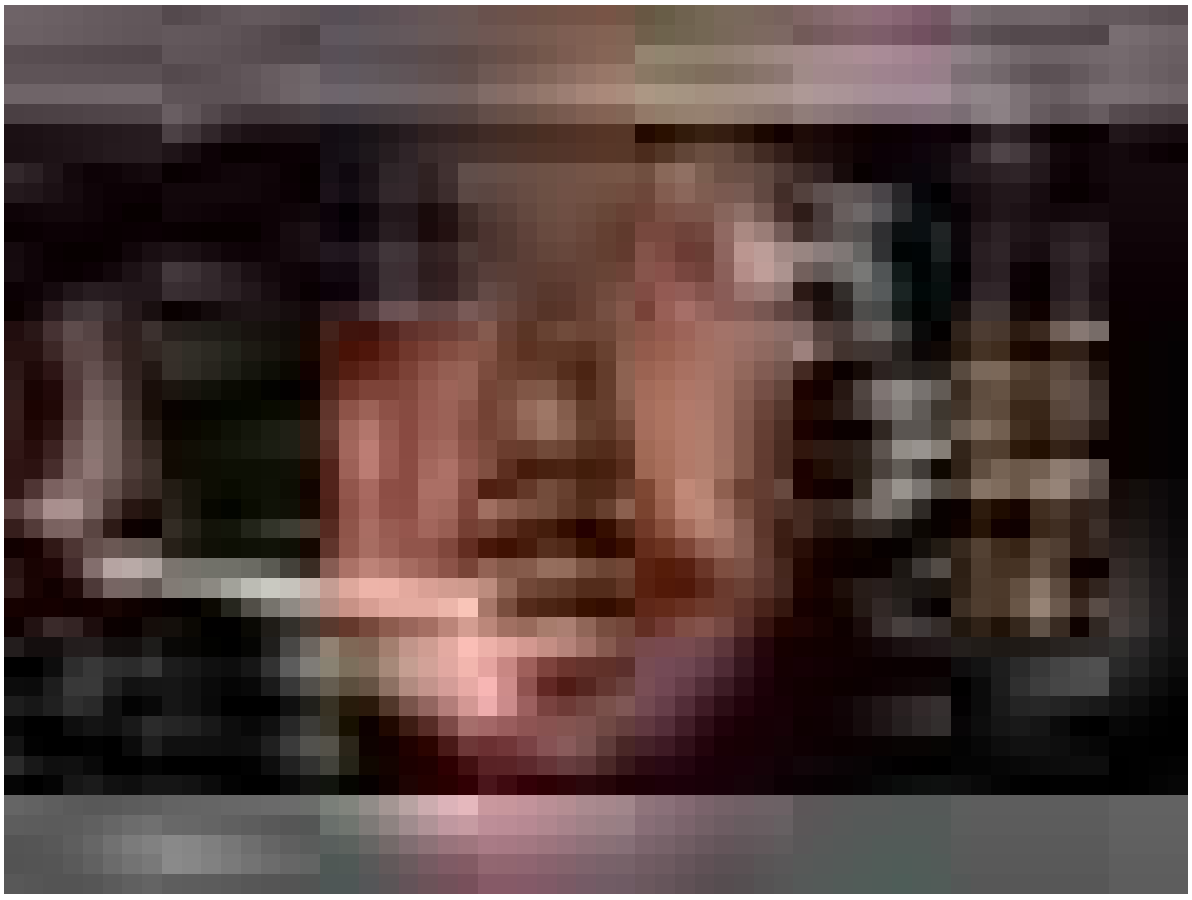
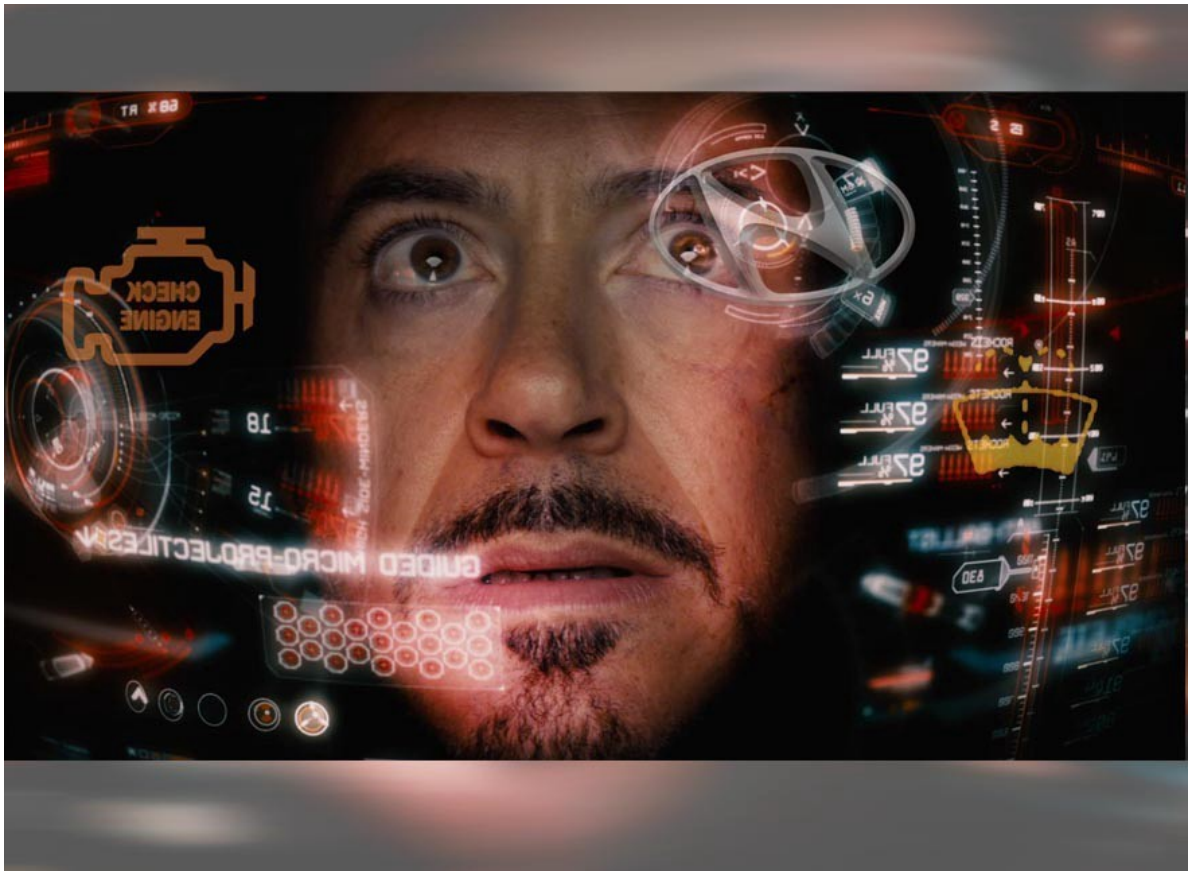
virtualrealitypop.com

Here's What It's Going to Take For Augmented Reality to Take Over The World

Andrew Kemendo

18-23 minutes





Tony seems…troubled. Image courtesy Marvel Comics

With rumors swirling about Apple's plans in Augmented Reality (AR), as well as recent teeth gnashing about the state of Magic Leap, it's beyond time for the AR community to have a real discussion about what it's going to take for Augmented Reality to become the primary computing environment worldwide.

> The question is, what will it take for an AR system to replace the smartphone and the desktop as the

ubiquitous computing device for everyday use?

I've spoken in the past about what I call the "AR stack." That is the set of technologies that are necessary to get to wide functionality and applicability for the average consumer. However I wanted expand on that and build a very basic reference that identifies key concepts necessary for AR, that people can easily cite when trying to understand this topic.

Form Factor

Image for post

I'd wear those

One of the key assumptions that we in the AR community can generally agree upon, is that consumers will demand AR displays that look as close as possible to the existing form factor for eyeglasses. That means that the user is looking through a transparent display. These are called Optical See-Through Head Mounted Displays. The general and most commonly used term which will be used from here on out is, Head Mounted Display (HMD). HMDs also include the category of Virtual Reality displays, which are completely opaque. That does not mean that there won't be significant adoption of AR before we reach that form factor, for example through smartphones, but AR HMD is arguably the idealized version.

I've heard on multiple occasions that even this wouldn't be good enough. The reasons most commonly cited: You can easily lose or break

glasses, they don’t work well while lying down (where many people use smartphones) and some people just don’t like wearing glasses.

Image for post

Wishful thinking

While I generally agree that there are constraints with the eyeglass form factor, I don't think that the contact lens will be where we see widespread adoption — partly because it's *probably* impossible in the coming decades.

Hopefully it will be clear after reading the following section, that it's hard enough to pack all the required elements into a pair of glasses. Adding the requirement of doing it on a contact lens, is not practically possible in the near future.

The Technical Requirements

Field of View (FOV): The idealized AR display will conform to the combined binocular and peripheral field of view of the human eye — which includes motion detection on the far edge of the FOV. That is approximately 200 degrees horizontal and 140 degrees vertical. Within this field, only about 140 degrees horizontal are binocular and the remaining 30–35 degrees are monocular peripheral vision.

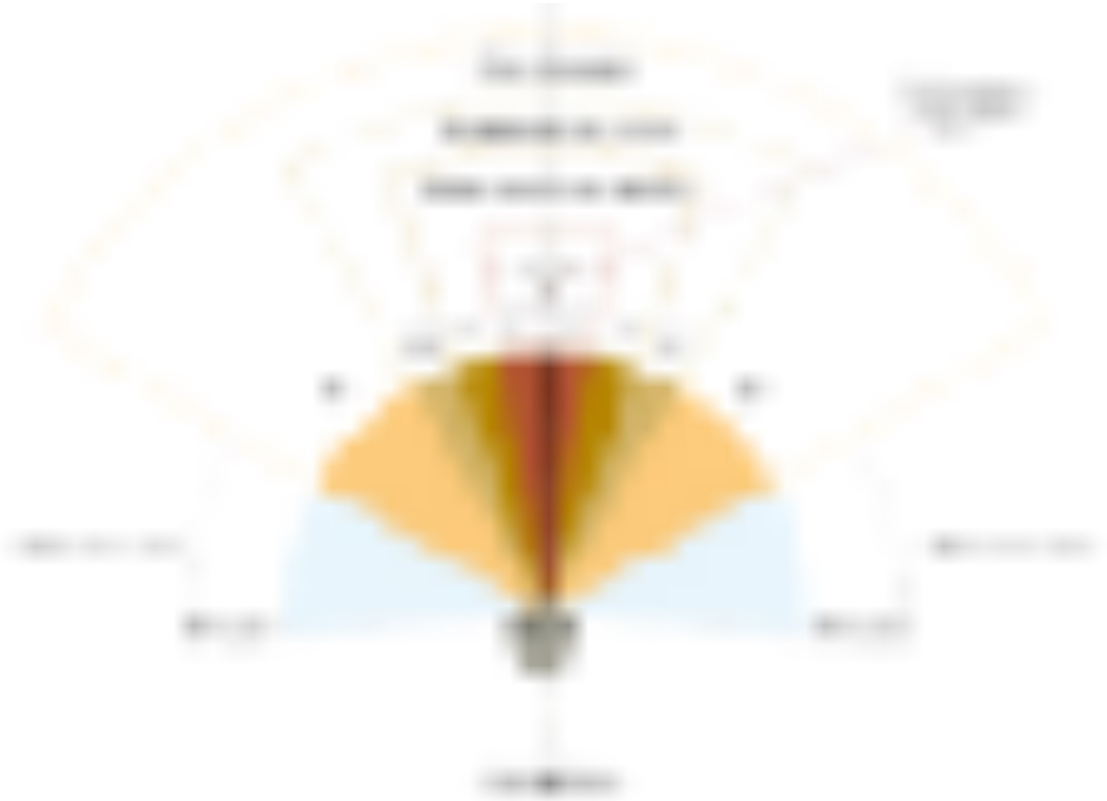


Image for post

Human Horizontal FOV

Pixel Per Degree (PPD): The idealized display would have equivalent pixel per degree resolution as the human eye at any depth of focus. That means a human can't discern a difference between rendering elements (ie. pixels) at that resolution. While PPD and human eye "resolution" is not a one-to-one comparison (we don't have pixels — though you could argue the point with respect to rods and cones), we can consider 60 PPD as the human equivalent necessary for a convincing display. This is the first key element required to eliminate the problem of eye fatigue with digital display systems.

Latency: In a similar nature as PPD, the latency of the visual pathway should mimic as close as possible to the human biological visual processing pathway. Experiments put humans at reliably discerning differences in visual stimuli at around 77 Frames Per Second (FPS) or 13 milliseconds between being shown a new image and recognizing it as such. To take a page from the VR world, if we want to be conservative we should

expect that anything worse than 60 Frames Per Second would be unacceptable.

Accommodation: When you want to focus on an object close to you, muscles around your eye, deform your eyeball lens ever so slightly to focus the incoming light (in the form of multiple and varying wavefronts) onto your retina. This process happens in reverse when you want to focus on something far away, and does so without conscious thought. This process is called accommodation, and allows us to change our focus simply by refocusing our gaze. Ideally, a display would give us unlimited focal points for which we can focus our gaze. In practicality I expect that an AR system would only need to provide between 5 — 10 different focal depths to be deemed sufficient (too little data to know with much confidence). This is the second key element required to eliminate the problem of eye fatigue with digital display systems.

Image for post

Depiction of wave-front distortion with distance

“Rendering Black”: One of the more contentious topics in AR is how, and even if it is possible, to successfully show black in a see through AR environment. Given that “black” is a relative lack of light coming into the eyeball from a specific vector, it’s not simply as easy as reducing the strength of

the emission that you are sending to the eye in a see through transparent display. The visual cueing that humans get from shadows and light effects, are very powerful, and are a key aspect of creating presence of a virtual object in AR. [Note: There are theoretical ways to do this but none have been proven or practical. The most promising technique to render black pixels that I know of is to have an optical element (a lens) that selectively blocks all light waves from a specific direction, regardless of wavelength.]

Power: Ideally this is a device that you can put on in the morning and take off at night without needing a re-charge. Matching iPhone 7 specs would put that between 10 and 13 hours of consistent use.

There are ungodly number of ways in which the community is trying to address these issues given the current state of technology.

The primary display method we see in practice, utilizes fixed micro OLEDs or LCDs, which are reflected through a series of waveguides presenting the image into your line of sight.

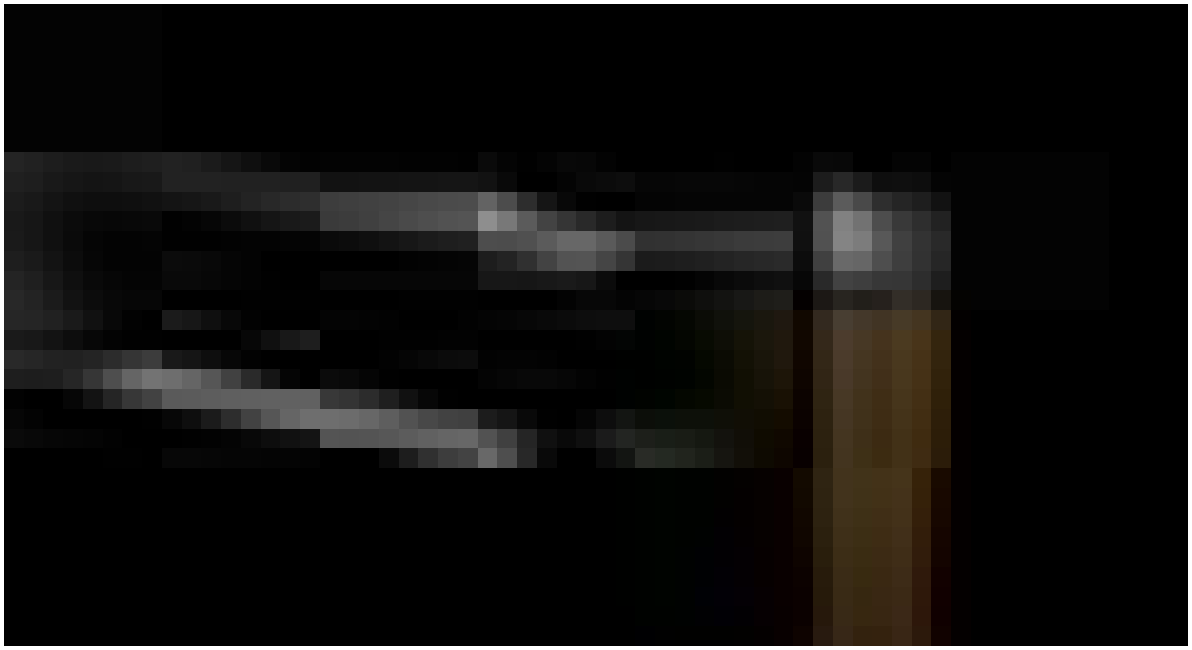


Image for post

Moverio BT-300 waveguide and LCD. Image Courtesy Epson

While nearly ubiquitous, this method has hard limits on FOV and PPD. Further, there is no known way to give accommodation utilizing this method.

I consider Virtual Retinal Displays (VRD) or Retinal

Scan Displays the best hope for achieving consumer level AR display technology. Introduced long ago, but becoming increasingly feasible in practice, these displays can project multiple wavefronts with various curvatures directly onto the retina, allowing you to focus at different depths. Done correctly, this can potentially give you a display with nearly unlimited FOV, accommodation and extremely dense PPD on the order of human acuity limits.

Image for post

VRD diagram. Courtesy: https://www.google.com/patents/EP0562742A1?cl=en

> BOTTOM LINE FORM FACTOR: For AR to be the primary computing device the form factor needs to be at least a pair of standard eyeglasses, which give human equivalent Field of View, Pixels Per Degree resolution, and Accommodation at a high enough refresh rate and a day worth of battery life . I'll hold judgement on whether "rendering black" is required for wide scale adoption.

Environmental Tracking

I wrote about environmental tracking on the Pair Blog last year but it is worth discussing briefly why this is important.

In the real world, physics dictates that objects generally stay in the place where they are set, relative to the objects around them. In the Augmented world we can't take this for granted.

In order for you to see an object in an AR environment and have it look convincing, there needs to be a way to "anchor" the virtual object to the real world. In an AR environment, this consists of tracking real-world objects (really image features) and translating their location to a virtual coordinate system. In practice this tells the system things like how far the viewer is to a real object, the height of the viewer with respect to a real object and so on. Implementations of this are generally captured in the discipline of Simultaneous Localization and Mapping (SLAM), though there are less robust ways to do tracking which will not be discussed.

> Think about it like adding gravity to the Virtual World

Once you combine the spatial awareness of the real world, with a representation of it in the virtual

world, then you can add new objects into the real world. As long as you maintain this combination, these worlds will synchronize such that when a viewer moves, the view of the object moves in-sync with the view, the same way as objects in the real world do.

Image for post

Monocular SLAM on an iPhone. Courtesy Pair3D

As mentioned in the previously published article, there are several ways to do this. You can either do it passively, by interpolating a 2D video feed (or feeds) you receive from the real world; or you can do it actively by sending out laser, infrared or ultrasonic emissions to return depth estimates immediately. The distinction between these is in the hardware necessary and the fidelity of depth estimates.

The best in class systems today utilize active IR or Laser depth systems to track the environment.

Ideally these systems will work faster than the display systems do, however when we look at size and power requirements of existing best in class systems, they are not yet able to do so in a very compact form that would go into the idealized form factors described above. It may be ideal to use two passive cameras (stereo RGB) for tracking in our idealized form factor, as they are small and need very little power, however advances in shrinking

active tracking systems hold significant promise.

> BOTTOM LINE ENVIRONMENTAL TRACKING: An AR system needs to be able to accurately maintain a virtual object's placement in the real world, such that virtual objects convincingly and reliably stay where placed, regardless of where or how the viewer moves.

Content and applications

Think of every application that you use today on your desktop or mobile device. Everything from spreadsheets to candy crush. Ok, now add an entirely new category of interactivity that you can have with real-world objects and places. Here's a brief list of those:

Everything from the User Interfaces to the 3D models, need to be created, curated and placed into the world in a way that people using AR will encounter, interact and collaborate with them in the right context. This content needs to be basically everywhere for it to work at scale such that 1. You don't need to rely on another device for your data

and 2. It actually makes your daily life easier or better.

Oh, and that means that the content can’t be a bunch of spam, AR popups and general noise that overloads your senses.

Image for post

Please no. Image Courtesy Keiichi Matsuda’s

Hyper-reality

Practically speaking it's probably not necessary to replace every possible application on the planet in order for AR HMD to have smartphone level penetration. However there needs to be enough content and applications to replace most of the core functionalities of smartphones: Messaging, Calling, Search, Navigation, Shopping and Video/Image playback.

This also means that, like every other computing platform, we need to build a strong community of developers, content creators, designers, hardware engineers and advocates, who will spread the word and build the experiences that people want.

> BOTTOM LINE CONTENT & APPS: An AR system needs enough content and use cases that it replaces or enhances current functionality of digital systems used today on smartphones, desktops and laptops

Ubiquitous Mapping

Speaking of having content everywhere and accessible to anyone, something that is massively overlooked, and rarely discussed in the AR community, is the need for a common georectified AR environment.

If you remember *way back* to when PokemonGo came out this summer, you might recall throngs of people going to very specific places to catch certain Pokemon:

Image for post

PokemonGo users swarming a park

That's because the creators wanted you to feel like you had to actually interact with the Pokemon where they were in the "real-world." The concept of contextual content is certainly not new. Many apps and products only work in certain geographic places (geofencing) and others require physically seeing some feature to activate certain content (most marker based AR). In practice, Niantic used geofencing to give Pokemon their homes, so it was possible to do at scale without big leaps in technology.

Imagine now, that you put on the perfect form factor AR HMD. It has great tracking capabilities and there is great content out there to be discovered. How then will the AR system know where the content you want to interact with is to high enough positional accuracy? GPS? Unfortunately GPS has wide margins of error,

averaging about 5 meters (16 feet) of accuracy. It also fails indoors.

Well the answer is a ubiquitous re-localization map. Effectively this looks like a copy of the real-world, but with all of the virtual layers and content georectified inside of it. It also needs to be so good that a user simply needs to looking at an object and you will find the associated digital content that lives around it. For an AR HMD system to work seamlessly, it needs to find content both indoors and outdoors and serve it to the user with inch level accuracy.

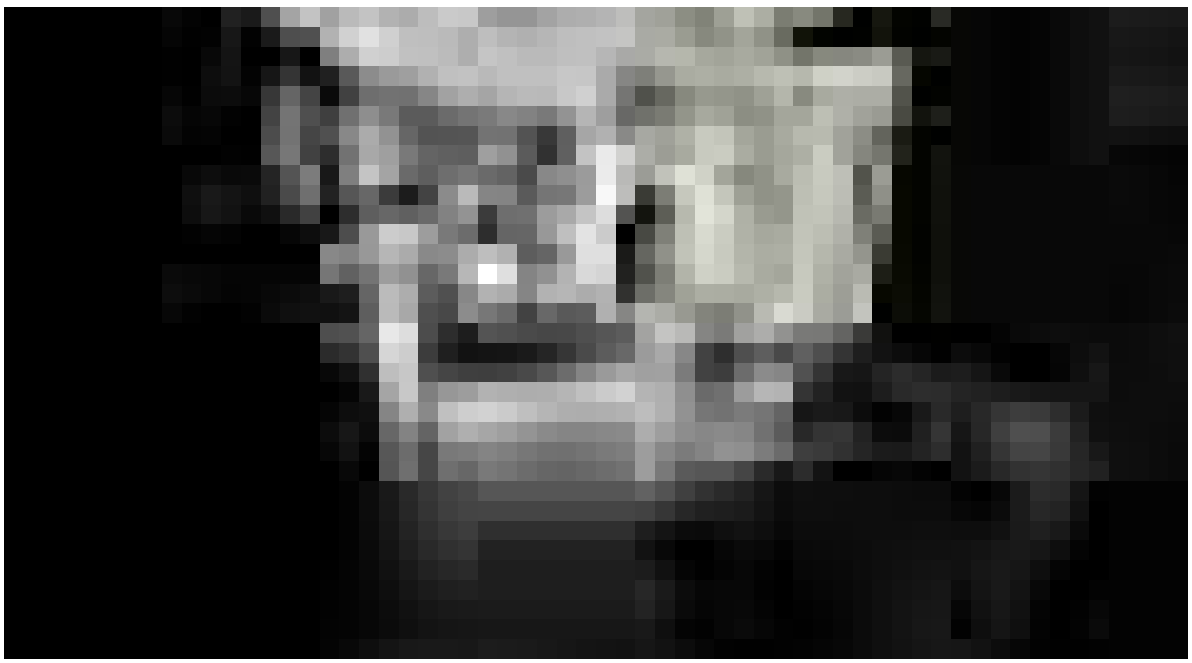


Image for post

PointCloud Map of UCD optics Lab. Image Courtesy Ruth Kerr

This is very much tied into the tracking capabilities, but broadens the scope of the localization portion such that every user has a precise geolocation that maps to the georectified map.

> BOTTOM LINE FOR UBIQUITOUS MAPPING: To have AR everywhere, there needs to be a massive consistent and accessible 3D map, consisting of virtual content that is georectified to the physical world down to inch level accuracy. I will note that I think this is the most important piece of the AR puzzle and has the highest potential value.

Human Factors Engineering

Something the AR community doesn't have much

research or practice in, are the human factors engineering problems for an AR environment. That's not to say people aren't thinking about it, but it's not well worn territory.

Gaze

When we look historically at the mechanics of how people consume and create content, some interesting trends are revealed.

Since the invention of parchment, we have gazed downward at our content. Whether we are writing with pen, typing on typewriter, or reading and watching, this trend has held true forever.

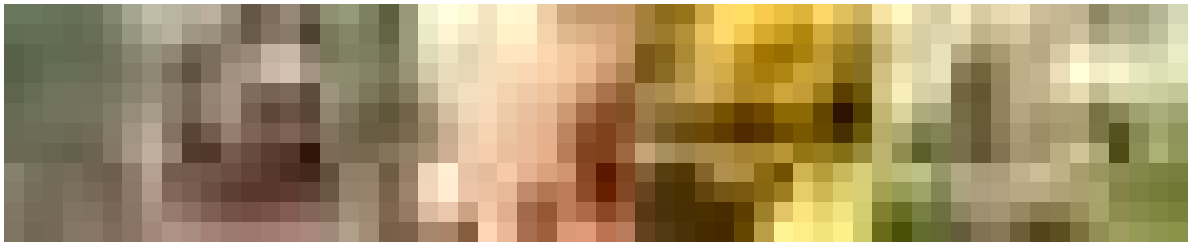


Image for post

Always looking down

Certainly we gaze forward and around while walking, talking, driving and passively interacting

with the world — for example watching theater or a movie — but nearly all interactions we have with digital or print content is while we are gazing downward. The reason for this is a mixture of social and ergonomic, but not worth getting into depth here.

Image for post

Downward gaze is best for Ergonomics

Should you be checking your twitter feed while you are engaging in conversation with someone? Probably not, and it will be obvious that you aren't actually listening to them. Nearly all proposed AR HMDs, follow in the wake of Virtual Reality HMDs,

in that they assume the user will predominantly be forward gazing — so they put the bulk of information and user interface directly in front of the user.

This is misguided. It ignores the fact that VR and AR are extremely dissimilar when it comes to how long you are in the environments (a few hours vs all day), and what "natural" actually looks like in the different contexts of day to day life.

Inputs

There seems to be this silly idea that gesture input will dominate AR interactions and input.

Flailing around

Try this: Get a stopwatch out and start it. Now hold your arms out in front of you and see how long it takes before you get uncomfortable. You can even bend your elbows if you like. How long did you last? Probably not more than a few minutes. Now imagine doing this gesture all day long for everything. I personally don't think it's practical, comfortable or reasonable.

But wait, don't we deal with real objects all day long with our hands? Well kind of. The actual amount of time you spend reaching, grasping and manipulating really depends on your job type, ability level and age. If you are a miner, construction worker, or other blue collar worker, you probably have years of repetition and strength built into your upper body — not so much for white collar workers, the elderly or the disabled.

> By proposing gesture based interactions as the core input mechanism for AR, we are telling white collar workers, the disabled and the elderly that to use these new computing devices they will have to gain strength in their shoulders, arms and chest.

There is a lot of woo-woo around the idea that forward gazing, gesture based interactions are "natural," and that all current AR displays are mimicking the natural way we do things. It would be useful for those stating as much as well as future AR UX designers, to take a look at ground truth ergonomics and human factors in current daily life, rather than looking to science fiction for

this.

> BOTTOM LINE FOR HUMAN FACTORS: AR needs to adequately factor in how humans naturally interact with digital content and the real-world in different contexts, so that we deliver the most seamless blending of the real and virtual worlds in the most comfortable ways.

Privacy and Security

What happens at the societal level when everyone has a camera array and depth sensor on their face that is constantly recording and mapping? How do our systems distinguish between private and public spaces? How do we protect private data in a pervasive mapping and tracking environment?

There are well documented cases of people being attacked for wearing a form of HMDs. Kyle Russell being attacked for wearing his Google Glass was a well publicized example. Steve Mann's assault at a McDonalds for wearing his EyeTap was also well publicized. It's clear that people aren't broadly ready to interact with such systems.

This is not a new concern, nor is it intractable. There are solutions proposed, but practically, we will probably see technologies deployed to the public before we have robust solutions. We shouldn't ignore these concerns completely, but there is a trend where we "figure that part out later" once we have sufficient penetration of these technologies. I generally agree with that approach, as we have adapted so far (citation needed) to the massive growth of social media and all that has come with it. It's certainly worth us having a public conversation about what the risks and benefits really are as AR is potentially orders of magnitude more impactful.

This is probably the hardest problem in the group of problems, as it takes a shift in how we collectively view this kind of technology and the trade offs in the broader scope of society. I fear that we will rush into this and see a backlash if not done correctly, which could set us back for widespread adoption.

> BOTTOM LINE FOR PRIVACY AND SECURITY:

> We in the AR community need to understand and explain the benefits and drawbacks to pervasive AR — with the proof being real-world usage.

Summary

For an AR system to become the ubiquitous computing device, the form factor needs to conform to how humans contextually interact with digital content and the real-world.

It should be similar in form to common eyeglasses, with display fidelity mimicking the real-world and the ability to be used all day uninterrupted.

It needs to convincingly display virtual content in the real-world, with a sufficient number of applications and content, accessible everywhere regardless of environment, such that it can replace the most common computing use cases with better quality than alternatives.

Finally, and critically, it needs to provide so much value that any social drawbacks are outweighed by the overall measurable improvement in quality of

life.