Yanxiang Zhang[1], Yi Qiu[2]

[1] University of Science and Technology of China

[2] University of Science and Technology of China



Web3D 2026

360° video is immersive - but still hard to analyze

Two problems appear when we move from watching to close inspection.

1. Interaction is still global

You can look around, but the whole video still follows one shared timeline. A small local event cannot be slowed down or scrubbed on its own.

2. Most pixels are often redundant

360° video needs high resolution, but in many analysis tasks only a small region is moving. The rest of the sphere may stay almost unchanged.

Example: frame-by-frame inspection of a craft worker's hand motion

We need local temporal control, while keeping the surrounding spatial context.



One change in the data structure unlocks two benefits

Spatio-temporal decoupling: store the background once, keep time only where it matters.



Static background
save once

Dynamic ROI
keep the timeline

Less data

Local time control

Key idea

The data structure is not only for compression. It creates a new interaction affordance.

Representation is interaction design.

Prior work solves pieces of the problem

The gap is at the representation level.

360° streaming

Viewport-adaptive methods save bandwidth, but the video still has one shared clock.

Interactive video

Rich temporal controls exist, but they usually operate on the whole 2D frame.

VR interaction

Modal interaction supports spatial tasks, but it does not create an independently timed local object.

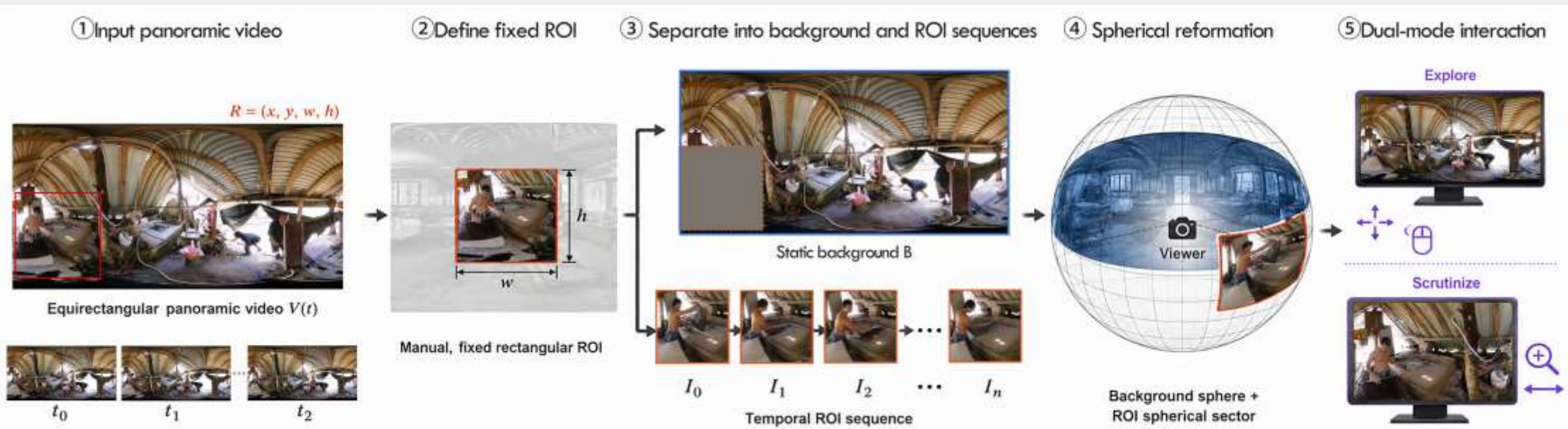


Explore-and-Scrutinize

separate content + local time control + web

From one 360° video to two layers

Offline separation; real-time interaction in a standard browser.



1 **Input panorama** 2 **Fixed ROI** 3 **Background + ROI video** 4 **Spherical mapping** 5 **Dual-mode interaction**

All-Intra ROI encoding keeps every frame directly reachable for frame-accurate scrubbing.

Explore space. Then scrutinize time.

Two modes separate spatial orientation from temporal analysis.



Explore = look around

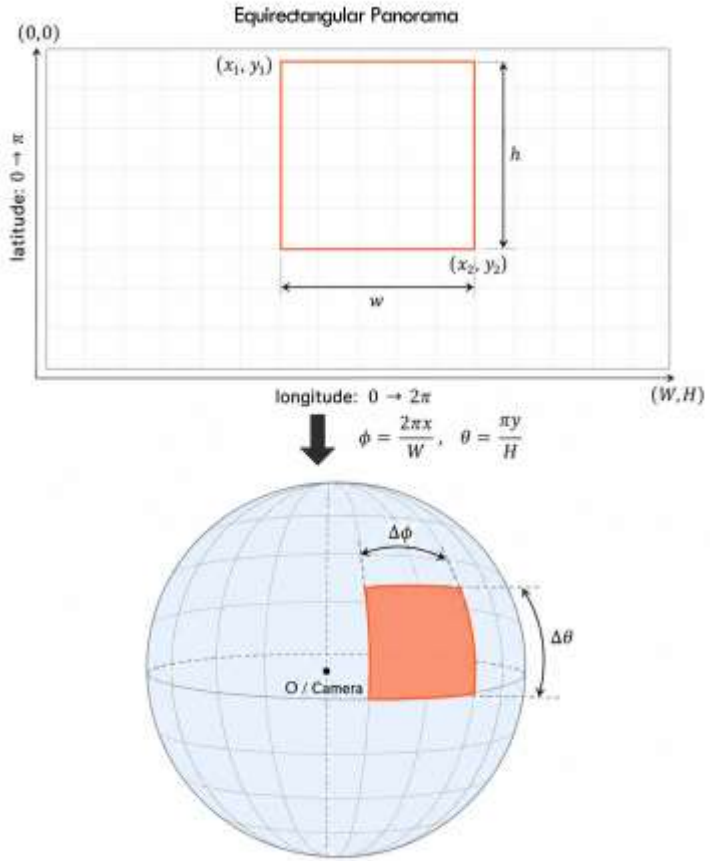
Scrutinize = move through ROI time

Context stays visible

Switch back anytime

The ROI stays inside the panorama

A spherical slice avoids the "floating video" feeling and visible seam distortion.



Why not use a flat video plane?

Flat plane

Different curvature -> visible seam and weak spatial alignment.

Spherical slice

Matches the background sphere in position, size, and curvature.

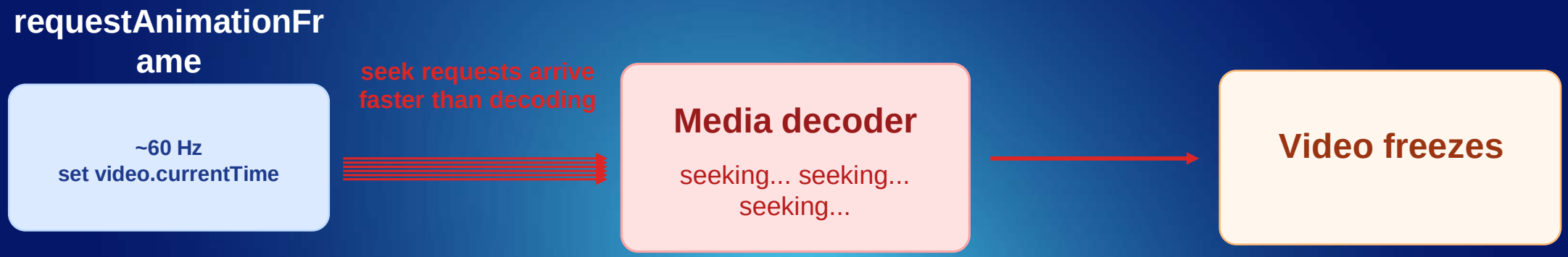
Result

The ROI can change in time while the surrounding scene remains stable.

pixel position -> longitude / latitude -> spherical sector

A browser problem we did not expect

Smooth UI updates can overwhelm the asynchronous media decoder.



Two failure modes

1 Decoder thrashing

New seeks interrupt unfinished seeks.

2 Micro-seeking

LERP gets closer forever, so tiny seeks never stop.

Let the decoder set the pace

Two-tier throttling: one outstanding seek, then stop inside a small deadzone.

Tier 1 - State flag throttling

Only send a new seek when video.seeking is false. The decoder becomes the clock for seek dispatch.

Tier 2 - Tolerance deadzone

Far away: ease toward the target. Close: snap to the target. Very close: stop seeking.

Measured result

17.3 ms → **2.6 ms**

mean seek latency

6.7×
lower mean latency

6.2 ms
P95 latency

~59 FPS
rendering stays smooth

LIVE Process

Explore -> Select ROI -> Scrutinize -> Explore

01

Look around

Find the work area

02

Select ROI

Enter Scrutinize

03

Scrub time

Move frame by frame

04

Go back

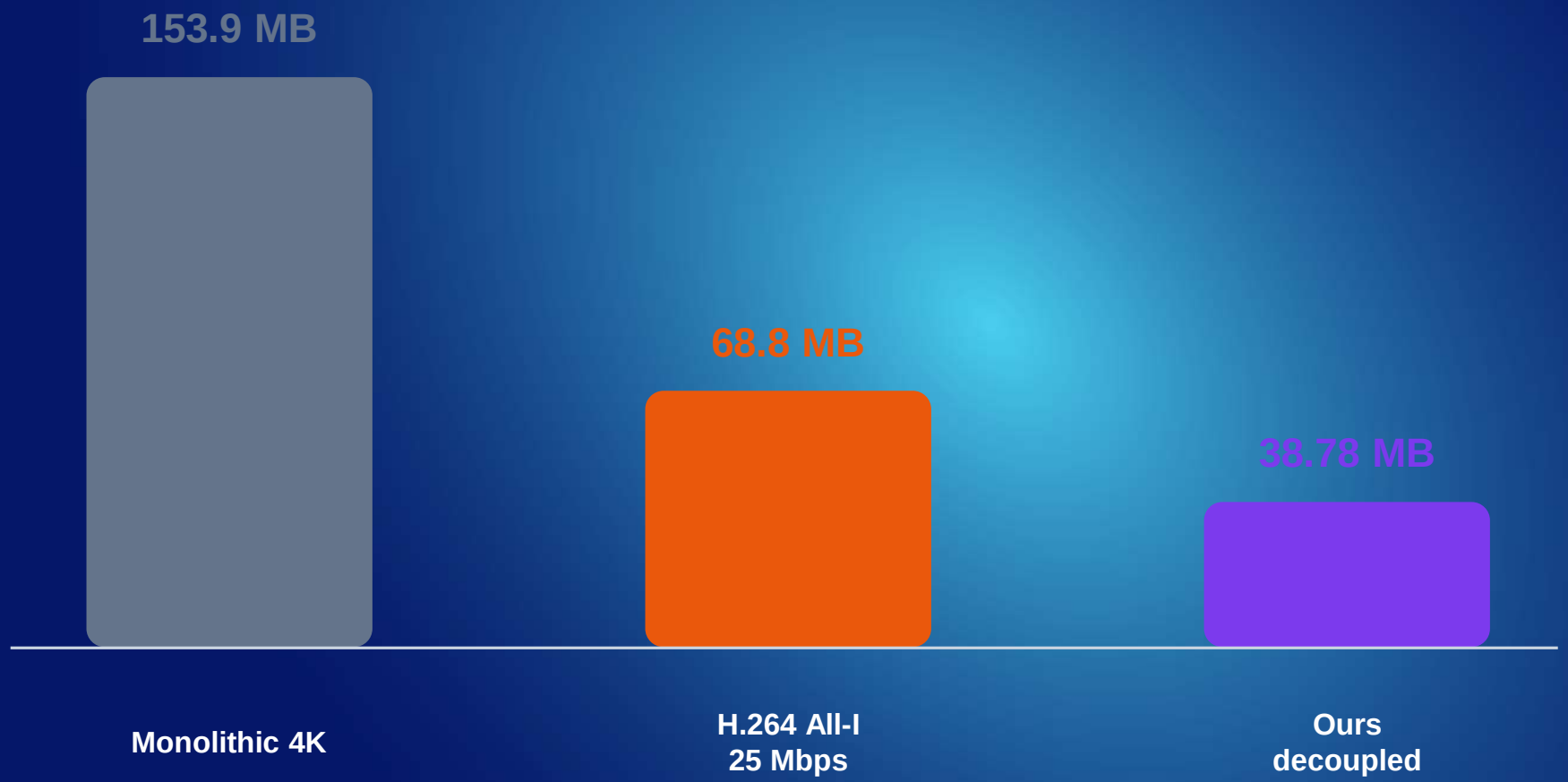
Continue exploring

Backup plan: use the recorded demo if the network is unstable.



Less data, while keeping random access

A fair codec comparison must respect the frame-accurate scrubbing constraint.



-74.7%

end-to-end data

-43.7%

vs. H.264 All-I

~15 s -> ~1 s

local-network load

Why All-I? Frame-accurate random access should not depend on decoding a long GOP first.

Users felt more control with low workload

Formative desktop study on a real 4K hand-papermaking task.

N = 12

participants

6F / 6M · ages 22-34

28.3

NASA-TLX

mean workload

78.5

SUS

mean usability score

Explore -> Scrutinize -> Explore

spontaneous workflow

“With the normal player I was fighting the timeline; with this, I was in charge of it.”

Common themes: temporal agency · cognitive relief · mode transition as a clear cue

Important: this is a formative study, not a controlled confirmatory experiment.

What we cannot do yet

The current design is intentionally narrow; the next step is to relax those assumptions.

Current limits

Single fixed ROI

one rectangular region

Mostly static background

moving cameras need stabilization/layers

Desktop user study

WebXR path not yet tested in-headset

Formative evaluation

small sample, no controlled counterbalancing

Next steps

Task-aware ROI detection

not only visual saliency

Multiple ROIs

selection + cross-ROI analysis

Hybrid cache + prefetch

scale to longer ROI sequences

Controlled headset study

domain experts + stronger statistics

broaden the content model, keep the same interaction principle

Representation can enable interaction.

By separating static context from dynamic ROI content, we get local temporal control and major data savings from the same design decision.

-74.7%
data footprint

6.7×
lower mean seek latency

78.5
SUS

Explore
↓
Scrutinize
↓
Explore

Thank you · Questions?

Backup - Positioning against prior work

What is unique is the combination of all five properties.

Approach	360°	Separation	Nonlinear time	Dual mode	Web
Viewport-adaptive streaming	Yes	No	No	No	Varies
Object-based immersive video	Yes	Yes	No	No	No
Video object segmentation	No	Yes	No	No	Desktop
TNIP	No	No	Yes	No	Desktop
Shot-orientation control	Yes	No	No	No	Yes
Explore-and-Scrutinize	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

The novelty is not any single column; it is the representation that makes the combination possible.



Backup - Seek latency details

Instrumented prototype: Chrome 132, RTX 3060 Ti, Windows 10.

Metric	Throttling ON	OFF	newA
Mean latency	**2.6 ms**	**17.3 ms**	**31.5 ms**
Median latency	1.7 ms	7.1 ms	32.5 ms
P95 latency	6.2 ms	33.3 ms	33.6 ms
Max latency	176.8 ms	116.1 ms	58.4 ms
Overflows >50 ms	5 (0.4%)	5 (0.5%)	1 (0.2%)
Severe >100 ms	3	1	0

Key point: throttling coalesces pending seeks and keeps queue depth at one.



Backup - Visual fidelity and data footprint

Representative slow, rapid, and complex-boundary conditions.



38.78 MB

our decoupled total

1.58 MB

background JPEG

37.2 MB

All-Intra ROI video

0.013 MB

web page



Thank You!

Web3D

2026

Doha, Qatar