PAN-EUROPEAN UNIVERSITY

Report no. 2.2

Guidelines for creation of intelligent avatars

Version: 1.0

Language: English

Project name: **Intelligent avatars for digital heritage applications**

Project acronym: **IntAvaVR**

Project no: **09I03-03-V04-00495**

Authors:

Lead researcher: **Prof. Dr. Selma Rizvić**

Other authors: **RNDr. Ján Lacko, PhD., doc. RNDr. Eugen Ružický, CSc., Ing. Juraj Štefanovič, PhD.**

Date: **30.8.2025**

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

Guidelines for creation of intelligent avatars

Creating 2D intelligent AI-based avatars in virtual reality (VR) applications requires careful consideration of both technical and user experience guidelines. First, the avatar's visual design must be consistent with the VR environment and the application's purpose. This includes selecting appropriate art styles, color schemes, and animations that ensure the avatar appears engaging and expressive without causing visual fatigue. Developers should also consider scalability, ensuring avatars can perform a wide range of expressions and gestures while remaining lightweight enough to run smoothly on different VR hardware. Optimization techniques, such as sprite sheets and vector-based animations, can help maintain high performance without compromising visual quality.

Equally important are the AI-driven behavioral and interaction guidelines. Avatars should possess adaptive intelligence capable of responding to user inputs, environmental cues, and contextual scenarios in real-time. This involves implementing natural language processing for verbal communication, emotion recognition to adjust expressions, and procedural behavior models that guide autonomous actions. To maintain immersion, interactions should feel intuitive and human-like, with responses that are contextually appropriate and timely. Developers should also ensure that the AI respects ethical and privacy considerations, avoiding behaviors that may seem intrusive or manipulative to users.

Usability and accessibility guidelines are crucial for maximizing the effectiveness of 2D AI avatars in VR. Developers should design avatars to accommodate diverse user needs, including options for adjusting visual clarity, speech output, and interaction complexity. Testing with a broad user base helps identify potential issues in communication, comprehension, and emotional engagement. Furthermore, avatars should support modularity, allowing developers to update or customize behaviors and appearance over time without disrupting the user experience. By combining visual coherence, intelligent responsiveness, and inclusive design, developers can create 2D AI avatars that enhance immersion and foster meaningful interactions within VR applications.

1. 2D Data acquisition

Step 1: Define Data Requirements

Clearly specify the objectives of the video dataset, including the types of actions, gestures, or expressions to capture. Determine the required resolution, frame rate, and duration of recordings. Consider factors such as the target application (e.g., AI training for avatar motion or gesture recognition) and the diversity of subjects, backgrounds, and lighting conditions needed to ensure robust model generalization.

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

Step 2: Setup of Green Screen Environment

Prepare a controlled environment with a uniform green backdrop. Ensure the green screen is evenly lit to minimize shadows, reflections, and color spill. Use diffused lighting from multiple angles to achieve consistent illumination across the subject and the background. Calibrate the camera to maintain accurate color capture and avoid motion blur or frame drops.

Step 3: Subject Preparation

Ensure subjects wear clothing and accessories that do not match the green screen color to prevent chroma keying errors. Instruct subjects on desired movements, gestures, or expressions according to the dataset plan. Maintain consistent distance from the camera and the green screen to avoid perspective distortion.

Step 4: Video Capture

Record the video at the pre-determined resolution and frame rate. Use multiple takes if necessary to capture variations in motion or expression. Monitor real-time video for issues such as shadows, color spill, or occlusions that could interfere with masking.

Step 5: Masking and Background Removal

Apply chroma keying to isolate the subject from the green screen. Generate accurate masks for each frame using automated or semi-automated methods, followed by manual refinement to handle complex edges (e.g., hair, clothing folds). Store the resulting foreground masks along with the original videos for downstream processing.

Step 6: Quality Control and Annotation

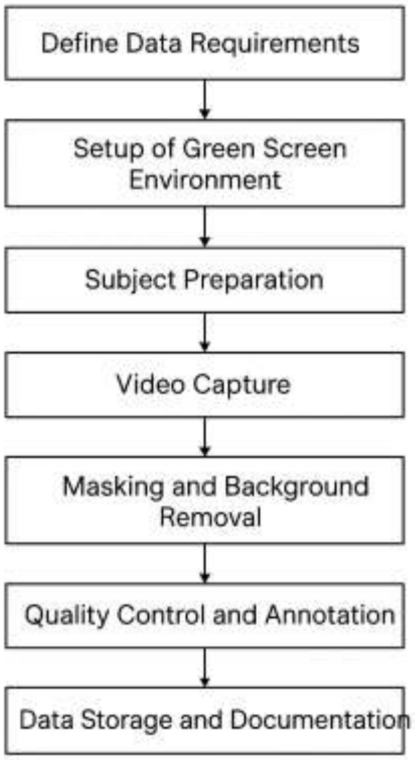
Inspect the masked videos for artifacts, inconsistencies, or missing data. Annotate frames with relevant metadata, including subject ID, action type, or temporal markers. Ensure the dataset meets statistical diversity requirements in terms of subjects, motions, and lighting conditions.

Step 7: Data Storage and Documentation

Organize the processed videos and masks systematically with clear file naming conventions. Document acquisition conditions, camera settings, lighting setup, and masking procedures to facilitate reproducibility and model training transparency.

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VAIA | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."


Define Data Requirements
Setup of Green Screen Environment
Subject Preparation
Video Capture
Masking and Background Removal
Quality Control and Annotation
Data Storage and Documentation

1.1 Scene setup

Green Screen Setup:

A flat, seamless green screen backdrop was used, positioned vertically behind the subject. The surface was stretched tightly to eliminate wrinkles and texture that could interfere with chroma keying. The green screen extended beyond the subject's movement boundaries both horizontally and vertically to ensure full coverage during recording.

- A 3 m (width) × 2.5 m (height) chroma green backdrop was mounted on a support frame, placed 1.5 m behind the subject.
- The bottom edge of the screen extended to the floor to allow full-body capture without floor reflections.
- The green surface was illuminated separately to maintain consistent luminance between 40–50% on the waveform monitor, avoiding hotspots or shadows.

Camera Position:

A single high-resolution digital video camera was placed at eye-level with the subject, centered along the frontal axis. The distance between the camera and the subject was set to approximately 2.5–3 meters, ensuring a clear field of view that captured the subject's full body without distortion.

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

The focal length was adjusted to minimize lens warping, and the camera was stabilized on a tripod to maintain consistent framing throughout the capture session.

- Camera: Single 4K digital video camera (e.g., DSLR or mirrorless with HDMI output).
- Tripod height: 1.6 m (aligned with the subject's eye level).
- Distance from subject: 2.8 m to achieve a full-body frame with a 50 mm focal length lens (minimizing perspective distortion).
- Angle: 0° (directly facing subject, perpendicular to green screen plane).
- Frame rate & resolution: 30 fps at 3840×2160 pixels.
- White balance: Fixed at 5600 K to match daylight-balanced lights.
- Shutter speed: 1/60 s, Aperture: f/4 for depth of field sufficient to keep subject sharp while slightly softening the green screen.

Lighting Arrangement:

A three-point lighting system was employed to ensure uniform subject illumination and minimize shadows on the green screen. Two softbox key lights were positioned at 45-degree angles to the subject from the front-left and front-right sides, providing even frontal illumination. A fill light was placed directly in front at a lower intensity to reduce harsh shadows on the face and body. Additionally, two backlights were aimed at the green screen to ensure even chroma saturation, and a rim light was placed behind the subject to separate the subject from the background and reduce color spill. All lights were diffused to prevent glare and maintain natural color rendering.

- Key Lights: Two softboxes with LED panels, each 100 W, placed at 45° angles to the subject (left and right), 2 m in front, at a height of 1.8 m, tilted downward at ~30°.
- Fill Light: 50 W LED panel with diffuser, directly in front of the subject, 2.5 m away, at chest height, providing ~50% intensity of the key lights to reduce shadow contrast.
- Back/Rim Light: A 60 W LED spotlight behind the subject, elevated at 2.2 m, pointing downward at 45°, creating a light outline around hair and shoulders to separate subject from the screen.
- Background Lights: Two 80 W LED panels directed at the green screen from each side, positioned 1 m from the screen edges, at a 45° inward angle. Light intensity adjusted to achieve uniform chroma saturation without exceeding 60% luminance.

1.2 Makeup and costumes

Costumes and Clothing Selection

Costume design plays a critical role in ensuring accurate chroma keying and high-quality video capture. Subjects must avoid clothing that matches or contains shades close to the green screen, as this produces transparency artifacts during masking. Similarly, reflective fabrics (e.g., sequins, metallic textures, silk) should be avoided since they reflect green light back into the camera sensor, creating color spill and false edges. Optimal clothing consists of matte, solid colors that contrast strongly with the green screen, such as dark blues, reds, or neutral tones. For full-body capture, shoes should also be non-reflective and distinct from the green floor extension to prevent edge blending.

Makeup Considerations

Makeup application is important for facial data capture, particularly when expressions are part of the dataset. Glossy or shimmery cosmetics introduce unwanted highlights that interfere with consistent chroma keying and may produce variable skin tones across frames. A matte foundation helps create uniform skin texture and reduces light reflection. Eyeliner and eyebrow definition may improve landmark detection for facial expression analysis, while lip color should be chosen to contrast naturally with skin tone without oversaturation. Excessive makeup alterations should be avoided to maintain a natural baseline representation of the subject, unless variations in appearance are explicitly required by the dataset design.

Influence on Data Quality and Masking

Improper costume or makeup choices can lead to chroma spill, blurred segmentation edges, or unstable mask boundaries, reducing the accuracy of the extracted subject silhouette. These artifacts require additional post-processing, increasing the time and computational resources needed for dataset preparation. Conversely, when clothing and makeup are optimized for green screen environments, segmentation accuracy improves significantly, reducing false positives in masking and maintaining fine details such as hair strands or hand gestures. Proper preparation

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

ensures the dataset is more consistent, reducing variability across subjects and enhancing the robustness of machine learning models trained on the video data.

Makeup and Costume Checklist for Green Screen Recording

Costumes / Clothing

- ✔ Wear matte fabrics in solid colors (dark blue, gray, red, or neutral tones).
- ✔ Ensure full-body clothing is non-reflective (including shoes).
- ✔ Choose outfits that contrast strongly with green backgrounds.
- ❌ Avoid green, neon, or pastel tones that resemble the screen.
- ❌ Avoid reflective or glossy materials (silk, sequins, metallics).
- ❌ Avoid complex patterns or stripes that cause visual aliasing.

Makeup

- ✔ Apply matte foundation to reduce shine and even out skin tone.
- ✔ Use natural, non-reflective lipstick or lip tint for mouth clarity.
- ✔ Light eyeliner and eyebrow definition to improve face tracking.
- ❌ Avoid glossy, glittery, or metallic makeup products.
- ❌ Avoid extremely bright or neon colors that could distort segmentation.

Accessories and Hair

- ✔ Keep hair neat and separated from the green screen (use subtle backlighting).
- ✔ Use matte hair products (avoid sprays or gels that create shine).
- ❌ Avoid reflective jewelry, glasses, or metallic accessories.
- ❌ Avoid hats, scarves, or loose items that obstruct facial capture unless required.

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

1.3 Data processing

Preprocessing of Raw Video Data

The first step in processing is the conversion of raw camera footage into a standardized digital format suitable for frame-by-frame analysis. Videos are typically transcoded into lossless or minimally compressed formats (e.g., ProRes, DNxHR, or PNG/TIFF image sequences) to prevent degradation of chroma information. Metadata such as frame rate, resolution, and color space are preserved for reproducibility. Color correction may be applied at this stage to ensure uniform green screen saturation and eliminate variations caused by lighting.

Chroma Keying and Foreground Extraction

Chroma keying techniques are applied to separate the subject from the green screen background. This process involves identifying pixels within a defined chroma range corresponding to the green screen and removing or masking them. Advanced algorithms (e.g., spill suppression, color difference keying, or machine learning–based segmentation) are used to handle fine details such as hair, semi-transparent materials, and motion blur. The result is a binary or alpha mask representing the subject's silhouette in each frame. The mask is then combined with the original video to generate masked video frames, where only the subject remains visible and the background is transparent.

Post-processing and Mask Refinement

Post-processing ensures that the masked frames are accurate and consistent across the entire video sequence. Common refinements include edge feathering to smooth jagged boundaries, morphological operations (dilation/erosion) to correct mask leakage, and temporal smoothing to reduce flickering between frames. In cases where chroma keying fails due to clothing, lighting artifacts, or shadows, manual correction may be applied on a subset of frames, followed by interpolation to minimize workload. The final dataset consists of synchronized pairs of masked frames and their corresponding raw frames, stored in an organized format with metadata annotations. These processed frames form the foundation for training and evaluation of AI-based 2D avatars, ensuring high fidelity and minimal noise in downstream applications.

Step-by-Step processing pipeline

1. **Raw Video**
   - Input video with subjects and background (e.g., green screen).
2. **Chroma Keying**
   - Detect and separate the background using color-based keying.
   - Outputs preliminary foreground mask.

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

3. **Mask Extraction**
   - Convert keying results to a binary mask (foreground vs. background).
   - May involve thresholding or alpha extraction.
4. **Refinement**
   - Clean edges, remove noise, smooth transitions (e.g., morphological operations, matting algorithms).
5. **Masked Frames**
   - Final output frames where the foreground is cleanly separated and ready for compositing or analysis.

2. Gaussian splatting data acquisition

Creating high-quality 3D AI-based avatars for virtual reality (VR) applications requires a careful balance between realism, computational efficiency, and user interactivity. One emerging technique, Gaussian splatting, represents complex 3D geometry and textures as a cloud of Gaussian primitives rather than traditional polygonal meshes. This approach allows for smoother rendering and easier integration of neural network-based enhancements, such as AI-driven facial expressions or dynamic clothing simulations. When developing VR avatars using Gaussian splatting, it is essential to define the purpose and context of the avatar, including whether it will be used in social VR, training simulations, or gaming, as these use cases impose different requirements on visual fidelity, animation complexity, and interaction responsiveness.

The first step in avatar creation involves data acquisition and preprocessing. High-quality 3D scans or multi-view images of the intended subject provide the foundation for Gaussian splatting representations. Data should be carefully cleaned and normalized to remove noise, correct lighting inconsistencies, and align the scans to a canonical pose. AI models, such as neural radiance fields or generative models, can then be trained to convert these inputs into Gaussian splatting clouds, optimizing for both geometric accuracy and realistic texture reproduction. Guidelines suggest maintaining a balance between the number of Gaussian primitives and rendering performance, as excessive detail can lead to VR latency issues.

Character design and aesthetic consistency are crucial when generating AI-based avatars. Designers should establish clear rules for proportions, style, and color palettes to ensure avatars remain visually coherent in the VR environment. Gaussian splatting enables subtle blending of surface features and realistic soft shadows, but these effects must be carefully calibrated to avoid uncanny valley artifacts. It is also important to consider modularity, allowing the avatar to adopt multiple outfits, accessories, or hairstyles without requiring a complete rescan. AI-driven procedural generation can assist in creating variations while maintaining consistent visual style.

Animation and interactivity guidelines focus on integrating the avatar into a responsive VR environment. Gaussian splatting supports real-time deformation, making it suitable for facial expressions, hand gestures, and body movements driven by motion capture or AI pose estimation. Developers should prioritize efficient skinning and blending algorithms to ensure smooth motion, especially in multiplayer scenarios where multiple avatars are rendered simultaneously. Haptic feedback, gaze tracking, and voice-driven lip synchronization can further enhance the immersion, but must be optimized to prevent latency and motion sickness. Rigging and weight assignments should be tested thoroughly to ensure that Gaussian-based avatars deform naturally under a wide range of motions.

Optimization and deployment guidelines emphasize performance, scalability, and cross-platform compatibility. Gaussian splatting avatars can be computationally intensive, so developers should employ level-of-detail techniques, dynamic Gaussian culling, and hardware acceleration to maintain high frame rates in VR. Networked VR applications require compression strategies for avatar data while preserving visual fidelity, which may involve AI-based encoding or streaming of Gaussian primitives. Regular user testing is recommended to validate the visual realism, interactivity, and comfort of the avatars, ensuring they meet the immersive expectations of VR users. By following these guidelines, developers can leverage Gaussian splatting to create AI-powered avatars that are both visually compelling and operationally efficient in virtual reality environments.

2.1 Scene setup

Step 1: Define the Capture Space

1. Choose a **well-lit, uncluttered environment** to minimize shadows and background noise.
2. Ensure the area is **large enough for the subject to move** if you plan to capture multiple poses.
3. Mark the positions for cameras and the subject with tape or markers for **precise alignment**.

Step 2: Select and Prepare Cameras

1. Use **three high-resolution cameras** with adjustable focal lengths and manual exposure control.
2. Ensure all cameras support **synchronized triggering** or remote control to capture frames simultaneously.

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495.”

3. Mount the cameras on **tripods or fixed rigs** to prevent movement during capture.

Step 3: Arrange the Cameras

1. Position the **primary camera** directly in front of the subject to capture frontal details.
2. Place the **second camera at ~45° to the left** and the **third camera at ~45° to the right**, forming a semi-circular layout around the subject.
3. Ensure the cameras are at **similar height**, ideally at the subject's chest or eye level, to maintain consistent perspective.
4. Verify that each camera **overlaps slightly in their fields of view** to allow for accurate 3D reconstruction.

Step 4: Synchronize and Calibrate

1. Use a **synchronization signal or hardware trigger** to ensure all three cameras capture frames at exactly the same time.
2. Perform **camera calibration** using a checkerboard or calibration pattern to obtain intrinsic parameters (focal length, lens distortion) and extrinsic parameters (relative positions and angles).
3. Test a few frames to confirm **alignment, focus, and exposure** are consistent across all cameras.

Step 5: Capture Data

1. Have the subject stand in the marked position and maintain the desired pose.
2. Trigger all cameras simultaneously to capture **multiple angles at the same time**.
3. Repeat the process for **different poses or expressions**, ensuring consistent lighting and positioning.
4. Store the captured images or video sequences in **organized folders** labeled by pose, timestamp, or camera angle.

Step 6: Post-Processing Check

1. Verify that all frames are **sharp, properly exposed, and synchronized**.
2. Remove frames with motion blur or occlusions.
3. Prepare the data for **3D reconstruction** using Gaussian splatting or other neural rendering methods.

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

2.2 Data processing

Creating a Gaussian-splatted 3D avatar begins with **capturing high-quality video** from multiple cameras or a moving single camera. Each video should cover the subject from various angles with consistent lighting and high resolution to preserve fine details. Preprocessing includes extracting frames, normalizing color and brightness, and segmenting the subject from the background to remove noise. This ensures that the subsequent 3D reconstruction accurately represents the subject without artifacts from shadows or background clutter.

Once the frames are ready, **camera calibration and pose estimation** are performed. Intrinsic parameters like focal length and lens distortion are determined for each camera, while extrinsic calibration defines their relative positions and orientations. Pose estimation identifies key landmarks on the subject, enabling accurate tracking across frames. These steps allow the 2D video data to be mapped into a coherent 3D space, which is crucial for generating a realistic avatar.

The core of the process is **Gaussian splatting-based reconstruction**. A dense point cloud is generated from the aligned frames, representing the surface of the subject. Each point is replaced by a Gaussian primitive that encodes position, color, and spatial uncertainty. Overlapping Gaussians are blended and smoothed to create a continuous 3D volume, avoiding the need for explicit mesh surfaces. AI-based refinement can further enhance texture detail, surface normals, and shading, resulting in a visually rich and realistic avatar suitable for VR applications.

Finally, the Gaussian avatar is prepared for **real-time VR integration**. Optimization steps include creating multiple levels of detail, compressing the number of Gaussians, and applying animation rigs for movement. The model is exported in a format compatible with the VR engine, preserving both geometric and color data. This pipeline transforms raw video footage into a fully functional, high-fidelity 3D avatar capable of dynamic interaction in immersive virtual environments.

Step 1: Camera Calibration and Pose Estimation

Before combining frames into a 3D representation, it's crucial to understand **where each camera is and how it views the subject**. This involves:

- **Intrinsic calibration:** Determines focal length, optical center, and lens distortion for each camera.
- **Extrinsic calibration:** Establishes the relative position and orientation of all cameras.
- **Pose estimation:** Detect keypoints or landmarks on the subject (e.g., joints, face landmarks) to track motion across frames.
  These steps allow the reconstruction algorithm to accurately map 2D pixels to 3D space.

Step 2: Multi-View 3D Reconstruction

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

Once frames are aligned in 3D space, the core Gaussian splatting process begins:

1. **Point cloud generation:** From multiple synchronized frames, generate a dense point cloud representing the subject's surface.
2. **Gaussian fitting:** Replace each point in the cloud with a Gaussian primitive, which encodes position, color, and uncertainty (size/shape).
3. **Blending and smoothing:** Merge overlapping Gaussians and apply smoothing to remove noise, producing a continuous 3D representation.

Gaussian splatting differs from mesh-based reconstruction because it doesn't require explicit surfaces; instead, **volumetric Gaussians approximate geometry and texture simultaneously**, making it ideal for neural rendering and dynamic lighting in VR.

Step 3: Neural Refinement and Texture Enhancement

AI can improve the visual fidelity of the Gaussian avatar:

- **Texture synthesis:** Neural networks can enhance surface detail, filling in regions that were poorly captured.
- **Normal and lighting inference:** Predict surface normals and reflectance for realistic shading.
- **Compression and optimization:** Reduce the number of Gaussians while preserving visual quality to maintain real-time VR performance.

At this stage, the avatar is **ready for animation and integration** in VR environments.

Step 4: Exporting and Integration

The final step involves preparing the Gaussian-splatted avatar for use in VR:

- **Level-of-detail (LOD) creation:** Generate multiple LODs to optimize rendering depending on the viewer's distance.
- **Animation rigging:** Apply skeletal or blend-shape rigs if real-time movement is needed.
- **File export:** Save the Gaussian splatted model in a format compatible with your VR engine, ensuring all positional, color, and shading data are preserved.

Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."

Quality of data

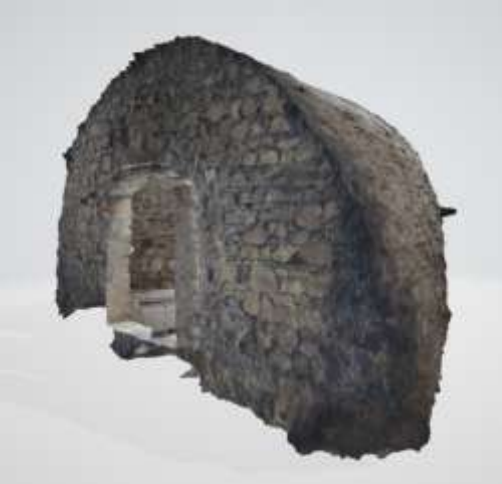
Data quality is critical for generating accurate and realistic Gaussian-splatted 3D avatars. High-resolution, well-lit, and stable video or image captures provide the foundation for precise reconstruction. Inconsistent lighting, motion blur, or occlusions can introduce noise into the point cloud, leading to artifacts or incomplete geometry. Multi-angle coverage is essential, as gaps in the subject's surface can reduce the fidelity of the final avatar. Additionally, consistent color calibration across all cameras ensures uniform textures, preventing mismatched shading or color discontinuities in the 3D model. Proper background segmentation also contributes to clean data, reducing errors when the Gaussian splatting algorithm maps points to the subject's surface.

Beyond raw capture, **temporal and spatial consistency** in the data further improves avatar quality. Videos should be synchronized when using multiple cameras, ensuring that all frames represent the subject in the same pose. Camera calibration and pose estimation accuracy directly impact the alignment of 2D frames in 3D space, so even minor calibration errors can distort the avatar. Noise reduction, artifact removal, and careful preprocessing maintain data integrity, allowing AI-based refinement to enhance rather than compensate for flawed inputs. High-quality, consistent data ultimately leads to smoother Gaussian splatting, better texture reproduction, and more immersive VR avatars.

3. Semi-automatic scene creation based on existing historical sites

Semi-automatic scene creation leverages photogrammetry to reconstruct historical sites into highly detailed 3D models with minimal manual intervention. The process begins with image acquisition, where multiple overlapping photographs of the historical site are captured from various angles. High-resolution images with consistent lighting and coverage are essential to capture fine architectural details. Drones or handheld cameras can be used to capture large or hard-to-reach areas. Tools like RealityScan (mobile) or RealityCapture (desktop) then automatically process these images to detect features, match points across images, and generate dense point clouds representing the surfaces of the site.

Once the initial point cloud is generated, the software provides semi-automatic alignment and optimization features. RealityCapture can automatically create meshes, optimize geometry, and generate textures, while allowing manual corrections where needed—such as filling gaps, removing noise, or refining misaligned sections. The semi-automatic workflow significantly reduces the labor-intensive steps of traditional 3D reconstruction, allowing users to focus on enhancing detail and correcting errors. The final output includes textured 3D meshes, which can be exported in standard formats for use in VR simulations, architectural studies, or historical preservation projects, effectively turning real-world historical sites into immersive, interactive digital environments.



Financované Európskou úniou NextGenerationEU | PLÁN OBNOVY | ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY | VA/A | MINISTERSTVO ŠKOLSTVA, VÝSKUMU, VÝVOJA A MLÁDEŽE SLOVENSKEJ REPUBLIKY | VÝSKUMNÁ AGENTÚRA

„Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00495."