Volumetric Motion Rigging

Volumetric Motion Rigging (VMR): a practical workflow for recovering motion from reconstructed meshes, rebuilding difficult surface detail, and preserving the quality of the final animation.

Author & contact: redazul on GitHub

Experimental assistance: GPT-6-Astra assisted during selected experimental runs. Research direction, final result selection, and Unity validation were provided by redazul.

Project on Hugging Face ↗ Download paper PDF ↗ Animated character ↓

INPUT VIDEO

The source performance, cropped to a square for comparison. Its 52-frame timing is unchanged.

FINAL ANIMATED RESULT

The character animation produced by Volumetric Motion Rigging (VMR).

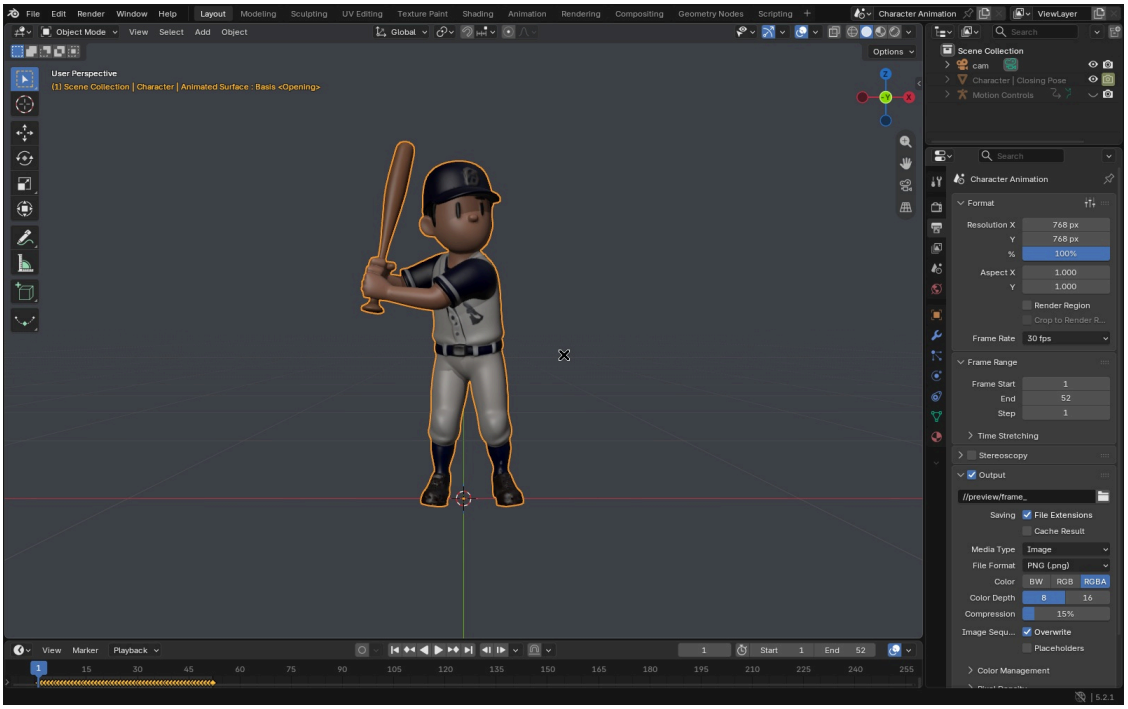
The two videos share playback controls. Normal playback is 30 fps; slow motion changes playback speed rather than adding poses. Open input video · Open final animation

---

INSIDE THE 3D APPLICATION

The character on Blender’s animation timeline

The result also exists as a 3D scene that can be opened, inspected from other angles, scrubbed and played in Blender. The view below shows the actual character in Blender’s workspace, with its timeline and pose changes visible.



Actual Blender workspace captures of all 52 timeline poses, replayed at 30 fps. The interface is not a mock-up. Capture was performed pose by pose, so this is not a real-time viewport-performance benchmark.

Download the Blender character animation Open workspace video

In Blender, use the timeline or press Space to play. The VMR asset uses baked mesh poses and retains a separate closing checkpoint. The FBX download below carries this motion as animated blend shapes; it is not a Humanoid retargeting rig.

DOWNLOAD THE ANIMATED CHARACTER

VMR character for Unity

The FBX contains the VMR character, embedded textures and animation for all 52 stored poses. Motion is exported as animated blend shapes.

Download animated FBX (155 MB)

Validated in Unity. The author confirmed that the FBX works in Unity and its animation is available. Separately, fresh FBX reimport reproduces all 52 stored poses within 4.85×10^{-8} scene units. Read the validation record.

Synchronized source video, reference meshes and VMR result.

01 / OVERVIEW

Recover motion without losing the character

> A video can be reconstructed into a sequence of 3D meshes, yet that sequence may still be unsuitable as a stable animated character. Independently reconstructed frames can change topology, surface detail and proportions. Playing them as a mesh flipbook reproduces the motion but can also reproduce visible identity changes.
>
> This work separates the problem into a coarse motion representation and a detailed surface representation. Simplified interior volumes guide the motion of a reference character. Local surface fitting corrects broad shape, while difficult regions such as a two-handed grip are reconstructed from actual reference geometry when deformation alone fails to preserve their detail. Volume constraints and endpoint locks then protect the reviewed result.
>
> The worked example is a 52-pose baseball swing. The VMR output was preferred for its lower visible jitter over tested framewise registration refinements and a bone-compressed approximation. The method produces a playable Blender animation, but a separate closing surface remains at the final pose. The broader goal of one continuous character surface over the entire sequence is therefore not yet fulfilled.

What is being reconstructed?

The input consists of the original video and a selected sequence of reconstructed 3D reference meshes. The demonstration uses TRELLIS.2-derived references, with corrected closing poses. The video supplies the observed performance; the mesh sequence supplies geometric fitting targets. Those targets are reconstructed estimates, not perfect ground truth.

The desired output is a detailed character whose fingers, clothing, bat and body remain recognizable as the motion progresses. Different triangulations in the reference frames are tolerable if correspondence can be recovered. Actual changes in visible surface connectivity or incompatible endpoint shapes are a harder problem and cannot be made continuous merely by keeping a constant vertex count.

Represent motion coarsely, preserve detail explicitly

Prepare

Align the video-derived meshes and select a detailed reference.

Track

Fit consistent body volumes through the sequence.

Reconstruct

Correct surfaces and recover difficult reference detail.

Validate

Check motion, local shape, endpoints and playback.

1. Establish a consistent geometric reference

Extract the source poses, inspect the reconstructed meshes, and place selected targets in a common coordinate system. Separate whole-object alignment errors from genuine motion before fitting the character. The demonstration uses 52 selected poses, with one output pose per reference frame. A detailed starting mesh supplies the character identity; independently reconstructed targets describe how that character should move.

2. Build a volumetric motion representation

The reference interior is voxelized and divided into a consistent set of regions using geodesic farthest-point seeds and Voronoi assignment. Distances through the solid interior define influence weights. In the demonstrated implementation, 64 regions guide the surface and weights are normalized to four influences per original vertex. The ellipsoids used to visualize the regions are approximations of these volumes, rather than ordinary Blender bones made visually thicker.

Each region carries rotation, translation and tightly regularized anisotropic scale. A surface point follows a weighted mixture of those transforms:

$$x_t = \Sigma_k \; w_k \; [\; A_{k,t}(x_0 - g_k) + g_k + t_{k,t} \;]$$

x_0 is the reference point, g a region center, A its rotation and scale, t its translation, and w its influence. The expression describes the motion representation; it does not include every later surface correction.

Because the controls have consistent identities, the target meshes need not share vertex indices. The motion fit is guided: mesh-derived skeleton motion supplies initialization and a rotation prior. Surface positions and normals, neighboring-region coherence and noncyclic

temporal refinement constrain the result. The head and bat receive rigid treatment to avoid destroying their recognizable shape.

3. Correct the surface without treating distance as the only objective

A close geometric fit can still produce a collapsed sleeve, a flat wrist or an implausible palm. Local shape-preserving correction limits stretch and compression while boundary constraints retain nearby accepted geometry. Particular care is needed where surfaces touch: a nearest-point match can attach a sleeve to a torso or one hand to the other.

Interpolation must also respect the kind of motion. Blending vertex positions across a large rotation can shorten a rigid object severely. Rotational treatment is therefore appropriate for the bat and other rigid regions. The current review format uses the original 52 stored poses; this observation is not a claim that arbitrary intermediate times or every possible export have been solved.

4. Introduce reference detail where deformation is insufficient

If repeated fitting destroys a feature instead of recovering it, obtain that feature from actual reference geometry and document the topology change. This is especially relevant to fingers around a handle, where small geometric and correspondence errors have a large visual effect. The reference-derived region must then be attached and transported coherently, rather than swapped indiscriminately between frames.

5. Preserve the reviewed motion and endpoints

Hold accepted regions and endpoint poses fixed while repairing localized defects. For the demonstrated late-wrist repair, only four stored poses changed; the other 48 were retained. The forearm and hand were fitted together using surface attraction, local shape constraints and internal distances supporting wrist thickness. This joint regional fit should not be confused with a new optimization over all time steps.

$$E \approx \lambda_{\text{fit}} E_{\text{reference}} + \lambda_{\text{shape}} E_{\text{local shape}} + \lambda_{\text{volume}} E_{\text{internal distances}} + \lambda_{\text{prior}} E_{\text{pose retention}}$$

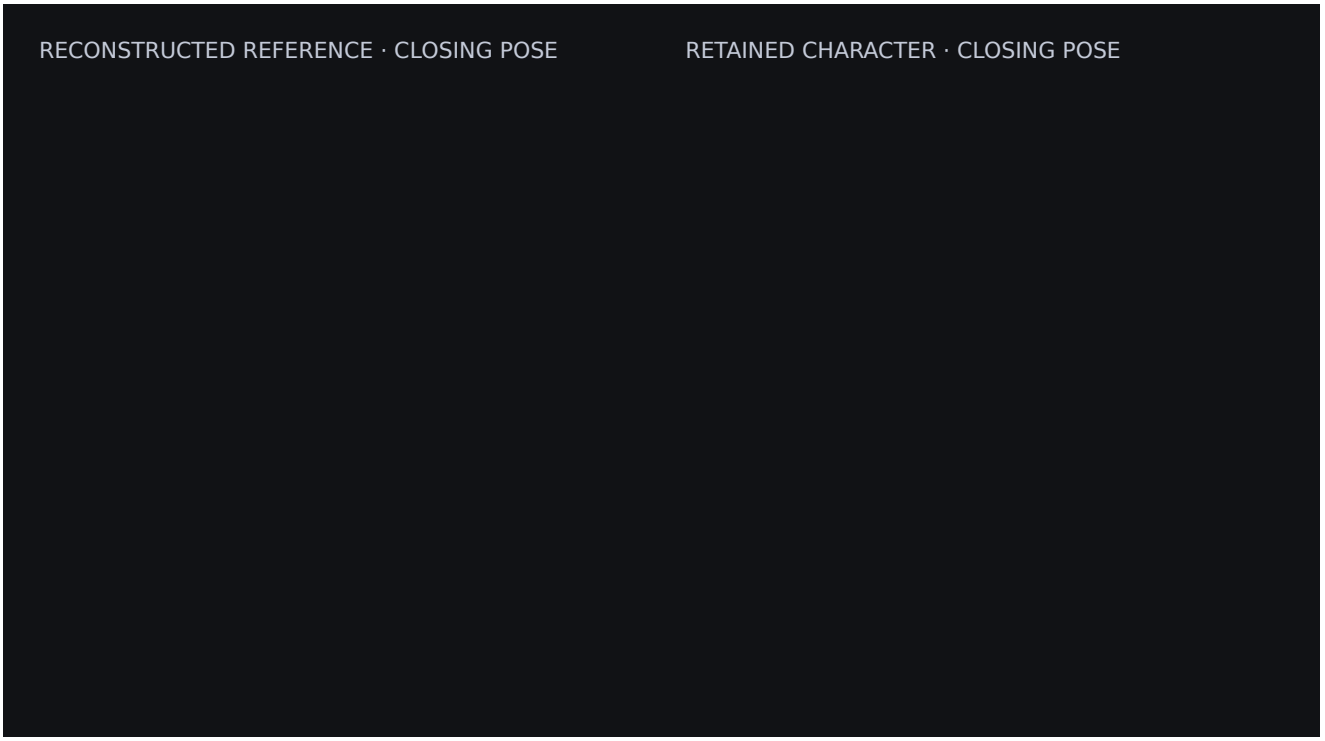
An explanatory summary of the late repair. Reference attraction is balanced against shape, thickness and retention of the existing pose. The implemented solver uses local/global updates and explicit boundary constraints.

03 / RECOVERING THE FINAL FRAME

Reconstruct the grip; retain the repaired body

The closing frame was the hardest local reconstruction in the example. Deformation-only fits could move the hands toward the target but also erase finger detail or distort the wrist. The successful closing reconstruction combined the repaired body with grip and bat geometry derived from the actual closing reference.

The reconstruction retained **41,058 body vertices** and added **10,250 grip and bat vertices**. Within the added region, **9,484 core vertices exactly matched their closing-reference coordinates** in the stored pose. Short wrist overlaps joined the region to the forearms. This exactness applies to the identified reference core, not to the whole character or to every image pixel.



The closing result combines retained character geometry with recovered grip detail. It is not a wholesale copy of the independent reference mesh. Wrist attachments use overlapping surfaces and are not certified watertight.

Carry the opening grip through the motion

Matching the closing hand configuration did not automatically produce the desired opening fingers, palm and exposed bat knob. A reference-derived opening grip was therefore established separately. Its 6,697-vertex hand/grip/bat region was transported as a rigid group through the first 51 poses, following tracked shaft and wrist motion. Forearm cross-sections were transported toward the wrists, and a screened displacement joined the boundaries.

This retained the same opening hand detail as the swing progressed and removed a discontinuity immediately after the first pose. It also introduced a constraint: the grip could look stiff and could differ from the late reference shape. The VMR result preserves the separately approved closing checkpoint at the final pose.

Current limitation. The VMR scene uses one visible character surface through poses 1–51 and a separate approved closing surface at pose 52. Exactly one is visible at a time. That is a preserved endpoint, not continuous surface identity across the entire sequence. A connected-surface alternative was built, but did not satisfy the same appearance requirements.

Preserve wrist thickness during local fitting

The final local repair constrained internal wrist distances while fitting the forearms and hands to the late references. Independent section measurements retained approximately **93–110% of the pristine wrist dimensions** across the sampled sections. These measurements address pinching and flattening; they do not certify anatomical accuracy, manifoldness or freedom from all self-intersections. Validation summary.

04 / EVALUATION

Lower geometric error does not guarantee smoother motion

Two open-source non-rigid registration methods were tested as refinements of the VMR animation: the Amberg method and the Sumner/Popovic correspondence method implemented in Trimesh. Both used the existing pose initialization, an 830-vertex proxy, 24,000 area samples per target, the same displacement transfer, protected regions and a maximum displacement of 0.008 scene units. Their solver regularization differed, so this was a comparison of practical configurations rather than an exhaustive method ranking.

Lower values indicate closer agreement with sampled target geometry. These are one-way distances on a shared body subset over poses 2–51, and the samples participate in fitting. They do not measure semantic correspondence, perceived jitter or held-out accuracy.

VISUAL REVIEW FINDING

The VMR was judged to have the least visible jitter among the compared outputs. The registration refinements reduced the reported surface-fitting error but did not supersede that motion-quality preference. This was a qualitative review, not a blinded study or a measured jitter benchmark.

Framewise fitting can follow small inconsistencies in independently reconstructed targets. Retaining more of an already coherent animation may therefore be preferable to repeatedly pulling its surface closer to each target. That is a plausible explanation of the observation; the individual contributions of source noise, fitting and compression were not isolated.

Inspect the refinement and compression outputs

These variants are provided as supporting comparisons, not as replacements for the VMR result shown at the top.

What skinning decomposition established

The actual EA Dem Bones library was used to compress a corresponding portion of the registered animation. Known wrist seams shared vertices and weights during decomposition. A 47,581-vertex surface across 51 poses was approximated with **64 bones and up to four nonnegative influences per vertex**. This produced actual skin weights and transforms, not a relabeling of the earlier volume controls.

Raw RMS error against the registered input was **0.003779 scene units**. At the last fitted pose, RMS error was approximately 0.009017 and maximum vertex error 0.039843. The native Blender armature matched the baked linear-blend reconstruction within approximately

2.73×10^{-7} scene units across the fitted poses. These are implementation and compression checks, not evidence that the compressed character is visually superior.

The final closing checkpoint was excluded from decomposition because it belongs to a different surface. Consequently, compression did not solve raw changing-topology correspondence or remove the closing handoff. A standard skeletal representation can be recovered for corresponding motion, but it cannot repair incorrect or missing correspondence by itself. Numerical evidence and library revisions.

05 / VALIDATION AND PLAYBACK

A playable scene is more than a rendered clip

The downloadable Blender scene contains the VMR character geometry, materials and timeline animation. A reader can open it, inspect the mesh in a 3D viewport, change the view, scrub individual poses and play the timeline. The VMR animation uses baked mesh poses; it should not be described as if the entire result were driven by the separately tested 64-bone approximation.

The workspace demonstration is taken from a presentation copy of that scene. Object labels, viewport framing, timeline markers and the display of reference controls were adjusted for clarity. It is a screen capture of the actual application, not a drawing of a Blender interface. Its 52 poses are captured individually and replayed at the clip’s frame rate, so it demonstrates scene content and timeline playback rather than hardware performance.

WHAT WAS CHECKED	WHAT IT ESTABLISHES	WHAT IT DOES NOT ESTABLISH
Stored and evaluated positions	Exported poses agree with the stored geometry; protected samples remain unchanged.	Correct anatomy or absence of every self-intersection.
Opening and closing checkpoints	Both endpoints pass five fixed camera checks each.	Continuous surface identity between them.
Wrist section dimensions	Measured local thickness survives the repair.	Watertight or universally valid attachments.
Frame counts and timing	52 poses; 30 fps normal playback; slower playback without added poses.	An objective score for smoothness.
Native armature comparison	The bone approximation reproduces its computed linear-blend result.	Faithful gameplay retargeting or improvement over the VMR animation.

Endpoint review uses immutable source and image hashes. A small fixed image tolerance allows at most 100 channels to differ by one 8-bit value in a camera image. Where fresh rendering produced small differences on unchanged geometry, complete evaluated render-input signatures were checked before reusing an earlier unchanged render. Newly changed poses were freshly rendered, fresh comparison images were retained, and no frozen oracle image or tolerance was replaced to make a candidate pass.

▸ Download the evidence behind the demonstration

06 / CONTRIBUTION AND LIMITS

An implemented workflow, with a clearly bounded result

The practical contribution is the combination of guided volumetric motion, reference-based detail reconstruction, localized shape and thickness constraints, endpoint preservation and

visual review. We use **Volumetric Motion Rigging** as a descriptive name for this workflow. The work does not establish a new class of deformation algorithm.

Its building blocks have precedents. Embedded Deformation blends local transformations organized in a graph; geodesic voxel binding computes influence through interior distances; skinning decomposition extracts weights and transforms from corresponding example poses. The demonstrated workflow applies related ideas to an inconsistent reconstructed mesh sequence and records where additional reconstruction is needed. [1–3]

- **Closing continuity remains unresolved.** The final pose uses a separately preserved surface.
- **Contact remains difficult.** Fingers, cuffs and rigidly transported grip regions can still differ from the reconstructed reference.
- **Playback is not game-readiness.** The FBX preserves baked animation, but does not certify a single skinned runtime mesh, Humanoid retargetability, collision behavior or runtime performance.
- **The evidence is case-specific.** The reported measurements and reviewer preference come from one sequence and the tested configurations.

What a stronger next result should demonstrate

A future revision should preserve the reviewed motion while solving correspondence over time, especially through contact. It should compare surface fit, local distortion and temporal stability together, with approved endpoints held fixed. A temporal metric must distinguish genuine rapid motion from unwanted jitter. Meeting those requirements with the same visible surface across the complete sequence would address the principal remaining limitation.

07 / TOWARD A LEARNED MODEL

From a reconstruction workflow to video-conditioned generation

VMR could provide the training pipeline for a system that receives a video and produces a character with editable animation. The present result establishes a reconstruction workflow and an animated FBX validated in Unity. A trained VMR model that generalizes to new videos has not yet been demonstrated. The architecture and training program below are proposed next steps.

Separate character identity from motion

A proposed system would estimate character appearance and camera motion from the video, recover one reference character, and predict its deformation across time. The output would include geometry, materials, a rig or blend shapes, and animation curves. Maintaining a consistent character representation is central to avoiding frame-to-frame changes in proportions and surface detail.

Video understanding

Separate identity, camera movement and observed motion.

Character representation

Recover a reference surface and deformation controls.

Temporal prediction

Predict motion and local corrections consistently over time.

Animation export

Fit editable curves and validate the animated FBX.

Recovering the motion shown in a video can begin with supervised prediction. Generating hidden geometry or alternative plausible motions introduces uncertainty that a conditional generative model could represent. A useful product can combine pretrained generation, learned prediction and geometric fitting; it does not require every stage to be one newly trained network.

What the existing meshes can teach

AVAILABLE EVIDENCE	POTENTIAL TRAINING SIGNAL
Video frames and reconstructed targets	Appearance-to-geometry relationships, with estimated geometry labeled as uncertain.
VMR controls and fitted poses	Motion and deformation prediction conditioned on the reference character.
Surfaces before and after grip or wrist repair	Local correction targets for specific failure modes.
Protected endpoints and contact regions	Constraints that preserve identity and important geometric detail.
Competing outputs and quality judgments	Preference supervision, once sufficiently varied comparisons are collected.
Validated animated exports	Output structure, timing and numerical checks for playable assets.

Mesh volume is not the same as dataset diversity. Many meshes from one swing still describe one character performing one motion. Intermediate failures should be retained with failure labels, rather than mixed into accepted targets. Dataset records should connect each video, pose, reference mesh, correction, coordinate system, correspondence mapping and validation result. The separate closing surface must be labeled explicitly; it is not a continuous-deformation training target.

Use VMR as a teacher first

The first proposed learning task is bounded: given a reference character and a short video segment, predict temporally consistent VMR controls and local surface corrections. The existing fitting pipeline would produce reviewed supervision and could refine predictions during early deployment. This allows direct comparison of fitting time, geometric error, contact preservation and temporal consistency before attempting a complete generative system.

Training would combine synthetic videos rendered from known animated assets, which provide exact cameras and geometry, with varied videos processed and reviewed through VMR. Entire characters and source motions should be held out for evaluation. Adjacent frames, alternative camera renders and revised outputs from the same source should remain in the same split to avoid misleading generalization results. Learning curves would determine how much additional data is needed.

A later system could compress motion into a learned latent representation and train a video-conditioned predictor or generative model over that representation. Training objectives should balance video agreement, geometric fit, temporal consistency, contact and local volume. Temporal constraints must preserve genuine fast motion rather than simply smoothing it away.

Build on existing generative components

TRELLIS.2 provides an image-conditioned 3D generation foundation. AniGen generates shape, skeleton and skinning from an image and provides an additional monocular-video fitting pipeline. These are candidates for evaluation or integration; their performance on this case's grip, contact and endpoint requirements has not been established here.

Editable keyframes would be a further output stage. Dense recovered motion could be reduced to animation curves within a chosen reconstruction tolerance, while preserving endpoints, contact events and rapid movements. The current baked animation demonstrates playback; it does not yet demonstrate sparse, artist-friendly skeletal keyframes.

08 / REFERENCES

Published methods and open-source implementations

1. Sumner, Schmid & Pauly (2007). Embedded Deformation for Shape Manipulation. Local transformations and graph coherence for detailed deformation.

2. Dionne & de Lasa (2013). Geodesic Voxel Binding for Production Character Meshes. Interior-distance skinning influence.

3. Le & Deng (2012). Smooth Skinning Decomposition with Rigid Bones. Compact linear-blend skinning recovered from example poses.

4. Trimesh non-rigid registration documentation. Implementations of the Sumner/Popovic correspondence method and Amberg non-rigid ICP; the demonstration used Trimesh 5.1.0.

5. EA Dem Bones. The actual decomposition library used in the comparison. Exact source revisions are included in the downloadable technical evidence.

Published references support the method context. Sequence-specific results are drawn from saved implementation checks, rendered comparisons and qualitative review. No general superiority or scientific novelty is claimed.

Technical white paper · 9 September 2026 · Author & contact: redazul.
The input/result demonstration and Blender scene describe the same VMR animation. The final closing surface handoff remains disclosed. Videos and figures accompany this publication. The animated FBX and Blender scene are available from the VMR repository.