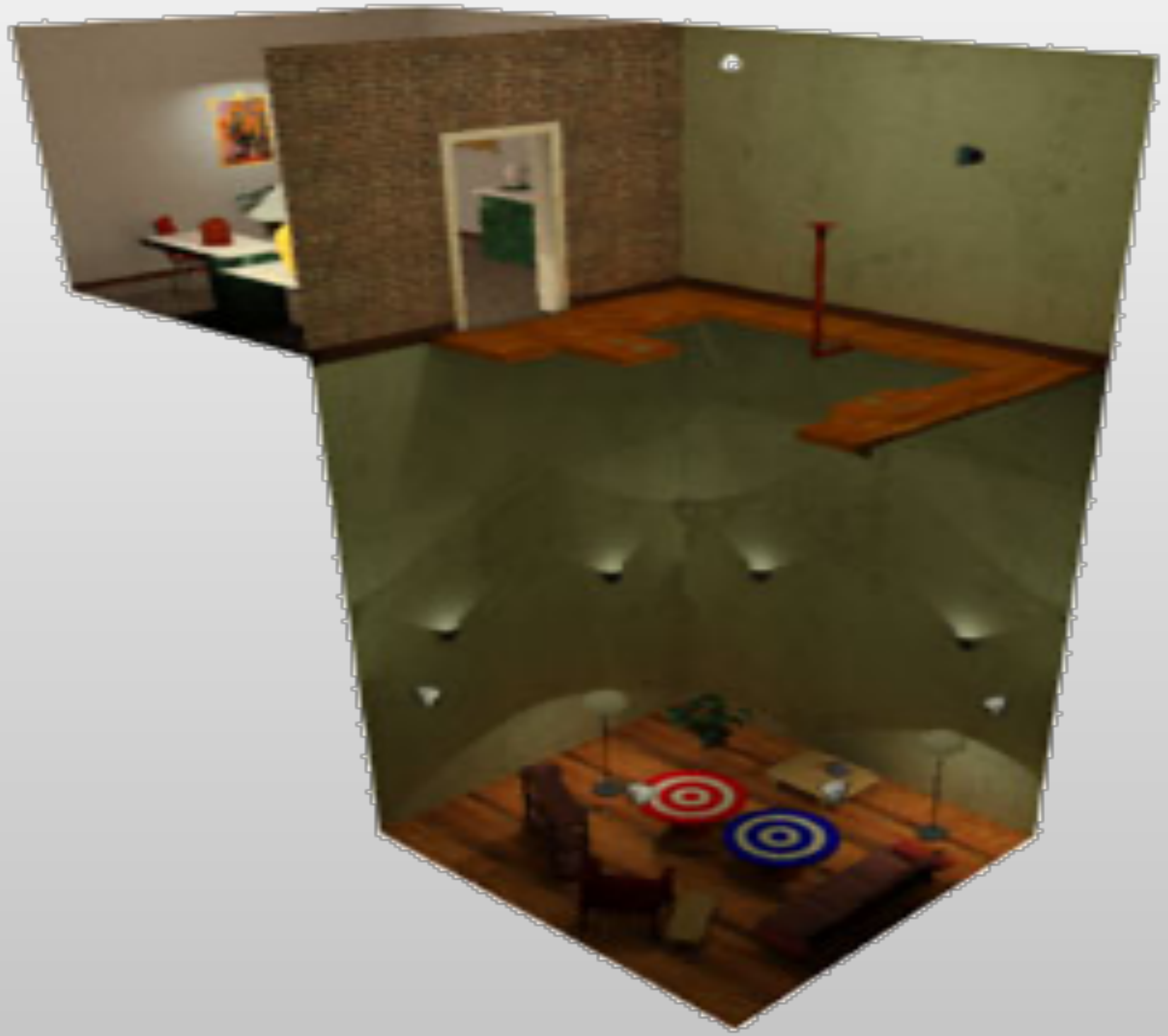
Sensors: These track the user's movements and translate them into movements within the virtual world.

Haptic feedback: This technology allows the user to feel sensations, such as textures, vibrations, and even temperature changes, within the virtual environment.

Audio: This is used to create realistic soundscapes that further immerse the user in the virtual world.

Key Elements of Virtual Reality Experience

- Virtual World - content of a given medium
  - screen play, script, etc.
  - actors performing the play allows us to experience the virtual world
- Immersion – sensation of being in an environment
  - mental immersion – suspension of disbelief
  - physical immersion – bodily entering the medium
  - Related to presence – (mentally immersed) the participant's sensation of being in the virtual environment (Slater)



Walking Experiment at UNC – Chapel Hill

Key Elements of Virtual Reality Experience

- ❖ Sensory Feedback – information about the virtual world is presented to the participant's senses
  - Visual (most common)
  - Audio
  - Touch
- ❖ Interactivity – the virtual world responds to the user's actions.
  - Computer makes this possible
  - Real-time

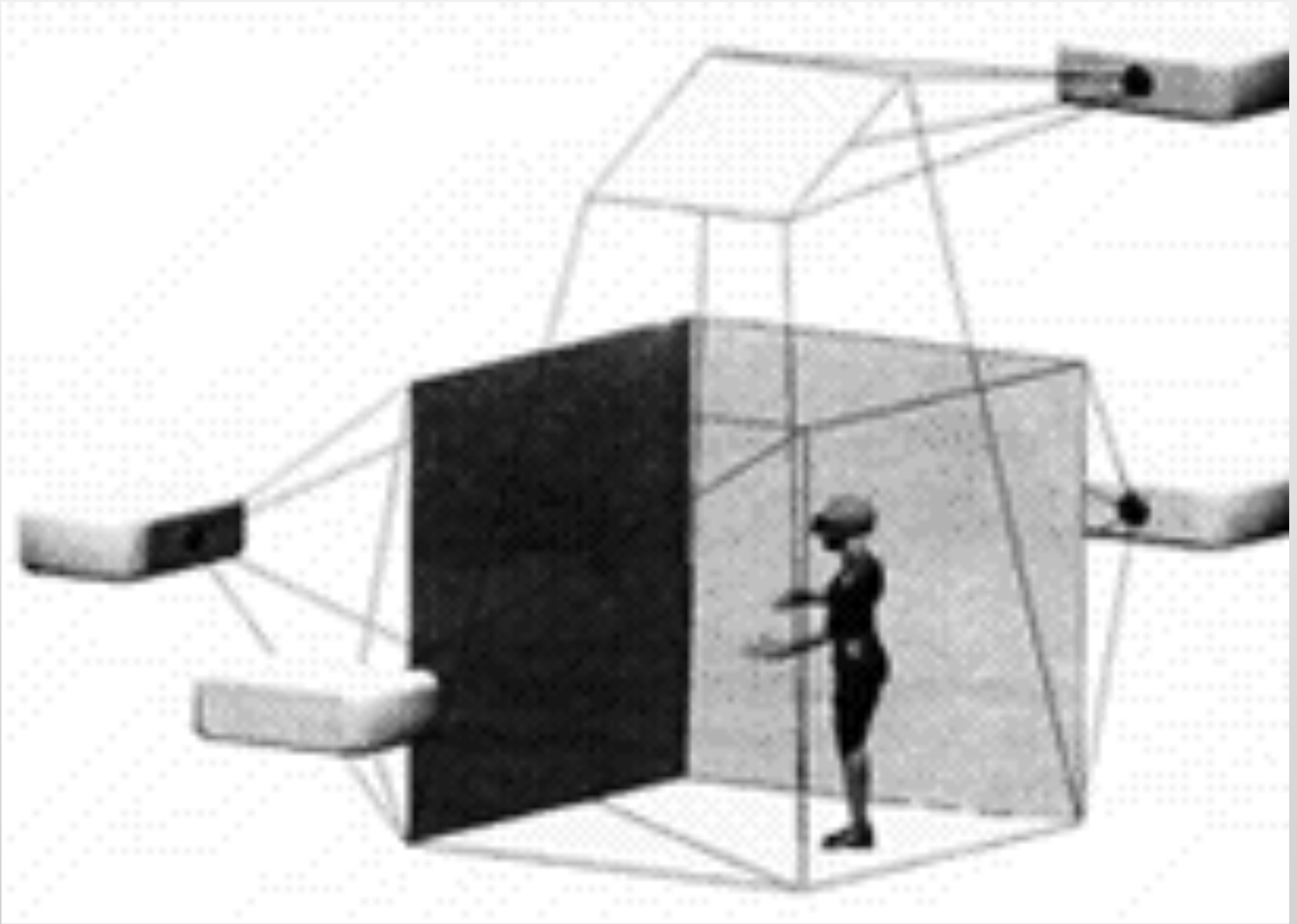
Walking Experiment at UNC – Chapel Hill

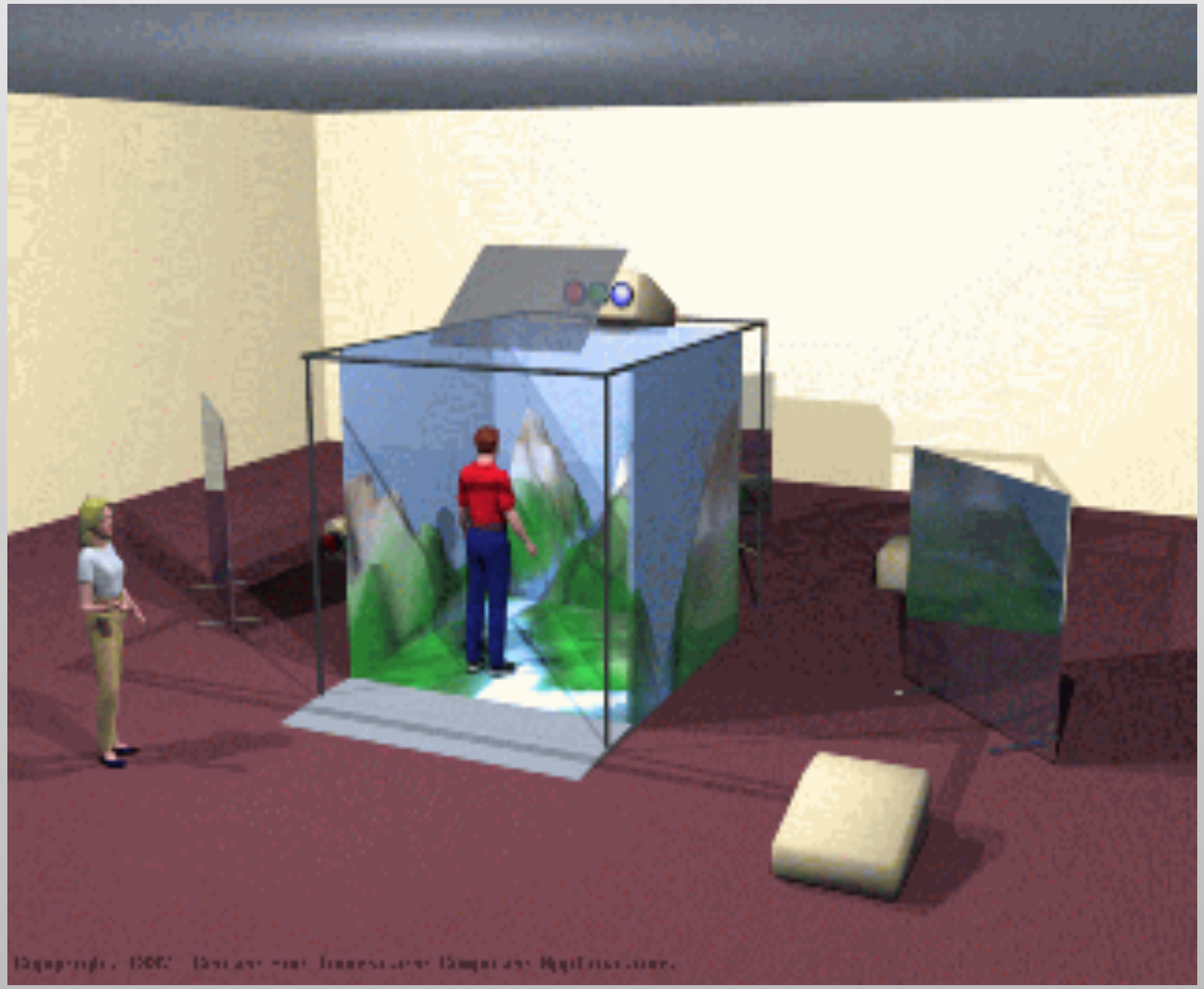
Immersive Technology

- ❖ Head-mounted Display
  - Optical System
  - Image Source (CRT or LCD)
  - Mounting Apparatus
  - Earphones
  - Position Tracker

Immersive Technology

Multi-screen Projection of stereoscopic images (CAVE)





Immersive Technology

❖Single large stereoscopic display
- Projection-based
- Head-tracked
- Possible tracking of hands and arms.
- Brings virtual objects into the physical world

Virtual Reality Applications

Entertainment: VR is used to create immersive and interactive gaming experiences, as well as virtual tours of real and imagined places.

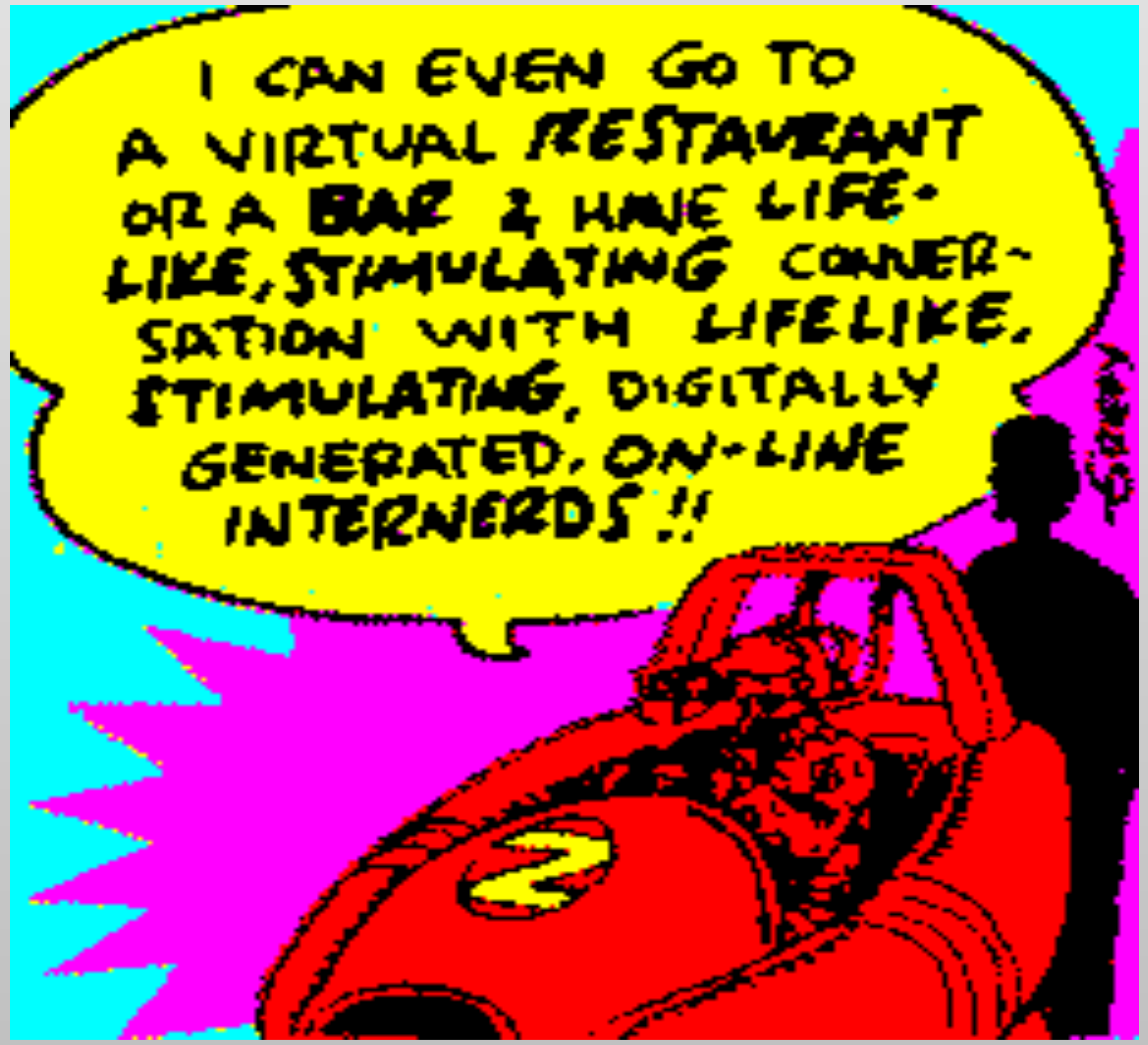
Education: VR can be used to create realistic simulations of real-world scenarios, such as historical events or scientific experiments. This can help students learn in a more engaging and memorable way.

Therapy: VR can be used to treat a variety of mental health conditions, such as phobias, anxiety, and post-traumatic stress disorder (PTSD).

Design and engineering: VR can be used to create 3D models of products and environments, which can then be reviewed and modified before they are built in the real world.

Augmented Reality vs Virtual Reality

Augmented reality (AR) and virtual reality (VR) are two emerging technologies that are changing the way we interact with the world around us. However, they are distinct technologies with different uses and applications.



VR Objects

AR Objects

Augmented Reality vs Virtual Reality

Feature	Augmented Reality	Virtual Reality
Definition	Overlays digital content onto the real world	Creates a completely immersive and virtual environment
Devices	Smartphones, tablets, smart glasses, head-mounted displays (HMDs)	Head-mounted display (HMD)
View of the real world	Maintains view of the real world	Replaces view of the real world
Interaction with the real world	Can interact with both real and virtual objects	Limited interaction with real world
Applications	Education, retail, manufacturing, healthcare, gaming	Entertainment, education, training, therapy, design and engineering

CV Algorithms Empower AR/VR Technologies

Computer vision algorithms play a significant role in empowering Augmented Reality (AR) and Virtual Reality (VR) technologies by enabling devices to perceive and interact with the surrounding environment.

Computer Vision contributes to enhancing AR and VR:

- ❖ Object Detection and Recognition
- ❖ Simultaneous Localization and Mapping (SLAM)
- ❖ 3D Reconstruction and Depth Perception
- ❖ Gaze Tracking
- ❖ Image and Scene Understanding

Object Detection and Recognition

AR leverages computer vision algorithms for recognizing and tracking objects in the real world. This allows for virtual content to be accurately overlaid onto those objects, creating a more believable and interactive experience. Ex. AR furniture apps use computer vision to detect and map the user's surroundings, allowing them to see how virtual furniture would look in their actual space.

In VR, object detection enables hand tracking and gesture recognition, allowing users to interact with the virtual environment using their hands. This removes the need for controllers and creates a more natural and intuitive interaction experience.

Simultaneous Localization and Mapping (SLAM)

This powerful computer vision technique helps AR/VR systems understand their position and orientation within the real world. SLAM algorithms continuously process camera data to create a 3D map of the environment, allowing AR/VR content to be realistically positioned and anchored in the real world.

In VR, SLAM enables room-scale experiences, allowing users to move freely within a virtual space without losing their bearings.

3D Reconstruction and Depth Perception

Computer vision algorithms can analyze multiple camera views to create detailed 3D models of the real world. This information is crucial for AR applications that need to accurately interact with the environment, such as games that require users to navigate virtual objects placed on real surfaces.

In VR, depth perception algorithms help create a more natural and realistic feeling of presence in the virtual world. By accurately calculating the distance to virtual objects, VR systems can create a sense of depth and scale that enhances the immersive experience.

Gaze Tracking

Computer vision algorithms can track the user's gaze direction by analyzing eye movements. This information can be used in both AR and VR applications to enhance user interaction and engagement.

In AR, gaze tracking can be used to control virtual objects or provide information about the environment based on where the user is looking.

In VR, gaze tracking can be used to select objects in the virtual environment, navigate through menus, and even control the direction of a virtual camera.

Image and Scene Understanding

Advanced computer vision techniques like scene understanding and semantic segmentation can analyze the content of images and videos to identify objects, landmarks, and other elements. This information can be used to personalize AR/VR experiences and make them more relevant to the user's context.

For example, AR travel apps can use scene understanding to identify landmarks and provide relevant information about them to the user.

Innovative uses of computer vision in AR/VR

Computer vision is a vital component of Augmented Reality (AR) and Virtual Reality (VR) technologies. By enabling machines to "see" and understand the world around them, computer vision unlocks a wide range of possibilities for creating immersive and interactive experiences.

- ❖ Enhanced Real-World Interaction in AR
- ❖ Personalized and Immersive VR Experiences
- ❖ Advanced Training and Simulation Applications
- ❖ Revolutionizing the Retail and Entertainment Industries

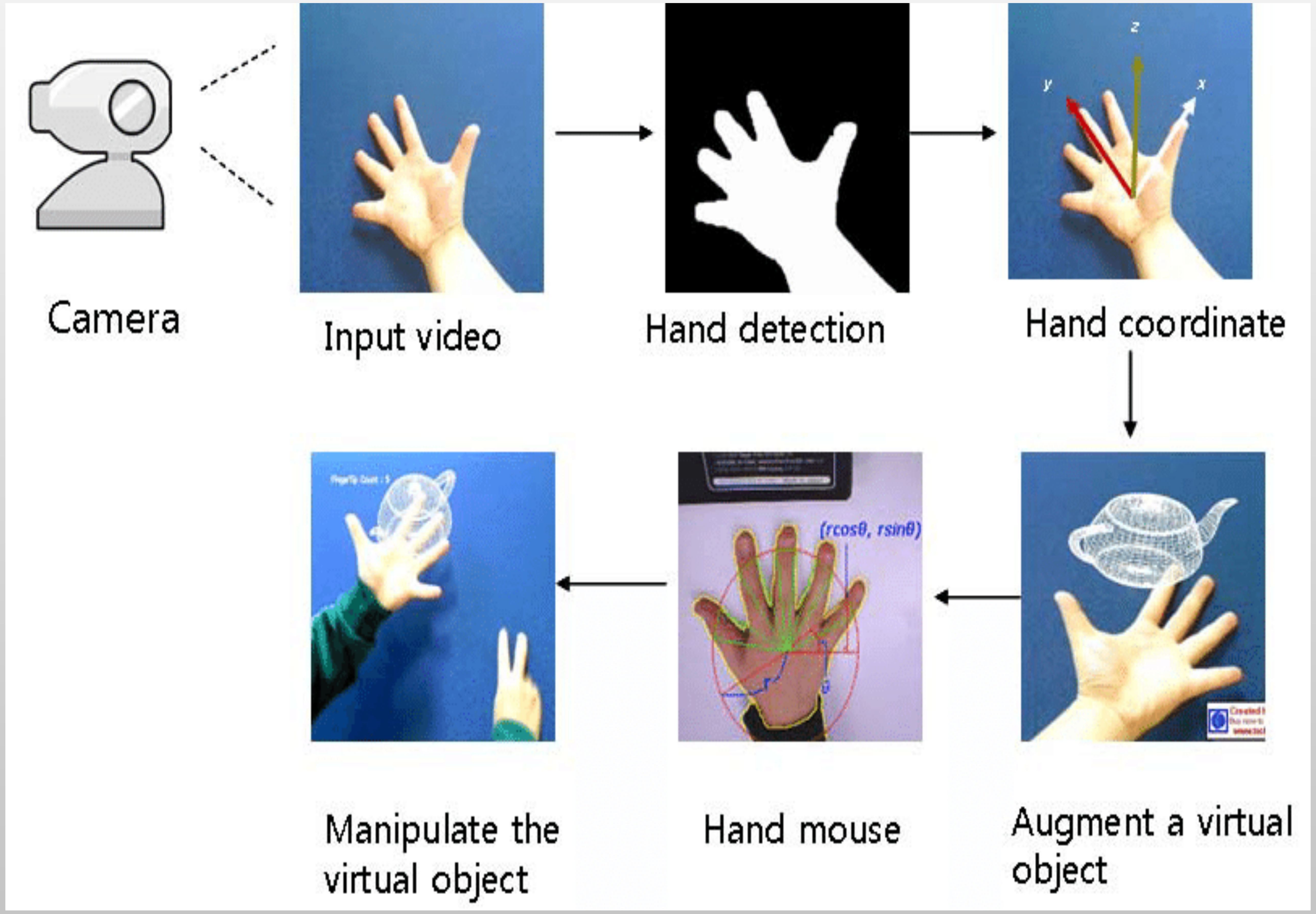
Enhanced Real-World Interaction in AR

Object Recognition & Manipulation: Imagine being able to scan a real object and have a virtual 3D model appear on your screen, allowing you to manipulate it and learn about it. This is possible with object recognition algorithms that can identify and track real-world objects in real-time.

Surface Recognition & Projection: Visualizing virtual content on any surface, from walls to furniture, can create interactive and informative experiences. Computer vision algorithms can analyze the surface and project content accordingly, adapting to its shape and texture.

Enhanced Real-World Interaction in AR

Real-time Occlusion: Seamlessly blending virtual and real elements is crucial for creating a believable AR experience. Computer vision algorithms can accurately detect and occlude virtual objects behind real ones, ensuring realism and depth.



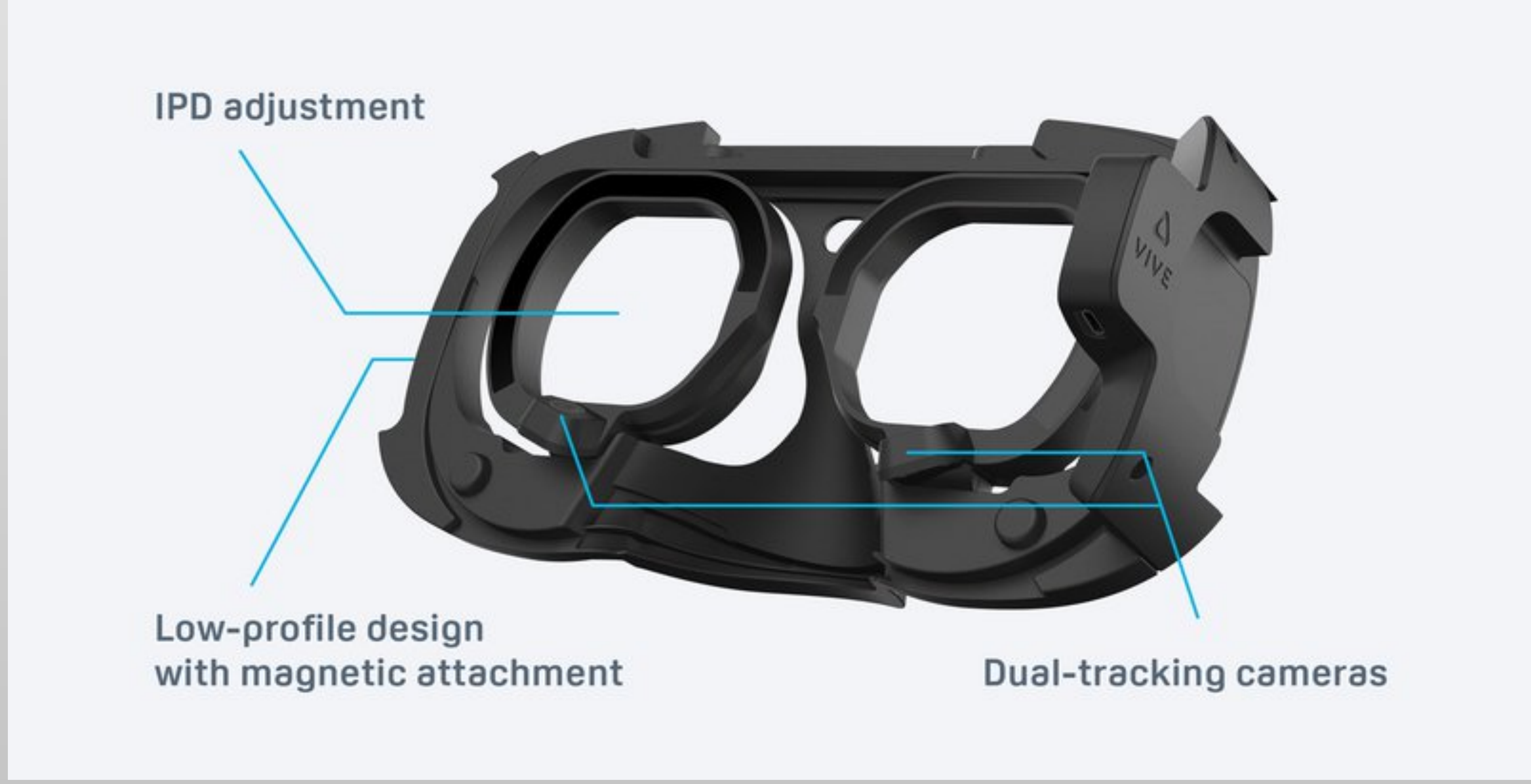
Personalized and Immersive VR Experiences

Gaze Tracking & User Interface Control: By tracking eye movements, computer vision allows for natural and intuitive interaction in VR environments. Users can select objects, navigate menus, and control the virtual environment simply by looking at what they want to interact with.

Facial Expression Recognition: Analyzing facial expressions in VR can personalize the experience and adapt to the user's emotional state. This could lead to more engaging and interactive VR applications, such as educational simulations or virtual therapy sessions.

Personalized and Immersive VR Experiences

Gesture Recognition & Natural Interaction: Computer vision enables users to interact with VR environments using their hands and gestures, eliminating the need for controllers and creating a more natural and intuitive interaction experience.



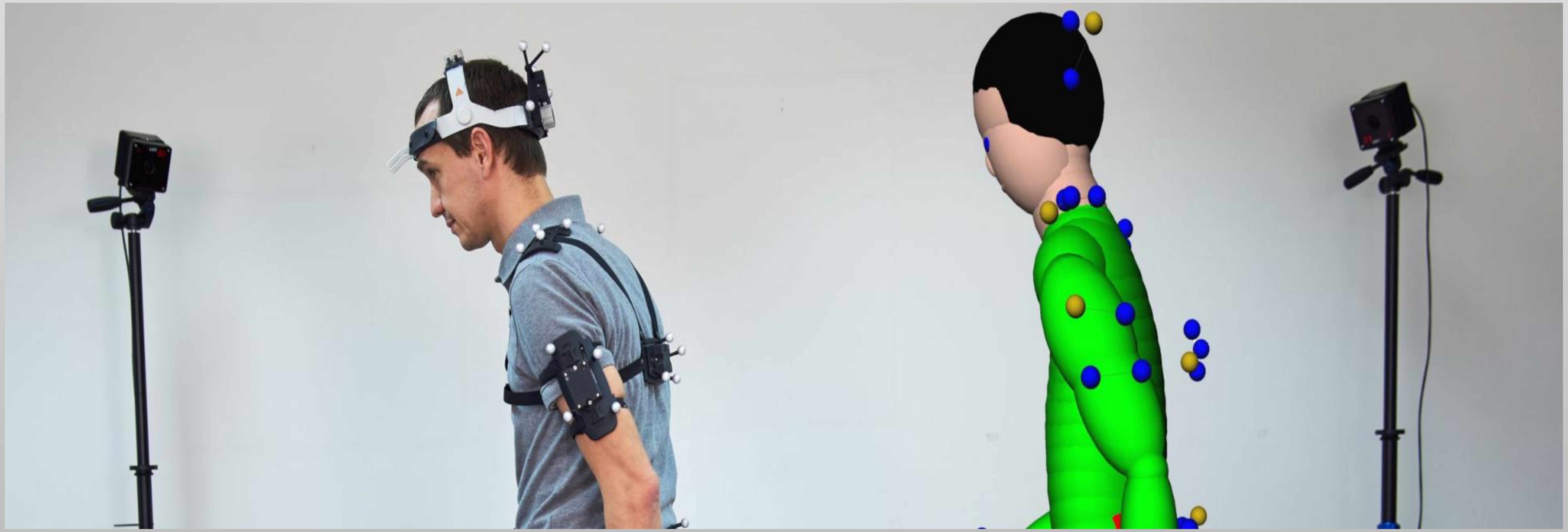
Advanced Training and Simulation Applications

Real-time Motion Capture & Feedback: Accurately tracking user movements in real-time allows for realistic simulations and training applications. This can be used for training surgeons, athletes, pilots, and other professionals in a safe and controlled environment.

Environment Mapping & Reconstruction: Creating accurate 3D models of real-world environments allows for immersive VR training simulations. Computer vision algorithms can analyze images and videos to reconstruct the environment with high fidelity, allowing users to interact with virtual representations of real-world locations.

Advanced Training and Simulation Applications

Interactive Learning and Educational Experiences: Combining computer vision with AR/VR can create interactive and engaging learning experiences. Imagine exploring historical sites in AR or dissecting virtual models in VR, bringing learning to life in a way that traditional methods cannot.



Revolutionizing the Retail and Entertainment Industries

Virtual Try-On and Customization: Imagine trying on clothes virtually before you buy them, or customizing products in real-time using AR. Computer vision algorithms can analyze your body and surroundings to accurately overlay virtual objects and provide a realistic experience.

Interactive Product Demonstrations: AR-powered product demonstrations can provide customers with a richer and more engaging experience. Imagine seeing a product in action in your own home or learning how to use it with step-by-step instructions overlaid in AR.

Revolutionizing the Retail and Entertainment Industries

Immersive Entertainment Experiences: AR/VR experiences powered by computer vision can take entertainment to a whole new level. Imagine exploring virtual worlds, interacting with characters, and participating in interactive games, all with seamless integration of the real and virtual worlds.

Ethical considerations in the use of computer vision in AR/VR

While computer vision unlocks numerous possibilities for AR/VR, its use also raises important ethical concerns that need careful consideration.

- Privacy Concerns
- Bias and Discrimination
- Mental Health and Well-being
- Safety and Security
- Transparency and Accountability

Privacy Concerns

Data Collection: Computer vision algorithms rely on large amounts of data for training and operation. This raises concerns about user privacy, especially when collecting sensitive data like facial expressions or body movements.

Surveillance: AR/VR applications with computer vision capabilities could potentially be used for surveillance without consent. This could lead to privacy violations and a loss of individual autonomy.

Data Ownership and Control: Who owns and controls the data collected by AR/VR systems? Ensuring transparency and providing users with control over their data is crucial to protecting privacy.

Bias and Discrimination

Algorithmic Bias: Computer vision algorithms can be biased based on the data they are trained on. This can lead to discriminatory outcomes, such as biased AR filters or VR experiences that favour certain groups of people.

Stereotyping and Cultural Insensitivity: AR/VR applications can perpetuate harmful stereotypes and cultural biases if not designed carefully. It's crucial to consider the potential impact on diverse communities and ensure inclusivity in design and development.

Accessibility and Inclusiveness: Not everyone has access to AR/VR technologies, which can exacerbate existing societal inequalities. Ensuring accessibility and inclusivity is crucial to avoid further marginalizing vulnerable groups.

Mental Health and Well-being

Addiction and Overuse: Immersive AR/VR experiences can be highly addictive, leading to excessive use and neglecting real-world responsibilities.

Social Isolation and Dissociation: Spending too much time in virtual worlds can lead to social isolation and difficulty reintegrating into the real world.

Psychological Effects: AR/VR experiences that are violent or disturbing can have negative psychological impacts on users, especially children and vulnerable individuals.

Safety and Security

Physical Safety: AR/VR applications can distract users from their surroundings, leading to accidents and injuries. Ensuring safety features and user awareness is essential.

Cybersecurity: AR/VR systems can be vulnerable to hacking and cyberattacks, compromising user data and privacy. Robust cybersecurity measures are crucial for protecting users.

Misinformation and Manipulation: AR/VR can be used to spread misinformation and manipulate users through immersive experiences. It's important to develop responsible content creation guidelines and fact-checking mechanisms.

Transparency and Accountability

Algorithmic Transparency: Users have a right to understand how computer vision algorithms work and how their data is being used. Transparency in algorithmic design and decision-making is crucial for building user trust.

Accountability for Ethical Issues: Developers and manufacturers of AR/VR technologies must be held accountable for ensuring ethical practices and addressing potential harms.

Public Discourse and Regulatory Frameworks: Ethical considerations in AR/VR require open public discourse and robust regulatory frameworks to ensure responsible development and deployment of these technologies.

Brainstorming session on innovative uses of CV in AR/VR

- Sharing Experience of AR/VR – Story Telling
- Construct a Tree of words that comes to your mind – Using Paint/Word
- Discussion of Advantages and Disadvantages