Various types of Motion Tracking, a comparison

Kinect & depth camera's	Vive Ultimate	AI-based Motion Capture

Depth-sensing / markerless camera-based mocap

How it works:

- RGB camera, infrared depth sensor
- Tracks body skeletons in 3D space without any wearables.

Strengths:

- **A**ll-in-one: depth + skeleton tracking
- Works out-of-the-box with good body tracking
- Widely used in interactive installations and prototyping

Limitations:

- Limited range and lighting conditions
- Skeleton tracking is less robust than pro systems
- Requires a (windows) PC and specific SDKs

In art, Kinect is great for:

- Interactive performances
- Visuals that respond to body movement
- Multi-user installations

See more info on 3d Depth camera's here

Inside-out inertial tracking with onboard cameras and IMUs (think of it as a hybrid between inertial and AI/vision-based tracking)

How it works:

- Unlike earlier Vive Trackers that rely on external Lighthouse base stations, the Ultimate Trackers use two onboard cameras and IMUs to track their position in space independently.
- They perform inside-out tracking, meaning they see the environment rather than relying on it.
- Designed to work with Vive XR systems, but are also being adopted for standalone tracking in XR, motion capture, and performance.

Strengths:

- No need for external base stations (fully wireless)
- Much more portable and scalable
- Accurate enough for many art/performance uses
- Easier multi-tracker setups

Limitations:

- Still relatively new — fewer integrations than legacy trackers
- Limited support in open-source or non-Vive environments (for now)
- Needs line of sight and light for the onboard cameras to function optimally

In art, Vive Ultimate is great for:

- Untethered performer tracking
- Object tracking in environments where base stations are impractical
- Mobile or temporary installations where quick setup is needed

How it works:

- Uses a single camera (or a small number of cameras) and AI algorithms to detect and track body, face, and hand movement.
- Examples include:
  - MediaPipe (Google): Real-time pose estimation in 2D or 3D
  - OpenPose : Widely used for body landmark detection
  - Move.ai: Advanced multi-camera AI mocap, often used with smartphones
  - DepthAI / OAK-D/ Zedi: Cameras with built-in AI processors that provide depth and pose data

Pros:

- No suits or markers needed — just a (web)camera
- Low cost, often free or open-source
- Quick to set up, highly accessible for artists and educators
- Can be embedded into web or mobile apps
- Good for gesture-based interaction, web-based artworks, or low-budget capture

Cons:

- Generally less accurate than optical or inertial systems
- Often limited to 2D or rough 3D estimation
- Struggles with occlusion, fast movement, or unusual poses
- Limited support for fine detail (like fingers or subtle facial expressions)

Revision #2
Created 2025-06-10 11:52:32 UTC by Astrid
Updated 2025-06-10 12:27:01 UTC by Astrid