

Web3D 2026

13-15 October 2026
Doha, Qatar

VisionSense3D

From physical objects to web-ready 3D assets: an affordable AI-assisted RGB-D scanning pipeline

Marwan Humaid, Ahmad Al-Moslemani, Hamad Aldous, Ejmen Al-Ubejdij, Umair Gavankar, Dr. Abdullatif Albaseer, Dr. Jens Schneider, Dr. Marco Agus

College of Science & Engineering, Hamad Bin Khalifa University

MOTIVATION

The bottleneck: object → web asset

Browser standards make textured 3D easy to **deploy**. Making the content from a real object is still the hard part.

Manual modeling

Needs 3D-authoring expertise and time.

Professional scanners

Costly, and hard to integrate into small institutions.

Who gets left out

Classrooms, museums, makerspaces, small labs and product teams.

EXISTING APPROACHES

Three routes, three blind spots

Approach	Strength	Weakness
Structure-from-Motion + Multi-View Stereo	Accurate under favorable conditions	Needs texture, stable lighting, view overlap and dense coverage
Consumer RGB-D sensors	Real metric observations	Noisy up close; poor on dark, reflective, transparent or fine surfaces
Generative 3D models	Visually complete meshes from sparse images	Arbitrary scale; may hallucinate unseen geometry

Their weaknesses are **complementary**, so we combine generative reconstruction with depth instead of picking one.

KEY IDEA

Key idea and contributions

RGB decides how it looks.
Depth decides how big it is.

CONTRIBUTION 1

An affordable end-to-end RGB-D + AI capture workflow from commodity hardware.

CONTRIBUTION 2

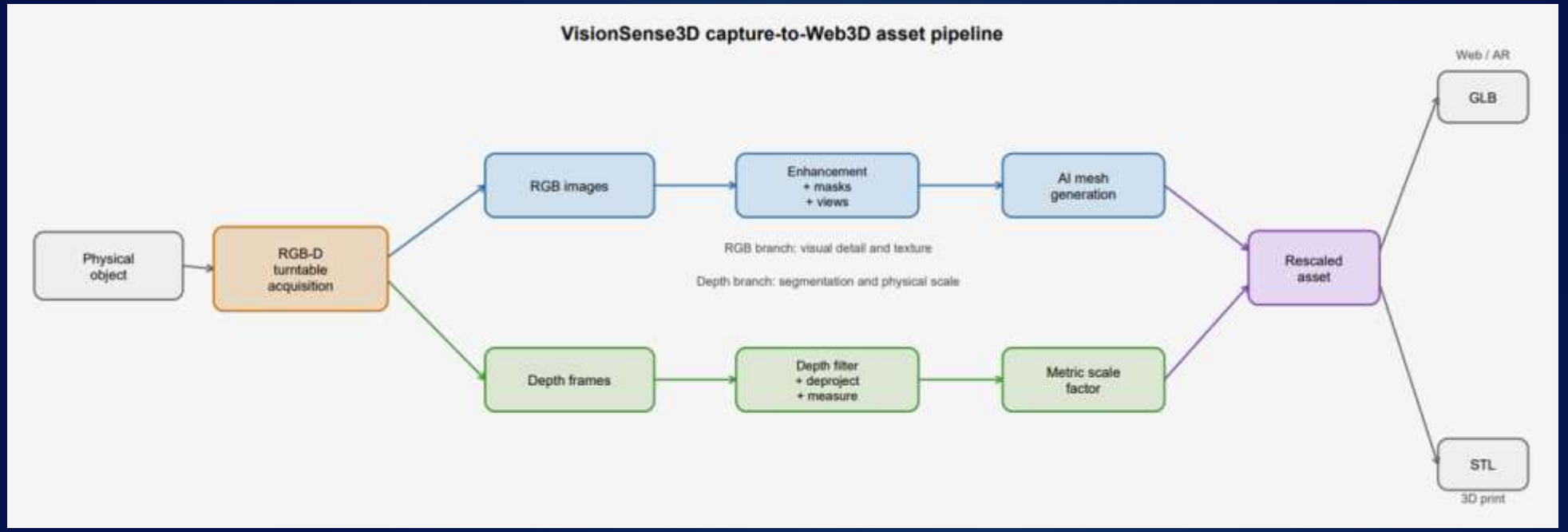
Depth-based global rescaling that keeps connectivity, normals, UVs, textures and materials.

CONTRIBUTION 3

A qualitative study of representative objects, including failure cases.

SYSTEM OVERVIEW

Two branches, one asset



- **RGB branch** - appearance and texture
- **Depth branch** - segmentation and physical scale

ACQUISITION

Fixed sensors, moving object

The rig - all off-the-shelf

- **RGB camera** - appearance
- **Intel RealSense** - depth
- **Motorized turntable** - regular angle steps
- **Diffuse, fixed lighting** - fewer shadows

Why this design

Sensors never move → **repeatable azimuth views** over a full revolution.

Object near the center → camera distance **barely changes**.

Depth only masks and measures - it is **never fused** into the surface.

Limit: one horizontal orbit misses the underside and weakly constrains the top, but the generative model infers those.

PROCESSING

From pixels to a physical size

1 · RGB

Enhance

CLAHE exposes weak local contrast.

2 · RGB

Mask

Background removal gives the object hypothesis.

3 · DEPTH

Filter

Histogram the masked depths; keep the dominant interval.

4 · DEPTH

Deproject

Camera intrinsics → points in physical units.

Neither cue alone: color proposes the object; depth removes similarly colored background.

Reference size: the vertical extent h_s, visible in every turntable view.

METRIC SCALING

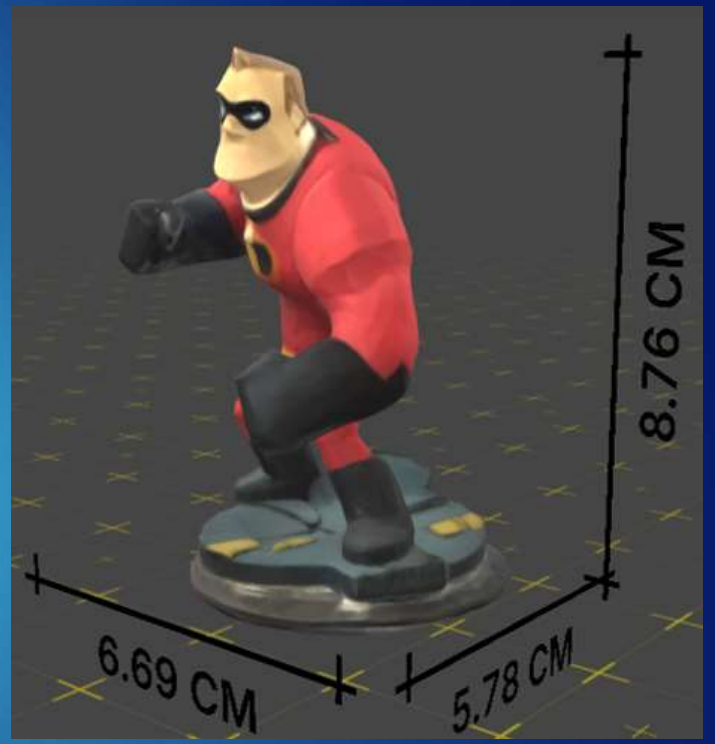
One number rescales the mesh

$$\alpha = \frac{h_s}{h_m}$$

h_s = height measured from depth (physical units)
h_m = height of the generated mesh (arbitrary units)
Every vertex is multiplied by **α**.

UNIFORM TRANSFORM → PRESERVED

- connectivity
- normals
- UV coordinates
- textures
- materials
- proportions



Depth-scaled mesh in physical units (Figure 4, left)

OPERATION MODES

Three ways to use it

Single Image

One RGB view, optional depth for scale. Fast; the back is inferred.

Multi View

Turntable views cut ambiguity. Slower, more consistent.

Studio

Capture, masking, depth view, reconstruction and export in one UI.



Integrated Studio interface (Figure 4, right)

Compact shapes reconstruct best





Buddha head - coherent

Compact and near-symmetric → stable silhouettes.

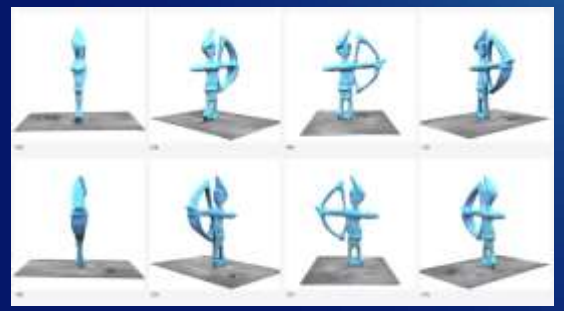




Articulated figurine - recognizable

Overall form holds, with some smoothing.





Archer - hardest

Thin, elongated parts → sensitive to mask errors and generative priors.

Shallow relief: the opposite extreme





EASY

Broad, simple silhouette.

HARD

Relief is finer than this low-cost workflow resolves.

USE IT FOR

Interactive presentation - not relief analysis.

RESULTS · LESSONS

Where it works, where it breaks

Works best with **clear silhouettes**, **moderate surface texture** and **stable illumination**.

FAILURE SOURCE 1 · DEPTH

Missing or noisy depth

Destabilizes segmentation and the physical size.

FAILURE SOURCE 2 · RGB

Ambiguous RGB evidence

The generative model smooths details or invents geometry.

The pipeline improves **deployability**, it does not turn a plausible reconstruction into a ground-truth scan.

SCOPE

What we claim, and what we don’t

We claim

- Recognizable, approximately scaled assets
- An integrated route from capture to GLB / STL
- Scale transfer that leaves appearance untouched

Not yet claimed

- Metrology accuracy - no error distribution yet
- Fixing local hallucinations - α is global
- Proven usability - no formal user study yet

CONCLUSION

Takeaways and next steps

The contribution is the **integrated route** from a physical object to a deployable Web3D asset, not a new reconstruction algorithm.

Low-cost RGB-D capture + generative reconstruction + depth-assisted global scaling + GLB / STL export.

Future work

- Compare against classical, AI-only and manual-scaling baselines
- Quantify repeated dimensions and deviation vs. reference scans
- Report cost, runtime, mesh size and mobile WebXR loading
- Better segmentation and elevated views for unseen surfaces

Thank You!

Questions?

SYSTEM DEMO

bit.ly/4uyCgUV

PAPER

doi.org/10.1145/3822516.3837793