Immersive Cinema is Ready for its Close-Up

Andrew Leitch

The woman looks simultaneously defeated and defiant as she sips her tea at the edge of a cliff, and the camera seems to capture every emotion that flits across her face. She's shot in a medium close-up, with a soft vignette enveloping her, but the camera that's capturing this powerful moment is not your run-of-the-mill cinematic setup: it's a custom-built 3D rig with an exceptionally high resolution (a pair of 8k lenses, one for each eye) and frame rate (90 fps), and 16 stops of dynamic range.

The end result of all of this technology is to make you, the viewer, feel like you're physically present, sitting right across from this woman in an intimate, reflective moment, bearing witness to her frustration, and her resolve.

The film is called “Highlining,” and it follows Faith Dickey, a world champion highliner, as she repeatedly attempts to cross a rope vertiginously strung 3,000 feet above the Norwegian Fjords. And it is shot in the Apple Immersive Video format (AIV), specifically designed to be viewed inside their Vision Pro headset.

“Highlining” was one of three Apple Immersive films that launched with the Vision Pro in February of 2024 and is, in my opinion, still one of the most persuasive examples of the power of this new medium.

The many AIV films that have followed since have, almost without exception, also been spectacles: back-country skiing in British Columbia, perilous rock climbing in Majorca, floating above majestic landscapes in Hawaii, or Iceland, or Turkey. And it’s not just people doing amazing things in and above majestic landscapes. Sports (NBA, NFL), music (Alicia Keys, Metallica) and even dog shows have all been given the same immersive treatment, and they all share the same uncanny sense of presence that I felt when I was “sitting across” from Faith Dickey.

The technical term for this effect is one-to-one world mapping, where the two camera lenses mimic your left and right eyes to form a stereoscopic view of the world, and everything that’s in front of you (if it’s shot correctly) feels like it’s from your point of view.

3D cinema has been around for about eighty years, and the more recent revival of the form, kicked off by the first installment of “Avatar,” came out in 2009. And cinematic 360 video (which is the medium I’ve been most engaged with) started emerging around 2014 for early VR headsets like the Oculus DK1.

But what you experience inside the Apple Vision Pro is an order of magnitude more immersive than the disposable-glasses-enabled 3D that most people have experienced at their local cineplex. There is a parallax depth that makes it feel like you could almost see around the edges of the

objects in front of you.

But although I've been completely transported by watching these immersive videos (and I've watched them all, many of them multiple times), I keep hoping for something more — something that moves beyond the spectacle (as impressive as it is) and towards what could become a new form of narrative drama.

Or let me put a bigger stake in the ground. I believe that the possibilities presented by Apple Immersive Video[1] point towards a fundamental shift in how stories could be told and represented in a cinematic environment.

Most of us have heard of the famous Lumiere Brothers moment from cinema's pioneering days, when the audience ran terrified from the theater as a train pulled into the station, convinced it was going to come barreling out of the screen and crash into them. Although the story is probably apocryphal, it points to a fundamental truth, which is that the power of a new medium can be astonishing when you first encounter it.

Blake Williams, in his excellent article[2] about Immersive Cinema, references the three stages that all new cinematic formats follow[3]:

Stage One is Demonstration (that Lumiere Brothers moment again). This stage shows off the new medium's amazing technical capabilities, and it's where the bulk of Apple's immersive videos fall.

Next comes Restraint, where there is an attempt to fit the new technology into the old forms. Again we can look to our cinema history, where some early practitioners would bolt a tripod to the floor in front of a proscenium arch and film a 10-minute version of Hamlet (10 minutes being the time limit that early camera technology allowed for).

The Apple-funded short film "Submerged" is a perfect example of this Restraint phase. While it had some powerful dramatic moments of a

WW2-era submarine under attack, it felt like a missed opportunity to more fully explore what could be done with this new form. It leaned heavily on tried-and-true cinematic conventions, like shooting with a shallow depth-of-field, or cutting away to a close-up of a scurrying cockroach. These techniques, which can be so effective in a "flattie," instead feel jarring and awkward in this new medium.

Finally, we come to Stage Three: Assertion, where a new language starts to emerge — one that's uniquely suited to this new medium.

And this is where we find ourselves. We are at the exciting "wet clay" moment, where the rules are still being worked out, and nothing has solidified into a coherent or established form.

As someone who has spent 10,000 hours in the dark[4], (a lot of it in New York's repertory houses) I have a deep appreciation of how cinematic language, when used to its greatest effect to tell stories, can transport and transform us.

I've also spent a good number of hours (if not quite 10,000) in various midtown and downtown New York theatres, impressed by the immediacy of watching some of the finest actors in the world ply their craft in real time.

So what I'm imagining for Immersive Video is a new language that exists in a sort of liminal space between Theater and Cinema, and which takes elements from both to define new ways of telling stories. I like to think of it as a virtual proscenium arch that has come alive with digital possibility; a theatrical canvas onto which cinematic imagery and spatial audio[5] are projected.

This new language can take full advantage of the vast array of cinematic and theatrical conventions that work within this format (while jettisoning the ones that don't) to develop new modes of expression that tell stories

in ways that have never been experienced before.

This is where understanding the one-to-one world mapping that AIV allows for is so important. It actually replicates how we see and hear the world. In their new book[6] about Apple Immersive Video, Ben Allan and Clara Chong call this "frameless storytelling." You no longer direct attention with cuts, focus pulls and camera movements within a fixed frame. Instead you direct the gaze inside the world, using elements like blocking, bodies moving in space, depth perception, and again, crucially, sound design.

To highlight one important aspect of this sense of presence their book details, the fidelity of the resolution in AIV allows the viewer to perceive a face's micro-expressions. Micro-expressions are very brief, involuntary facial expressions that can reveal a person's genuine emotions, even when they're trying to conceal them. They typically last less than half a second, and we're instinctively wired to pick up on them. (Trained interrogators are very good at spotting them, particularly if you're lying.)

It is possible to see micro-expressions in regular cinema of course, especially in an extreme close-up on a big screen. But there is an intimacy and immediacy to them in the AIV format that feels new and unique.

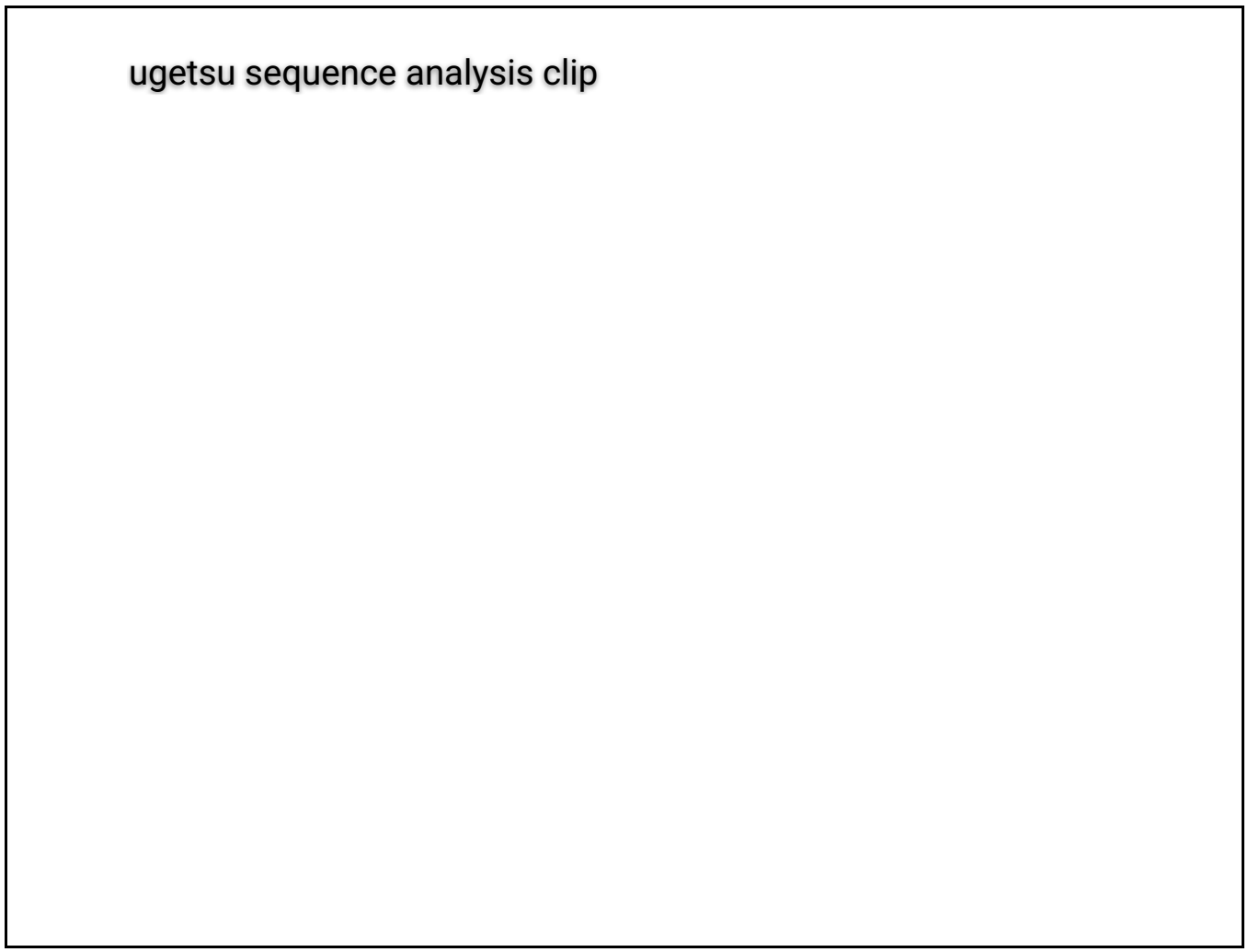
This is what makes the moment as we sit across from Faith Dickey so powerful. Without even being aware of it, we are instinctively picking up on her emotions and empathizing with them. And it reinforces my central thesis, which is that the possibility of using this new medium for narrative drama could be truly transformational.

It is not just about the micro-expressions of course, but the whole experience of feeling like you're inside the scene, witnessing the drama unfold right in front of you. Going back to my digital proscenium arch, you are in the first row, middle seat, and every other viewer who engages with the experience is also in that exact same excellent seat. (This proximity

effect is not limited to drama. One of the most exciting recent developments in AIV is the live-streaming of NBA games, where every spectator gets a court-side view of the action[7].)

Pulling from my 10,000-hour archive, I can imagine how some of my all-time favorite cinematic moments could be transposed into fully immersive experiences.

The first example is from Ugetsu Monogatari (Kenji Mizoguchi, 1953).



A husband, Genjurō (Masayuki Mori), returns to his war-ravaged village to search for his wife, Miyagi (Kinuyo Tanaka). He runs into his home, now an

abandoned wreck, and vainly calls out for her. The camera pans away from the center of the house to follow him walking around, desperately trying to find her, and then, when the camera circles back to the center of the house (all in a single shot), Miyagi, who is plainly now a ghost, is sitting there, tending to a pot over a roaring fire.

In this scene, I can imagine the spatial sound design playing a crucial role in directing attention. The camera still pans around the room, but with one-to-one world mapping it feels like you're exploring the space, slowly turning your head to take in the devastation. As you gaze around the dilapidated space, you suddenly hear the crackle of a fire coming to life, seemingly behind you. When you turn your head to locate the source, Miyagi has miraculously reappeared.

Another moment is from Five Easy Pieces (Bob Rafelson, 1970).

At the end of the film, the main character, Bobby Dupea (Jack Nicholson) returns to the family home after a long absence to visit his father, Nicholas (William Challee), who has been paralyzed by a series of strokes. After pushing his father in his wheelchair to a nearby field, Bobby is futilely trying to engage with him.

The bulk of the scene cuts between Bobby and his mute father, both shot in medium close-up. The juxtaposition between Bobby's poignant recounting of his life's struggles and disappointments and his father's impassive face is what gives this scene such emotional power.

Again, by taking this sequence into our frameless world, there is no longer a need to intercut, but rather Bobby and his father would both be visible for the entire scene. You could follow the interaction in real time, choosing which actor to focus on, and viscerally feeling the intensity of Jack Nicholson's performance as he starts to break down in front of his mute father.

My third selection is from a more recent film, May December (Todd Haynes, 2023).

Natalie Portman plays an actress, Elizabeth Berry, who is preparing for a role based on the life of Gracie Yoo (Julianne Moore), an older woman who had infamously started a relationship with and then married a teenage boy.

After spending time with her family throughout the film, Elizabeth has absorbed Gracie's characteristics, and reads from a letter that Gracie wrote to her then "boyfriend" when they were first getting together. It is shot in a single take, with Elizabeth speaking directly into a mirror as she adopts Gracie's persona.

It is an acting tour de force from Natalie Portman. She is able to uncannily mimic the distinctive mannerisms and speech patterns of Julianne Moore's character, which we've come to know through the course of the film.

In this example, rather than Natalie Portman addressing the mirror, it would feel like she is speaking right at you, making direct eye contact. As she transforms herself into Gracie with such eerie precision you would be right there with her, witnessing her complete metamorphosis into another character's persona.

To be clear, these examples are already compelling works of dramatic art, and I am not suggesting they could be "improved" by being rendered in an immersive environment. It is more of a thought exercise to imagine how some of my favorite cinematic moments could be reinterpreted into this new form of storytelling using elements of this new language.

Of course, there are countless other examples that could be equally captivating. I'm thinking of Gloria Swanson descending the stairs at the conclusion of Sunset Boulevard (Billy Wilder, 1950), which is also, of

course, where my title is borrowed from.

Or Robert De Niro as Travis Bickle in Taxi Driver (Martin Scorsese, 1976), repeatedly asking "You talkin' to me?" as he stares psychopathically into the mirror. (The idea of feeling like you're in the same room with Travis Bickle is particularly terrifying, and points towards the myriad possibilities that the thriller or horror genres could explore in this new medium.)

I know there are a lot of talented filmmakers[8] out there already working with immersive video, and the recent arrival of a new camera specifically built to shoot Apple Immersive Video[9] is opening up even more possibilities.

And I'd like to think that one of those filmmakers will soon be releasing a narrative drama that will fully take advantage of this new language in this new medium, and bring to life the tremendous storytelling potential of immersive cinema.

Let's watch this space.

Notes:

1. At the time of writing, the Apple Vision Pro is the only headset I've tried that is capable of fully rendering these highly immersive experiences, but there will inevitably be competing video formats that emerge in the coming years.
2. See Blake Williams' full article here. It is highly recommended reading, and I'm heavily indebted to his acute analysis as a point of departure for this piece.
3. The three stages referenced here are laid out in Scott Higgins book on the history of technicolor, Harnessing the Technicolor Rainbow.
4. David Thomson, author of one of my cinema bibles, The Biographical Dictionary of Cinema, remarked that he initially wanted to call his massive tome 10,000 Hours in the Dark.

5. Using the new Apple Spatial Audio Format (ASAF), which builds on well-established spatial audio protocols, AIV can also replicate (to a stunning level of fidelity) a real-world soundscape. This is a rabbit hole for another time (and another article) but having a fully immersive 360-degree sound field that literally wraps around your head is a foundational building block of the medium.
6. Cinematic Immersive for Professionals is a comprehensive overview of this new language, coupling practical recommendations with a strong theoretical grounding. More highly recommended reading.
7. Ben Thompson has a great take here on the Vision Pro viewing experience, which echoes my central thesis but applied to sports, not drama.
8. A shoutout to my friends at Step Into Vision, a smart of group of folks who are all acutely aware of the transformative possibilities of the Apple Vision Pro.
9. The release of the Blackmagic URSA Cine Immersive camera and the workflow around the capture and editing of AIV has made producing content for the Apple Vision much less onerous.