zappar

glossary of terms

Helping demystify the word soup of AR, VR and MR

Zappar Ltd. 2017

Contents

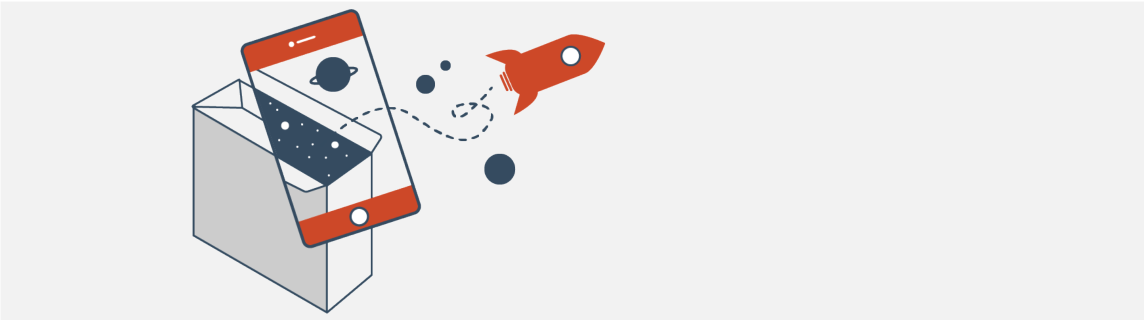
Objective

With every day the world gets a little more digital. And with that digitization comes a whole new language and set of terminology to grasp whilst also trying to piece together how these new terms and technologies fit together and impact our lives and businesses.

Given we live and breathe this stuff we thought it might be helpful to give a guide to what it all means. We're in the fortunate position that we have some of the smartest engineers on the planet who can give you the technical low-down. But what really makes them clever is their ability to translate this digital voodoo into plain English so you can explain it to your mum or boss (the two aren't mutually exclusive).

So here's a glossary of terms and overview of various platforms, tools and products on the market as a guide to the crazy world of AR, MR, VR and more...

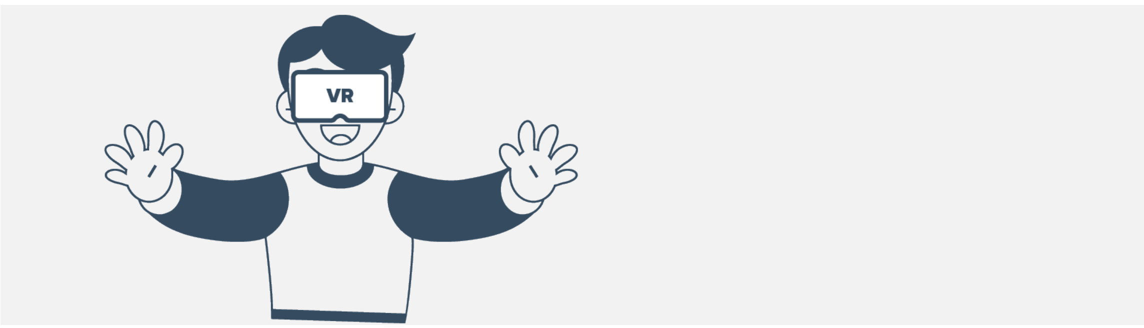
Types of Reality



Augmented Reality

In literal terms, AR takes your experience of the world around you and adds simulated, or virtual content. Much of the industry uses the term AR to refer to adding 3D or interactive content onto a live camera feed on a handheld device, such that the content looks to be present in the real world.

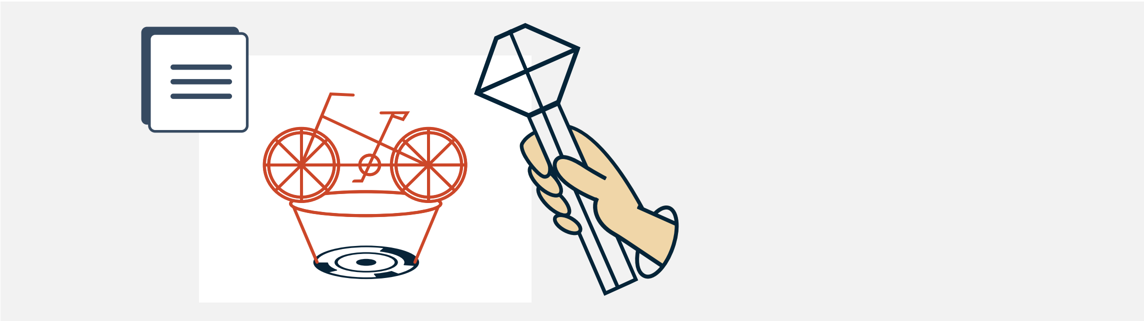
Show me



Virtual Reality

Virtual Reality replaces the real world around you with completely computer-rendered content. Often VR is experienced using a headset that can track the movement of the viewer's head to simulate an immersive 360-degree environment. Examples of VR headsets include Samsung Gear VR, Oculus Rift, HTC Vive and Google Cardboard.

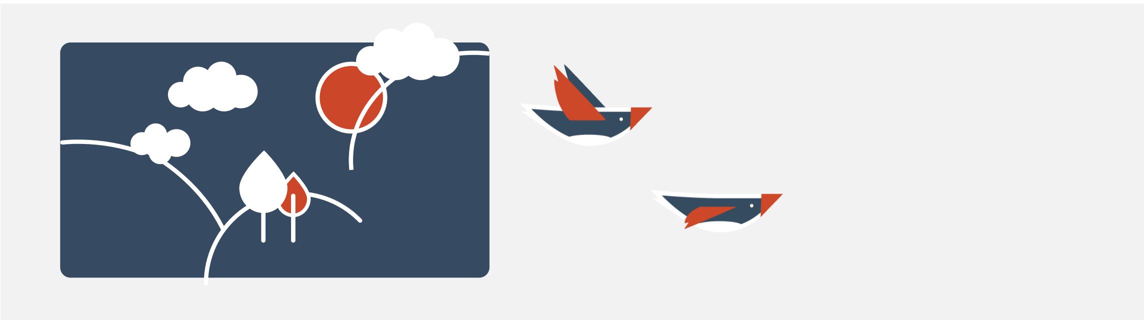
Show me



Mixed Reality

Like AR, Mixed Reality adds simulated content to the world around you. Many in the industry use the term to refer to experiences where the computer-rendered content is tracked to the environment, and viewed in an immersive manner, such as with a head-mounted display. The display allows the viewer to see both the physical world and the virtual content at the same time. In this light, MR is a subset of AR, albeit with more immersion, than is typical of handheld AR. Examples of MR headsets include ZapBox, and Microsoft Hololens.

Show me



Physical Reality

PR predates AR, VR, and indeed MR. Users experience the world around them using a variety of senses that come fitted as standard on most humans.

Show me

AR Tools

ARKit

This is a software framework shipped with iOS 11+ that provides a number of high-quality techniques (e.g. visual-inertial odometry, see below) for tracking 3D space using the camera and other sensors fitted on recent Apple devices. App developers can use the framework in combination with a 3D content platform to build AR experiences in App Store apps. As ARKit is an iOS-only library, alternative technologies are required to serve equivalent content to users of other mobile devices, such as Android.

ARCore

This is an implementation of visual-inertial odometry for Android devices, developed by Google. It exposes a very similar set of functionality to ARKit on iOS Devices. ARCore is still in pre-release as of December 2017 and only supports a limited number of high-end Android devices although increasing the number of supported devices is planned in time for the first public release of the framework.

Facebook Studio

Facebook Studio is an AR content platform and creation tool. Designers can use it to build filters and AR experiences to be served through the Facebook app.

Unity

Unity is a 3D game engine that supports both desktop and mobile devices. Content developers can build games that can then be submitted to the mobile app stores for distribution to end users. Apps and games built with Unity are often large to download due to the complexity of the runtime, and the visual and interactive fidelity of the gameplay experiences.

Vuforia

Vuforia is a software framework that developers can use to track real-world objects in 3D space. It is often used as the tracking library backing AR experiences built using the Unity game engine.

Snap Lens Studio

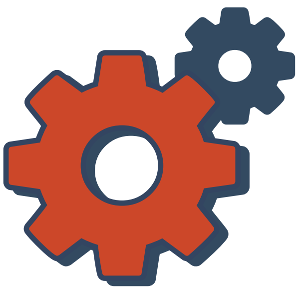
The newest kid on the block which allows developers to create "lenses" for users to place animated 3D objects in their photos and videos. Currently a small group of approved agencies can produce content for brands to appear in Community Lenses. Others can surface via Snapcodes which unlock the effect for 24 hours. Filters that use face tracking are not currently a feature of the Snap Lens Studio tool.

ZapWorks

Zappar's award-winning toolset for creating, publishing and hosting AR, VR & MR content along with other bite-sized mobile content experiences. Experiences can be richly animated and interactive, and can be downloaded on-the-fly without app updates. Content can be instantly published and viewed in the free Zappar app and can be accessed by scanning a unique zapcode or clicking a deep-link.

ZapWorks experiences can also be published into third-party apps, and the zapcode design can be completely customized to incorporate a branded icon. One example is the Shazam app which offers an Augmented Reality function powered by ZapWorks and using a custom Shazam code.

Show me



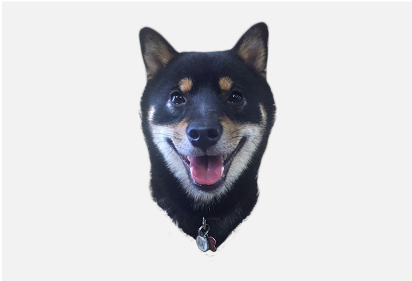
AR Elements / Assets

3D Models

This is another term for 3D objects - the things that you see on screen that appear to have 3D shape but are actually computer generated.

This football player **is a 3D model**

Spike is **not a 3D model** (he's a real dog!)



3D Model Formats

Like any form of computer content, 3D models, such as an animated 3D character, are shared and distributed as files between artists and then the devices that end up displaying them on screen. There are a plethora of different formats for this purpose, each with its own advantages/ history/drama. Some popular formats include:

FBX - this prevalent and featureful format is well supported across many 3D editing tools.

OBJ - despite lacking many features (such as animation), its simplicity affords the confidence that models will be displayed correctly between different 3D editing tools.

Dynamic Lighting / Real-time Lighting

This refers to techniques used by computers when displaying 3D content to simulate the effects of changing lighting on that 3D content. Content developers can set the properties of virtual lights, including their positions and colors, to achieve a desired aesthetic. These techniques often model reflection, specular highlights, shading and shadows.

Polygons, textures, maps, bones, skins, rigs

These are all technical terms related to 3D models (see above). Different configurations of these properties for a model will determine where a 3D model will work well (or at all). A 3D model designed for a movie will have many millions of polygons and huge high-resolution textures, and so will not be able to be displayed on devices with limited processing power, such as mobile phones. 3D models for use on phones must either be designed specifically for that purpose, or undergo a conversion process (often referred to as re-topologization, re-texturing, or optimization).

Alpha Channel Video

This is where a video has the set background removed so that the foreground elements (e.g the actors) can appear on an arbitrary background later on. Visual effects techniques like “green screen” make it easy to remove the background from live video. In the movie industry (and VR) the background is often replaced with a dramatic scene (e.g. an alien planet). In AR/MR it’s often a camera view, and therefore the user’s physical surroundings that sit behind the foreground video content.

Show me

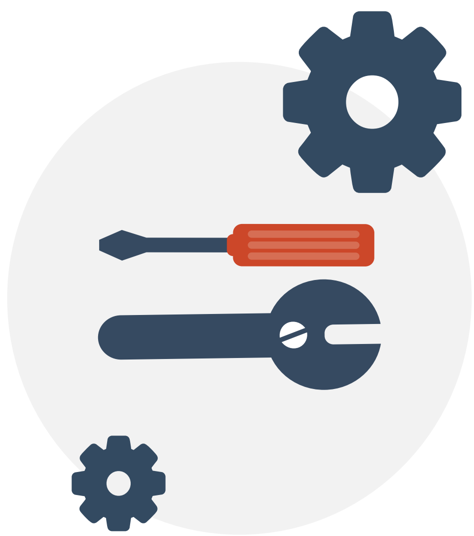
Photospheres / 360 Panoramas

This is a technique where an image shows the entire environment around a point in space. Users viewing a 360 panorama can look in any direction, either by dragging on the screen with their finger, or rotating a device that can track their motion (e.g a VR headset or a mobile phone with a gyroscope, see below).

Show me

360 Videos

See 360 panoramas, above, but where you read “image” replace it with “video”. It’s worth noting that high quality 360 video can be very bandwidth intensive and only devices with a good network connection will be able to reproduce a reasonable 360 video to its user without extensive loading or reduced visual quality.



Computer Vision and Mobile App Terminology

Image Recognition / Image Lookup

This is the term for when a computer identifies an image from a known set of reference images. An example use case would be where a user points their camera at an image in the world around them, and an image recognition algorithm identifies the image and provides relevant content to that user.

Cloud Image Lookup

Image recognition algorithms (see above) identify images from a known set of reference images. Cloud image lookup systems store the reference set on the Internet while end users' devices upload images to the Internet for detection. This makes it possible for the system to have a huge known set, perhaps millions of images, without having to store them all on the users' devices. Downsides of this technique include the requirement for good network connections, and that systems cannot discern between images that look similar or identical.

Object Classification

This is a computer vision term for algorithms that can identify the type of a pictured object. For example an algorithm shown the picture of an animal may be able to determine that it's a dog, and indeed its breed.

SLAM

This term stands for Simultaneous Localization and Mapping. It's a technique for working out where in an environment a camera is (localization), and the 3D structure of that environment (mapping). It accomplishes this using views of the environment from multiple locations, and by solving some tricky maths problems. While the results can be impressive, the algorithms have requirements both on the amount and type of visual detail present in the camera images, and on the motion of the camera needed to obtain the required views. These requirements make it tricky to build robust end-user experiences using a pure-SLAM solution.

Visual Odometry

Odometry relates to measuring how far something has moved rather than exactly where it is; and visual odometry is the process of working out this motion by processing camera images. Unlike SLAM (see above), visual odometry techniques do not aim to recover an exact map of the environment, which can lead to more robust performance in real-world use cases. Visual odometry allows for AR experiences where content appears in arbitrary space in front of the user, and keeps its position in the world as the user moves around, without requiring a specific image or object to track.

Depth Camera

The cameras that most people know about detect light - they produce pictures where the color at a point on the image represents the color of the light that the camera observed in that direction. For depth cameras, however, each point on the image shows the distance to the nearest surface that the camera detected in that direction. Depth cameras can help to relax some of the restrictions of SLAM or Visual Odometry algorithms - for example they allow flat surfaces to be detected even when the surface lacks visual detail.

Face Tracking

This is when a computer vision algorithm is able to recognise and follow a human face in 3D space. AR applications (for example Snapchat's Face Filters) are able to display virtual content on top of a tracked face, e.g. face paint, whiskers, comedy hats / masks. A depth camera (such as the TrueDepth front-facing camera on the iPhone X) can help robustness of face tracking algorithms.

OCR - Optical Character Recognition

Optical Character Recognition is when a computer system uses the visual appearance of a piece of text to recognise its constituent letters and numbers (a.k.a. the characters). Examples include the system banks use to automatically read the funny-looking numbers along the bottom of cheques (do you remember cheques?), or when apps like Uber can read your credit card numbers in from a photo.

Tracking

In the AR / VR / MR industries, tracking is when a system understands the 3D position and rotation of an object or environment, typically for the purpose of adding virtual content that appears attached to it. Computer vision algorithms can track a large variety of things these days, for example: images, faces, controllers, and arbitrary objects.

Target Images

When AR algorithms recognise and track a given picture, it's often referred to as a target image.

1D Barcodes

These are the standard barcodes that we all know and love - they're typically a set of bars and gaps, arranged in a line. Each barcode encodes a small amount of data, often an identifying number. Supermarkets often use these to know which products you are purchasing at checkout. Examples include UPC, EAN codes, and zapcodes.

2D Barcodes

Like 1D barcodes (see above) these encode a small amount of data. Unlike 1D barcodes they arrange that data over an area of space, rather than in a line. This means they can store more data than a 1D barcode, but may require more complex hardware (or software) to detect and decode correctly. Examples include QR codes, and DataMatrix Codes (DMC).

Gyroscope

The gyroscope (or gyro) is a sensor, present in many smartphones, that can detect rotational speed (changes in the direction a phone is pointing) in 3D space. Coupled with accelerometer sensors, the gyro allows software to display a 360 degree world on the screen that the user can look around by turning and tilting their phone. While most smartphones have a gyroscope, some lower-end devices may only have an accelerometer and magnetometer, resulting in very poor 360 degree experiences.

Accelerometer

This is a sensor that can detect the 3D forces felt by a device. It can be used to detect when the user shakes a device, or drops it. Since gravity is also a force (thanks Isaac) the accelerometer can be used by a phone to know which way "down" is. The accelerometer, coupled with the gyroscope (see above) are a means by which smartphones can understand their orientation in 3D space.

Magnetometer

This sensor detects magnetic fields. This allows the device to sense the Earth's magnetic field in the same way as a compass so it can identify which direction a phone is facing (North or East for example). Unfortunately the Earth's magnetic field is quite weak and especially indoors the signal can be swamped by metal or other objects in buildings. When outdoors the magnetometer provides a more accurate reading, but the update rate is still quite slow. Therefore the sensor data is often combined with (or "fused") with gyroscope data to provide smooth 360 degree experiences.

Geo-location

Your geographic location is where, on Planet Earth (or indeed on any other planet) you are. It's a pair of coordinates (longitude and latitude) that pinpoint your position. There are a number of ways that mobile devices can work out your geo-location; these include:

GPS - an elaborate system of satellites in space are used by a chip in your phone to triangulate your position (see below). For an app to get a user's GPS location it has to ask for the user's permission to do so.

Your IP address - when your phone is connected to the Internet it's possible to get an approximation of your location based on where your phone appears in the Internet's global infrastructure. While considerably less accurate than GPS, it does not require explicit consent from the user.

GPS

The Global Positioning System allows devices to have an accurate understanding of their position on the planet. It's used in a very large number of applications, from aviation to the maps app on your phone (see geo-location, above). GPS is an astonishing technology; and will be recognised as one of humanity's greatest achievements for centuries.

Mobile Web

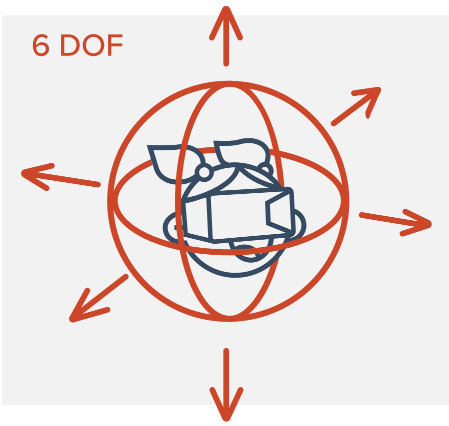
While the actual infrastructure of the Internet you view on a desktop computer is basically the same as that of the Internet on a smartphone, the mobile web refers to websites that have been specifically optimised for viewing on mobile devices. Mobile websites are visited by users using "web browsers" (such as Safari on iOS, and Chrome on Android). Sites can leverage the limited technologies and standards exposed by web browsers to provide end users with interactive experiences. Mobile websites are distinct from native apps (see below) which have deeper access to the various features and functions of the device they are installed on.

Native App

Apps are programs that can be installed on smartphones and typically provide the user with functionality that goes beyond what is provided by the core phone operating system itself. Apps can be more closely integrated with the features of a smartphone than a mobile web site (see above) and can generally provide a faster and more seamless user experience. This is particularly true of AR - access to the camera, and the ability to process the frames it produces fast enough, is only feasible in native apps at this time.

6DOF

This stands for 6 degrees of freedom, not to be confused with the 5SOS, an Australian pop band that toured with One Direction. It's actually a maths term. In order to describe exactly where an object is in 3D space, and which way it's facing, we need to know six things about it: three numbers to describe its position (e.g. X, Y and Z cm from the center of our room), and three to describe its rotation (e.g. it's rotated X, Y and Z degrees in various directions). The term is used in AR and MR when a device knows the full position and rotation of an object (perhaps itself, or a controller) in a 3D space.



Examples of 6DOF:

- HTC Vive and ZapBox know the full position and rotation of their user and controllers in 3D space

Not examples of 6DOF:

- The Nintendo Wii controller - the Wii only knows the direction the controller is pointing, not its full position in space. It has only 3DOF
- Google Daydream controllers - as above, only rotation is known
- Google Cardboard / Google Daydream Content - Just the user's head orientation influences the rendered content, so again only 3DOF

FPS - Frames per second

When computers show video on screen they do so by displaying one still picture after another very quickly, giving the illusion of motion. FPS refers to the number of such still images that are shown every second for a given video. In the AR/VR/MR industry it's a particularly important metric for video recording through cameras. Most smartphone cameras can capture at least 30 frames every second. This is good enough for most video watching (it's typically the FPS used for internet video) but a higher FPS (e.g 60) is recommended for VR or MR since human eyes are more sensitive to FPS when the video takes up a large part of the field of view.

Spatial Audio

This is when a computer system is able to play audio to a user such that the user perceives it coming from a point in 3D space. It's a technique often used in VR experiences to give a greater degree of immersion.

Video see-through

One of the two primary mechanisms by which MR headsets show the world to the user. A camera captures the world around the user and displays that, along with the virtual content, on a screen that sits in front of the user's eyes. The user only sees light from screen - the inside of the headset (including the screen) takes up their entire field of view. Example - ZapBox.

Optical see-through

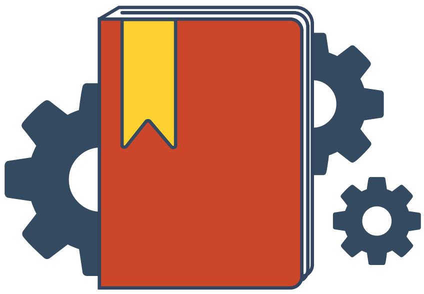
The other primary mechanism by which MR headsets show the world to the user. The actual light from the real world passes through a semi-transparent screen inside the headset to the user's eyes. The virtual content is then projected onto that screen from some clever contraption inside the headset. Example - Hololens.

Video see-through vs Optical see-through

Each of the techniques has its own advantages and disadvantages. Handily we've arranged some of them in the table below.

	Video see-through	Optical see-through
Field of view (i.e. how much of your vision can the virtual content appear over when you move your eyes around)	Large - lenses inside the headset mean that the screen can appear to take up a lot of your field of view.	Limited - the screens in current generation headsets are relatively small. This leads to the virtual content appearing only over a small area of your vision in front of you.
Opacity of content (i.e how fully can virtual content hide the real world behind it)	Complete. Virtual content can totally hide the real world behind it	Limited - current generation screens inside optical see-through headsets can't properly hide the real world behind the screen. Virtual content appears a bit like a hologram that you can see through a bit.
Latency of the real world in the user's view (i.e when you move your head, how soon do your eyes see the new place you're looking at)	Variable - the camera has to capture images then display them on the screen inside the headset.	Instant - since you're literally viewing the real world with your eyes.
Latency of the virtual content over the real world view (i.e. when you move your head, how well does the virtual content "keep up" with the motion)	Very good - the headset can ensure that the same latency applies to both the camera view and the virtual content.	Variable - the sensors and screen projection will always lag a bit behind the real world that the user can see through the semi-transparent screen.

The combination of all these factors affect how well a user will experience the system they are using. Higher latency, in particular, can lead some users to feel nauseous and may limit the amount of time they can spend in an AR/VR/MR headset experience.



THANK YOU

www.zappar.com