Computer Vision for Augmented Reality

Computer Vision:

Computer Vision (CV) refers to the use of machine learning to process, analyze and understand digital images or videos in the real world to derive meaningful decisions and take action based on that. In simple words, computer vision allows machines to recognize and understand visuals just like humans. It is used in medical visualization, maintenance, and repair, etc in recent years. One of the complex examples of computer vision is in self-driving cars.

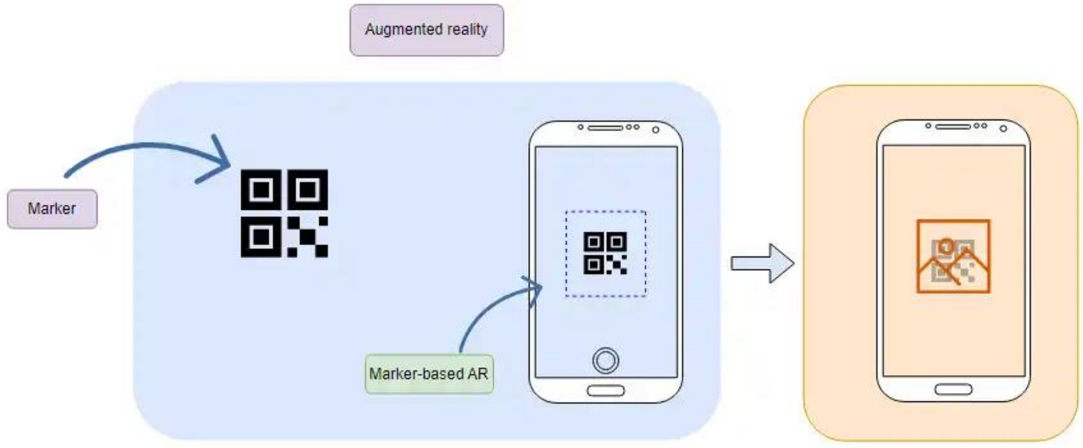
We, human beings gather almost all information through visual perception. Our intelligence enables us to identify and make inferences from any observation. Even though computers can't identify objects the same way we do, with careful development and training and with the help of computer vision, machines can be taught to identify objects and can be used to address individual challenges in almost every industry vertical.

Computer Vision in Augmented Reality

Augmented reality technology enables computing devices such as smartphones, tablets, and smart glasses to overlay objects on the real world. Computer vision plays an important role in augmented reality. Though the real potential of computer vision can only be leveraged with Artificial Intelligence (AI) and machine learning, it helps to build more accurate augmented reality environments and enables detection and identification of objects and locating real-world elements. In other words, it helps in object detection and tracking. It also supports augmented reality with simultaneous localization and mapping (SLAM).

Introduction-Marker Tracking:

Maker-based AR works by scanning a marker which triggers an augmented experience (whether an object, text, video or animation) to appear on the device. It usually requires software in the form of an app, which enables users to scan markers from their device using its camera feed.



How does it work?

Marker-based AR functions through the combined efforts of cameras, computer vision algorithms, and the marker itself. This process involves a few fundamental steps, which are:

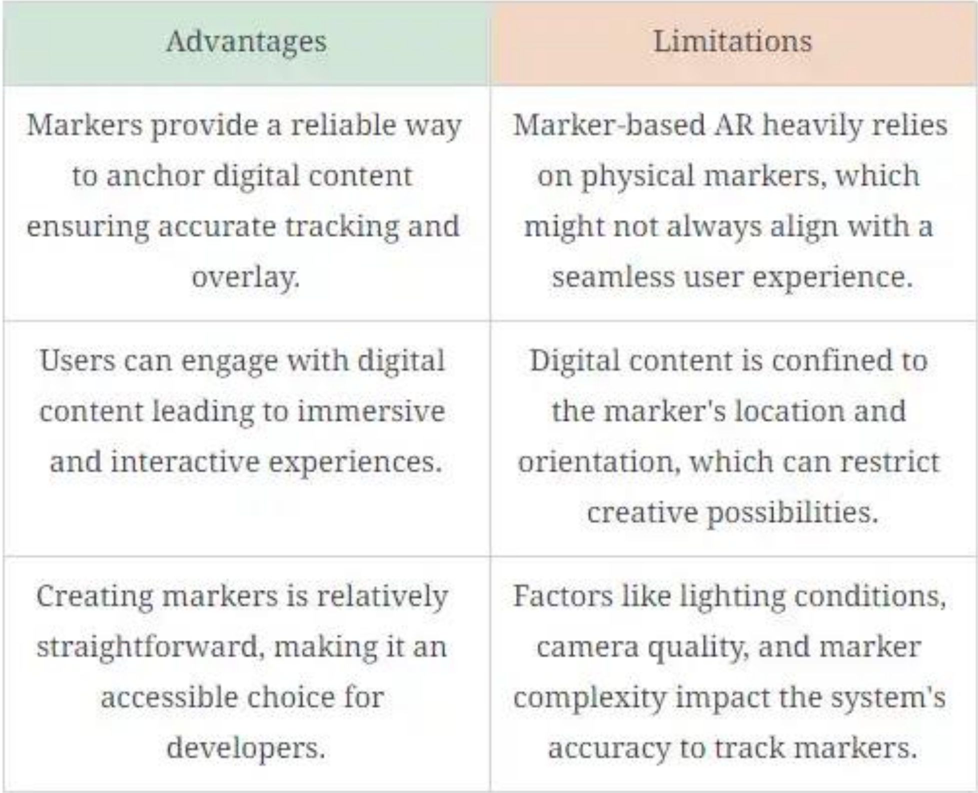
- **Marker detection**: The AR system's camera scans the environment for markers. These markers are usually simple, high-contrast patterns that can be easily recognized.
- **Pose estimation**: Once a marker is detected, the AR system calculates its position and orientation in relation to the camera.
- **Content overlay**: Based on the marker's pose, the AR system overlays digital content onto the marker in real-time. This content could be 3D models, videos, animations, or textual information.

Applications of marker-based AR

Marker-based AR finds its application in various domains, making experiences more interactive and informative. Here are some examples of its applications:

- **Education**: Imagine a biology textbook coming to life with 3D models of organisms when viewed through a smartphone. Students can explore these models from different angles, gaining a deeper understanding of the subject matter.
- **Marketing and advertising**: Brands often use markers on product packaging to unlock additional content, such as promotional videos or interactive experiences, when scanned with a mobile app.
- **Interior design**: Marker AR can let users visualize how furniture or decor items would look in their living space before purchasing.

Advantages and limitations of marker-based AR



Advantages	Limitations
Markers provide a reliable way to anchor digital content ensuring accurate tracking and overlay.	Marker-based AR heavily relies on physical markers, which might not always align with a seamless user experience.
Users can engage with digital content leading to immersive and interactive experiences.	Digital content is confined to the marker's location and orientation, which can restrict creative possibilities.
Creating markers is relatively straightforward, making it an accessible choice for developers.	Factors like lighting conditions, camera quality, and marker complexity impact the system's accuracy to track markers.

Multiple-Camera Infrared Tracking

In general, the known points in the world will not be constrained to a plane, as assumed in the previous section on tracking of flat markers. For tracking arbitrary objects, we require general pose estimation, which addresses the problem of determining the camera pose from 2D-3D correspondences between known points q, in world coordinates and their projections p, in image coordinates.

we describe a simple infrared tracking system designed to track rigid body markers composed of four or more retro-reflective spheres. It uses an outside-in setup with multiple infrared cameras [Dorfmüller 1999]. A minimum of two cameras in a known configuration a calibrated stereo camera rig is required. With this strategy, the additional input and wider coverage of the scene from multiple viewing angles will improve the tracking quality and the working volume. In practice, four cameras set up in the corners of a laboratory space are a popular configuration. Use of more than two cameras will improve the performance of the system, but is not fundamentally different from the stereo case.
The stereo camera tracking pipeline consists of the following steps:

1. Blob detection in all images to locate the spheres of the rigid body markers

2. Establishment of point correspondences between blobs using epipolar geometry between the cameras

3. Triangulation to obtain 3D candidate points from the multiple 2D points

4. Matching of 3D candidate points to 3D target points

5. Determination of the target's pose using absolute orientation (as described, for example, by Horn [1987] and Umeyama [1991])