

A Mixed-Initiative Procedural Workflow for Fulldome Scene Assembly from Layered 2D Illustrations

Katmanlı 2B İllüstrasyonlardan Fulldome Sahne Kurulumu için Karma-Inisiyatifli Prosedürel Bir İş Akışı

Halim YENTÜR¹

Department of Cartoon and Animation, Faculty of Fine Arts, Anadolu University, Eskişehir, Türkiye



Received / Geliş Tarihi 30.04.2026
Revision Requested /
Revizyon Talebi 23.06.2026
Last Revision / Son
Revizyon 26.08.2026
Accepted / Kabul Tarihi 22.09.2026
Publication Date / Yayın
Tarihi 10.10.2026

Corresponding author / Sorumlu Yazar:
Halim YENTÜR

E-mail: halimyentur@anadolu.edu.tr

Cite this article: Yentür, H. (2026). A Mixed-Initiative Procedural Workflow for Fulldome Scene Assembly from Layered 2D Illustrations. *PLANARCH - Design and Planning Research*, 10(3), 436-449. DOI: 10.54864/planarch.1940932



Content of this journal is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License.

ABSTRACT

Transferring layered 2D illustrations to immersive fulldome environments is challenging: manual 3D modeling does not scale to scenes containing several hundred elements, while AI-based monocular depth estimation (MDE) suffers from domain shift on stylized and panoramic content and lacks deterministic artist control. Existing approaches address procedural automation, artist-driven calibration, and dome-specific layout separately rather than in one mixed-initiative pipeline. This study proposes a mixed-initiative procedural workflow that transforms layered 2D illustrations into 3D fulldome scenes with stable placement and billboard orientation under a fixed dome camera, requiring no recomputation under camera rotation. The system uses a horizon-parameterized depth mapping heuristic converting pictorial height into radial depth, cylindrical mapping for 360° layout, and figure-derived terrain fitting via implicit surface reconstruction. Hierarchical parametric calibration preserves artist control over depth and scale at global and category levels. The workflow was tested on a 400-element historical visualization project. It achieved full procedural placement coverage across eligible scene elements with no per-element manual correction, completed in three parameter-adjustment iterations. Across 400 automatically identified overlapping figure pairs, determined by pixel-level alpha-mask intersection without manual selection, occlusion-order consistency with source-layer relations was 357/400 (89.25%). The mean ground-contact correction after raycasting was 0.62 m, with 78% of corrections within ± 1 m. Production time was estimated to improve approximately 30-fold. Rule-based depth mapping with mixed-initiative calibration thus offers a viable, deterministic, artist-controllable approach for fulldome scene assembly from stylized panoramic content.

Keywords: Procedural content generation, fulldome visualization, mixed-initiative design, depth mapping heuristic, scene assembly, digital heritage visualization.

ÖZ

Katmanlı 2B illüstrasyonların fulldome ortamlarına aktarımında iki temel güçlük vardır: manuel 3B modelleme yüzlerce sahne ögesi içeren sahneler ölçeklenmez; yapay zekâ tabanlı monoküler derinlik tahmini yöntemleri ise stilize ve panoramik içerikte alan kayması nedeniyle nesne sınırlarını koruyamaz ve deterministik sanatçı kontrolü sağlayamaz. Mevcut yaklaşımlar prosedürel otomasyon, sanatçı güdümlü kalibrasyon ve dome'a özgü yerleşimi tek bir karma-inisiyatifli akışta birleştirmek yerine ayrı ayrı ele almaktadır. Bu çalışma, katmanlı 2B illüstrasyonları sabit dome kamerası altında yeniden hesap gerektirmeyen kararlı yerleşim ve yönelimle 3B fulldome sahnelerle dönüştüren karma-inisiyatifli prosedürel bir iş akışı önermektedir. Sistem, ufuk parametrelili bir derinlik haritalama stratejisiyle figürlerin resimsel yüksekliğini radyal derinliğe çevirir; 360° düzen için silindirik haritalama, zemin için örtük yüzey çıkarımına dayalı figür tabanlı arazi uyumu uygular. Hiyerarşik parametrik kalibrasyonla sanatçı, derinlik ve ölçeği hem global hem kategori düzeyinde denetleyebilir. İş akışı, 400 sahne ögesi içeren tarihsel bir canlandırma projesi üzerinde geliştirilip sınanmıştır. Sistem, uygun sahne öğelerinin tamamında prosedürel yerleşim sağlamış, öğe bazlı manuel düzeltmeye gerek kalmamış ve yalnızca üç parametre ayarlama iterasyonu yeterli olmuştur. Alfa-maskesi piksel kesişimiyle otomatik belirlenen 400 örtüşen figür çiftinde (manuel seçim olmadan) kaynak katman ilişkileriyle oklüzyon sıra uyumu 357/400 (%89,25); raycast sonrası ortalama zemin temas düzeltmesi 0,62 m (düzeltmelerin %78'i ± 1 m aralığında) olarak ölçülmüştür. Üretim süresinin yaklaşık 30 kat azaldığı tahmin edilmektedir. Bu bulgular, kural tabanlı derinlik haritalama ile karma-inisiyatifli parametrik kalibrasyonun stilize panoramik içerikte fulldome sahne kurulumu için uygulanabilir, deterministik ve sanatçı kontrollü bir yaklaşım sunduğunu göstermektedir.

Anahtar Kelimeler: Prosedürel içerik üretimi, fulldome görselleştirme, karma-inisiyatifli tasarım, derinlik haritalama sezgiseli, sahne kurulumu, dijital kültürel miras görselleştirme.

Introduction

Full-dome projection systems are becoming increasingly prevalent in museums, planetariums, and experience centers, offering a panoramic experience that envelops the viewer within a hemispherical image (Bourke, 2005; Lantz, 2007). Whereas flat-screen content is typically framed within a frontal viewing plane, dome content surrounds the viewer. This difference is not limited to projection geometry but fundamentally transforms the spatial relationship between the audience and the scene. In cultural heritage and historical visualization projects produced for these environments, visual narrative often begins not with 3D models but with layered 2D illustrations. This is primarily because historical subjects cannot be directly photographed and because artistic interpretation is an integral part of the storytelling. However, when these layered 2D sources are transferred directly to the dome environment, figures are perceived as overlapping planar layers lacking depth, occlusion, and ground contact, significantly weakening the sense of spatial envelopment that is the core promise of the dome. Consequently, for such content to function within the dome environment, scene elements must be repositioned in 3D space with consistent depth ordering, ground contact, and perspective scale relationships.

Procedural content generation (PCG) methods, widely used in animation and game design, offer a potential solution to this scene assembly problem. However, existing approaches lie at two extremes, and both prove inadequate at the production scale of scenes containing hundreds of elements such as human figures, architectural structures, vegetation, animals, and various objects. In the manual approach, each asset is individually positioned, scaled, and oriented in 3D space. This process requires many hours for a single scene – estimated at approximately 27 hours in the present production case – and demands substantial reworking upon any compositional change. At the opposite extreme, AI-based approaches such as monocular depth estimation (MDE) face structural limitations on stylized content. MiDaS (Ranftl et al., 2022) and Depth Anything (Yang et al., 2024) are trained on natural photographs and, due to the domain gap on illustration inputs, tend not to preserve edge fidelity, determinism, or artist control. This limitation was also observed in preliminary experiments conducted during the development of this study, motivating the use of an object-based, deterministic approach over pixel-based estimation.

This study proposes a mixed-initiative procedural workflow that transforms layered 2D illustrations into 3D scenes with stable placement and billboard orientation under a fixed dome camera. This placement and orientation require no recomputation when the camera rotates. The system converts the height-in-picture-plane depth cue described by Cutting and Vishton (1995) into a procedural mechanism. Combined with cylindrical layout, figure-derived terrain fitting, and parametric calibration, it delivers a deterministic, repeatable, and artist-controlled scene assembly. The study makes three core contributions: (i) a horizon-parameterized, vertical position-based depth mapping heuristic together with cylindrical scene layout; (ii) a hierarchical parametric calibration mechanism that provides production reliability and artist control; and (iii) figure-derived terrain fitting, which reconstructs ground geometry from the figures' base points. The workflow was developed and tested within a historical full-dome visualization project containing 400 scene elements.

Procedural Generation and Mixed-Initiative Design Systems

Procedural content generation (PCG) refers to the practice of creating digital content through algorithmic methods. In their comprehensive taxonomy, Shaker et al. (2016) classify PCG methods into three categories: constructive methods, which produce deterministic content under a defined set of rules; search-based methods, which explore the solution space through optimization algorithms; and machine-learning-based methods, which generate content from patterns learned from datasets. The approach adopted in this study falls within the constructive category. Rule-based systems are favored in commercial production settings because their outputs are predictable and repeatable, properties that also help maintain consistency at crowded scene scales. However, automation alone is not sufficient for large-scale scene assembly. It must be preserved alongside artist control. For this reason, the mixed-initiative design literature is as central to this study as the PCG literature.

Mixed-initiative co-creativity (MI-CC) addresses this problem as a collaborative model in which the human designer and the algorithmic system share initiative within the creative process. The Tanagra system developed by Smith et al. (2010) is one of the pioneering efforts in this area: the designer makes a change to the level layout, the algorithm evaluates the change through a constraint solver and returns consistent suggestions, and the iterative loop continues. Liapis et al. (2013) applied a similar approach to map design with Sentient Sketchbook. Talton et al. (2009), in turn, extended this framework beyond game-level design and introduced the concept of collaborative design spaces, in which the user and the algorithm jointly explore the space of 3D modeling and spatial layout. More recently, Karadağ (2025) examined human-AI co-creativity in architectural education and showed how the MI-CC framework transforms studio culture and design pedagogy. The common thread across these studies is a sliding scale between full manual control and fully autonomous generation. The application of this scale to domains such as full-dome visualization and cultural heritage visualization, however, remains underexplored. This study is positioned at the parametric-control end of this scale: the algorithm generates the scene in a rule-bound, deterministic manner, while the artist expresses artistic intent through batch calibration over the parameter space rather than intervening on individual assets.

Depth Estimation and 2D-to-3D Scene Assembly

Early work on generating 3D scenes from 2D images established the foundational approaches. Horry et al. (1997), with their Tour Into the Picture (TIP) method, offered a simple 3D navigation from a single photograph based on a user-defined horizon point. Layered Depth Images (LDI), proposed by Shade et al. (1998), provided a structural solution to the occlusion problem by storing multiple depth layers from a single viewpoint. By preserving the pixels behind foreground objects, LDI eliminated the need for inpainting. Both TIP and LDI, however, provide only as much parallax as their stored data allows, and in environments such as full-dome that require wide-angle, depth-rich navigation, they produce occlusion gaps and prove insufficient.

Deep-learning-based monocular depth estimation (MDE) methods aimed to overcome this limitation but introduced new problems for stylized content. Models such as MiDaS (Ranftl et al., 2022) and Depth Anything (Yang et al., 2024) are trained on natural photographs and standard-perspective images, and they suffer from domain gap on stylized illustrations. Maruani et al. (2025) argue that illustrations often appear on flat media such as

book pages, posters, or paintings, which can lead monocular depth estimation models to treat them as non-volumetric or flattened content. A complementary diagnosis is offered by van Dijk and de Croon (2019), who analyzed how such models infer depth and found that neural networks disregard apparent object size and instead rely on the vertical image position of the ground-contact point. This finding becomes central to the mechanism presented in the Vertical Position-Based Depth Mapping Strategy.

When synthesizing novel viewpoints from a single image, the most critical challenge is the emergence of occluded regions behind foreground objects. Yu et al. (2025), with their WonderWorld system, offered interactive scene generation using layered Gaussian surfels, while Szymanowicz et al. (2024) proposed a solution based on 3D Gaussian Splatting (Kerbl et al., 2023) with Flash3D. However, these methods carry a risk of hallucination and can produce geometric inconsistencies during reprojection. The present study sidesteps this problem by treating the layered structure itself as the artist-prepared source-layer reference, an approach that can be understood as an object-based adaptation of LDI logic. None of the approaches above, however, addresses dome-specific projection constraints or the problem of preserving artistic fidelity in stylized content.

Immersive Environments and Artistic Fidelity

Standard MDE models assume a planar image and cannot handle the spherical warping present in 360° or domemaster projections (Szeliski & Shum, 1997). Industry-standard compositing tools support warping pixels into an equirectangular projection but do not provide automatic depth computation. The placement of large numbers of scene elements in 3D space remains entirely a manual task for the artist, and this approach does not scale to crowded scenes. Fulldome projection systems project imagery onto hemispherical surfaces using spherical-mirror or fisheye-lens techniques. Bourke (2005) provided the foundational technical references for this area, while Lantz (2007) classified large-scale immersive displays from historical and technical perspectives. Yet the existing fulldome literature focuses on projection techniques and image-warping algorithms, and no content-generation workflow for producing 3D scenes from layered 2D illustrations was identified in the reviewed literature. The film and visual-effects industry, by contrast, offers an established practice in which 2D matte paintings are projected onto simple 3D geometry (camera projection) and 2.5D sprite representations (Szeliski et al., 1999) are used as efficient, aesthetically controlled production strategies that target perceptual plausibility rather than full geometric reconstruction. This study adapts a similar logic to fulldome scene assembly.

Beyond these projection constraints, artistic composition in

illustrative content demands a perceptual space: the artist's deliberate scale and position decisions form an aesthetic coherence independent of physical accuracy. Young (2018) argued that hand-drawn 2D animation aesthetics afford a more organic reading in immersive 360° environments than the precision-oriented 3D tradition, and showed that preserving the artistic qualities of stylized content is a narrative necessity. The height-in-picture-plane depth cue defined by Cutting and Vishton (1995) forms the perceptual foundation of this study, and Gardner et al. (2010) experimentally confirmed this relationship. A critical tension arises, however, in converting this perceptual principle into a procedural mechanism. The original illustration deliberately adopts an "incorrect" perspective, and this deliberateness is part of the artistic narrative. When a mathematically correct linear perspective is applied, figures near the horizon line shrink disproportionately in 3D space, distorting the scale relationships of the original composition. Renner et al. (2013) demonstrated that distance perception in immersive environments exhibits systematic bias, indicating that the tension between metric accuracy and artistic intent is not merely aesthetic but also affects perception. Accordingly, a placement and calibration approach in this area should account for both metric accuracy and artistic intent. Heritage visualization also raises transparency expectations. Frameworks such as the London Charter (Denard, 2009) and the Seville Principles (López-Menchero Bendicho & Grande, 2011) emphasize documentation of the production process and a clear separation between historical evidence and interpretation.

Literature Gap

The studies reviewed above point to four intersecting gaps. First, monocular depth estimation models suffer from domain gap on stylized illustrations and, particularly in crowded panoramic-scale scenes, fail to preserve edge fidelity. Second, existing 360° compositing tools do not provide automatic depth computation and do not scale to scenes containing hundreds of elements. Third, automatic reconstruction systems fit the scene into metric space and do not provide a parametric calibration mechanism that preserves artistic composition. Fourth, mixed-initiative design tools focus on game design and architectural education and do not propose a content-generation workflow for fulldome visualization. The common gap is that existing approaches address procedural automation and artist control separately rather than combining them for large-scale, stylized dome scene assembly. Table 1 situates the proposed workflow at the intersection of these approach families. This study addresses this intersection by developing a mixed-initiative procedural workflow specifically for fulldome scene assembly from layered 2D illustrations.

Table 1. <i>Positioning of the proposed workflow relative to related approach families.</i>						
Approach family	Layered 2D source	Artist-controlled parameters	Dome-specific layout	Object-level placement	Rule-based reproducibility and auditability	Heritage communication / spatial storytelling
Image-based 2D→3D methods (TIP, LDI)	Primary focus	Partial	Not primary focus	Partial	Partial	Not primary focus
Learned depth estimation (MDE)	Not primary focus	Not primary focus	Not primary focus	Not primary focus	Not primary focus	Not primary focus
Procedural / mixed-initiative tools (PCG, MI-CC)	Not primary focus	Primary focus	Not primary focus	Partial	Partial	Not primary focus

Table 1.
Positioning of the proposed workflow relative to related approach families. (continued)

Approach family	Layered 2D source	Artist-controlled parameters	Dome-specific layout	Object-level placement	Rule-based reproducibility and auditability	Heritage communication / spatial storytelling
Full dome projection workflows	Not primary focus	Not primary focus	Primary focus	Not primary focus	Partial	Partial
Proposed workflow	Present study focus	Present study focus	Present study focus	Present study focus	Present study focus	Application context / interpretive framing

Cells indicate each approach family's primary design emphasis rather than absolute capability. "Primary focus" denotes a central aim of the family; "Partial" denotes a capability or concern addressed indirectly or in limited form; "Not primary focus" denotes a concern outside the family's typical scope; "Present study focus" denotes the intersection targeted by the proposed workflow.

Methods

Research Design and Scope

The study is grounded methodologically in practice-based research (Candy & Edmonds, 2018) and constructive design research (Koskinen et al., 2012). In the research-through-design approach as defined by Zimmerman et al. (2007), the design artifact is not merely a technical output but also a vehicle for knowledge. Koskinen et al. (2012) locate constructive design research within the Lab, Field, and Showroom contexts. By focusing on system consistency under controlled conditions, this study corresponds to the Lab context. Accordingly, the focus of this article is not exhibition success or audience experience but the technical consistency of the procedural workflow. The production project functions as the proof-of-concept case for this workflow.

The technical assessment of the system is structured around five criteria. Three of these are measured empirically: (i) occlusion consistency, that is, the alignment of depth ordering between figures with the visual hierarchy of the original illustration; (ii) ground-contact correction magnitude, that is, the vertical adjustment applied to seat figures onto the terrain during raycasting; and (iii) production efficiency, that is, the time savings relative to the manual process. The remaining two properties are reported as structural characteristics of the system: (iv) perspective scale consistency, that is, the match between depth-dependent size variation and perceptual expectation, which was visually verified across the parametric calibration iterations; and (v) determinism, that is, the property that identical inputs produce identical outputs, which is a structural guarantee of the rule-based architecture.

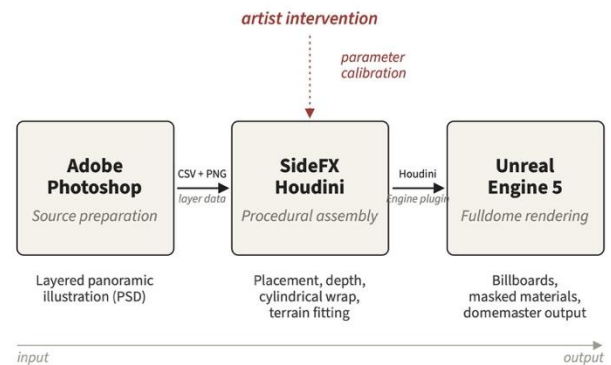
The method operates within three scope constraints. First, the scene is assembled under the assumption of a single horizon point, and as shown by Horry et al. (1997), this simplification is valid for many natural and architectural scenes. Second, occluded-surface reconstruction is excluded from the scope, and figures retain their original texture information as billboards. Accordingly, the term "scene assembly" as used here refers not to full geometric reconstruction but to the deterministic placement of 2.5D billboard representations in 3D coordinate space. Third, the full dome camera is fixed at eye level (1.70 m) near the base of the dome, and scenarios involving camera movement are not considered.

Toolchain and Design Principles

The workflow is built on three core tools: Adobe Photoshop (Adobe Inc., San Jose, CA, USA) for layer decomposition and asset preparation; SideFX Houdini (SideFX, Toronto, ON, Canada), whose node-based non-destructive architecture supports

procedural placement, depth computation, and terrain fitting; and Unreal Engine 5 (Epic Games, Cary, NC, USA) for its high-resolution real-time full dome visualization capability. Data transfer between the tools is handled through a comma-separated values (CSV) intermediate format and the Houdini Engine plugin. This configuration keeps the workflow non-destructive (Figure 1).

Figure 1.
The procedural workflow pipeline.



Note: Layered source material prepared in Adobe Photoshop is exported as a CSV table together with individual PNG textures; SideFX Houdini performs procedural placement, depth computation, cylindrical wrapping, and figure-derived terrain fitting; the resulting 3D scene is transferred to Unreal Engine 5 through the Houdini Engine plugin for full dome rendering. The dashed line indicates the parameter feedback loop available to the artist.

The defining principle in the selection of this toolchain is the mixed-initiative co-creativity (MI-CC) framework. In the model set out by Smith et al. (2010) in the Tanagra system and extended by Liapis et al. (2013), the algorithm proposes content and the human designer refines it. In this study, the division of labor works as follows: the algorithm handles position computation, depth offset, cylindrical wrapping, terrain fitting, and scene-wide updates, while the artist sets the horizon parameter, the ramp curves, and the category-level scale and offset calibration. Intervention capacity is a design principle, not a usage requirement. The system occupies a position of controlled autonomy between full autonomy and full manual operation.

Data Preparation and Layer Data Export

The production case consists of a panoramic layered Photoshop document (PSD) illustration containing 400 scene elements — human figures, animals, trees, architectural elements, and various objects (Table 2). A layered PSD is used as the source format because it preserves the artist's compositional layer order and per-element transparency, both of which the procedural pipeline depends on.

Table 2.*Production project specifications.*

Property	Value
Project / location	Full-dome visualization of a historical scene
PSD resolution	24,000 × 6,000 pixels (flat panoramic projection)
Total layer count	~520 (including base + clipping-mask chains)
Procedurally placed scene elements	400
Human figures (red tag)	320
Trees, vegetation, grasses (green)	38
Architectural elements, primarily houses (blue)	22
Animals (orange)	10
Objects / accessories (yellow)	6
Atmospheric elements and rocks (purple)	4
Export resolution	512 px long edge (preview / layout)
Reference figure height	2.4 m (stylized upper bound)
Cylindrical wrap radius (R)	~30 m

A two-stage automation script detects the visible layer stacks in the Photoshop source file and exports them as PNG and CSV.

Each exported stack includes a sequence number, pixel position, size information, composition dimensions, color tag, and blend mode. The color-tag system (red, orange, green, and so on) is used to label figure categories, and these tags form the basis for category-level parameter differentiation in the procedural placement stage.

Transfer of the layer data to the procedural system is handled through this CSV table, with each row representing one discrete figure layer. The CSV and geometry fields used by the workflow are listed in Table 3. The export contains 308 such layers, fewer than the full scene-element total in Table 2, because not every scene element corresponds to a separately exported layer. This subset of 308 layers, for which alpha-mask exports and final procedural depth values were available, forms the basis of the alpha-mask overlap-based occlusion-order audit. The pixel-to-meter conversion coefficient is computed automatically by dividing the 2D-canvas pixel height of the designated reference human figure by the metric target (2.4 m). In this project, the 2.4 m value is a design decision consistent with the illustration's stylized scale and does not claim anatomical realism. Because this value sets the pixel-to-meter conversion coefficient, it also conditions the scene's metric scale, terrain-seating magnitudes, and perceived depth relationships, and is therefore treated as a stylized project parameter. The use of a CSV intermediate format keeps the workflow tool-agnostic and non-destructive. This data-preparation step converts the artist-organized layered and categorical structure of the illustration into a machine-processable form and provides the data foundation for the subsequent procedural placement steps.

Table 3.*CSV and geometry data fields used by the workflow.*

Field	Type	Description
index / basename	integer / zero-padded identifier	Layer identifier and PNG-matching key; basename is preserved as text to retain leading zeros
xpx, ypx	integer (px)	Top-left pixel position of the layer bounding box in the source composition
wpx, hpx	integer (px)	Pixel width and height of the layer bounding box
compW, compH	integer (px)	Source composition dimensions
blend_mode_base, opacity_base	string / integer	Layer compositing mode and opacity
label_base / label_any	string	Color-tag category labels used for category-level calibration
basecolor_texture	string (path)	Reference to the exported PNG texture
P_x, P_y, P_z	float (m)	Procedurally computed 3D world position
orient_x, orient_y, orient_z, orient_w	float (quaternion)	Billboard orientation toward the dome center
scale_x, scale_y, scale_z, pscale	float	Procedurally computed per-figure scale values

One row per discrete figure layer in the geometry/audit subset (308 rows). The full production case contains 400 procedurally placed scene elements. The export holds 38 fields in total, of which the workflow-essential fields are listed.

Vertical Position-Based Depth Mapping Strategy

In the human visual system, the vertical position of an object in the visual field functions as a strong monocular depth cue (Cutting & Vishton, 1995). Objects close to the horizon line are perceived as far, while objects near the bottom edge of the image are perceived as near. Gardner et al. (2010) experimentally confirmed this relationship and showed that objects on the ground plane and those in the sky region exhibit distinct depth behaviors. The depth mapping strategy of the present study operationalizes this perceptual cue as a rule-based, artist-controlled ramp function.

The horizon parameter (y_{horizon}) defines the break point of the ramp curve in the normalized vertical position space ($t \in [0,1]$). This point marks the transition between the ground band (near figures, low push) and the sky band (far figures, high push). In functional terms, y_{horizon} defines the location of the ground-sky transition, while the ramp profile controls the sharpness of that transition. Whereas the physical horizon in the scene is determined geometrically by camera position and angle, y_{horizon} serves as the artist-controlled threshold through which that physical relationship is translated into the scene's depth behavior. By editing the ramp profile, the artist can therefore control both the position and the sharpness of this transition.

The depth generation mechanism operates in three sequential steps. First, figures are placed on a flat plane according to the pixel positions read from the PSD. Second, cylindrical wrapping is applied. Third, the normalized vertical position of each figure's ground-contact point is passed through an artist-shaped ramp curve. Here, Y corresponds to the lower edge of the figure's alpha-trimmed bounding box, that is, the ground-contact point. Van Dijk and de Croon (2019) showed that even neural networks performing monocular depth estimation rely on the vertical position of the ground-contact point rather than on the apparent size of objects as their primary depth cue. This finding suggests that the rule-based approach of the present study relies on the same cue that neural networks use implicitly, but does so in a deterministic and transparent form. The core design decision at this stage is to translate the vertical composition established by the artist in the illustration into depth information for the immersive environment, converting the height of figures in the visual field into distance in 3D space. The push amount is computed by passing the normalized vertical position through the ramp curve (Eq. 1):

$$P' = P + N \cdot d_{max} \cdot \text{ramp}(\text{normalize}(Y)) \quad (\text{Eq. 1})$$

where N is the radial surface normal after cylindrical wrapping and d_{max} is the maximum push distance set by the artist. In the present study, scene depth is defined along the normal axis of the placement surface. Accordingly, "depth" is defined not by the Z axis of the world coordinate system but by the radial direction of the cylindrical surface. This formulation allows figures in the lower region of the image (the ground band) to be pushed less and to remain in the foreground, while figures in the upper region (the sky band) are pushed more and placed in the background. The final procedural depth key is defined as the layer-order-based

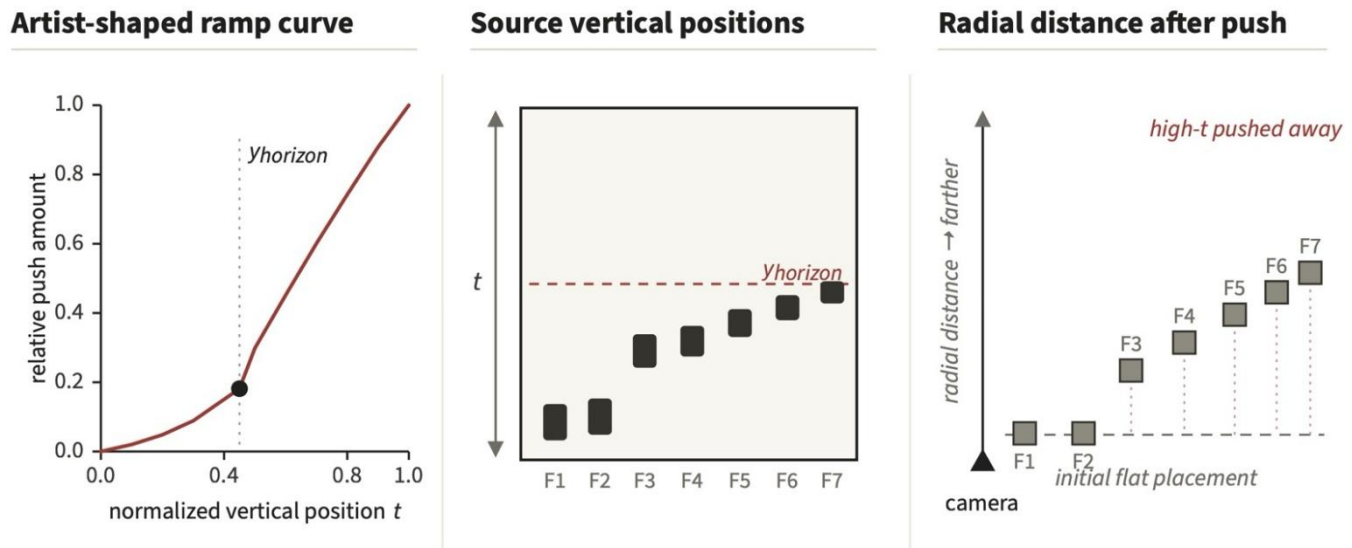
base depth plus the vertical-position-driven push along the radial surface normal. For pairwise ordering, it is operationalized as the radial distance of each figure from the fixed dome-camera origin, with a smaller radial distance corresponding to a nearer figure. The horizon line is determined indirectly by camera position and angle in the scene, and the break point of the ramp curve leaves this relationship under the artist's control.

Shaping the ramp curve with a non-linear profile allows the artist to adjust the depth transition between foreground and background according to the narrative of the scene. This approach is the procedural counterpart of the height-in-picture-plane cue: rather than a direct linear mapping, it operates on the amount by which figures are separated from the placement surface (Figure 2). As figures are pushed deeper, they shrink in camera space. The inverse-compensation mechanism is introduced to prevent this perspective shrinkage from distorting the original composition and the artist's intended scale relationships. The system computes a scale ratio by comparing each figure's camera-space depth before and after the push (Eq. 2):

$$S = (1 - \alpha \cdot w_{scale}) + \alpha \cdot w_{scale} \cdot (z_1/z_0) \quad (\text{Eq. 2})$$

where z_0 is the pre-push camera depth, z_1 is the post-push camera depth, α is the compensation ratio set by the artist, and w_{scale} is a second falloff curve drawn independently of the depth-push ramp. This arrangement allows the intensity of the perspective compensation to be adjusted separately from the push amount. Since the scale change preserves the billboard's center, a small vertical compensation is applied to keep the feet from separating from the ground. This mechanism prevents disproportionate shrinkage of pushed-away figures and preserves the perceptual coherence of the original illustration composition.

Figure 2.
The vertical position-based depth mapping mechanism.



Note. The left panel shows the ramp curve in normalized vertical position space ($t \in [0,1]$), with the horizon parameter $y_{horizon}$ marking the transition between the ground band and the sky band. The center panel shows the source illustration in which figures occupy varied vertical positions across the picture plane. The right panel shows the radial-distance result after the ramp-driven push: figures originally low in the picture remain near the camera, while figures originally high in the picture are displaced farther away, producing the depth behavior described by Cutting and Vishton (1995).

The system has two distinct reference points. Depth and scale computations are tied to an in-scene camera, the fixed dome camera. Billboard orientation, by contrast, is computed toward the geometric center of the dome, so that figures face this fixed center rather than the camera. In this restricted sense, camera-invariant placement and orientation means that the world-space

placement and billboard orientation of figures remain stable under the fixed dome camera setup, and that scale compensation is computed once under that assumption. No recomputation is required if camera rotation changes at render time, while camera position changes remain outside the scope of the study.

Parametric Calibration System

In the production case, all 400 scene elements were procedurally placed and no per-element manual intervention was required. The illustration, however, is not a one-to-one representation of physical reality: a systematic tension exists between the artist's fictional scene logic and the geometric rules of 3D space. From a production standpoint, automated placement cannot be treated as sufficiently reliable unless parametric intervention remains available, and final aesthetic control must remain with the artist.

The parametric calibration system is organized as a two-level hierarchy. The global ramp adjusts the radial placement depth and scale behavior of the entire scene at once, and the category ramps then overlay separate parameter sets for each figure type on top of this base behavior. The color tags assigned to figures

provide the basis for this category-level parameter differentiation. Through this structure, the artist can, for example, change only the scale behavior of trees, or adjust the depth ramp of a particular category as a group. The depth ramp allows radial depth values to be adjusted in bulk across bands, while the scale ramp adjusts the perspective scale (pscale) as a function of depth, inclusive of inverse-compensation scaling. The core principle of this approach is to provide bulk, non-destructive control through the parameter space rather than through individual figure edits. The workflow parameters and their operational definitions are summarized in Table 4. The motivation for this design is that the gap between physical correctness produced by the algorithm and fictional correctness embedded in the narrative of the illustration can be closed parametrically by the artist.

Table 4.
Workflow parameters and operational definitions.

Term	Symbol	Definition	Role
Horizon	y_horizon	Break point (ground-sky band transition) of the ramp in normalized vertical space ($t \in [0,1]$); together with the ramp profile, governs the depth-transition behavior	Artist parameter
Depth-push ramp	ramp(·)	Non-linear curve mapping normalized vertical position to push fraction; global and per-category	Artist parameter
Maximum push distance	d_max	Maximum radial displacement at the top of the ramp	Artist parameter
Inverse-compensation ratio	α	Strength of perspective-shrinkage compensation	Artist parameter
Compensation falloff	w_scale	Independent falloff setting how compensation varies with depth	Artist parameter
Cylinder wrapping radius	R	Radius converting horizontal world position to angle ($\theta = X_w/R$)	Artist parameter
Wrapping ratio	—	Continuous blend between flat plane and full cylindrical layout	Artist parameter
Category scale / offset	—	Per-element/category-type (color-tag-based) overrides on global depth and scale	Artist parameter
Perspective scale	pscale	Per-billboard scale, driven by the scale ramp as a function of depth	Computed (ramp-tunable)
Raycast lift offset	—	Small offset along the surface normal at the ray intersection; prevents z-fighting and sinking	System
Ground-contact point	Y	Lower edge of the figure's alpha-trimmed bounding box	Operational definition
Final depth key	—	Layer-order base depth plus vertical-position push along the radial normal; for pairwise ordering, the radial distance from the fixed dome-camera origin (smaller = nearer)	Operational definition
Eligible scene elements	—	Production elements for which automatic position, scale, and orientation transformations were generated	Operational definition

Artist parameters are set per project; this table defines each term rather than prescribing fixed values.

This parametric calibration mechanism is directly tied to the mixed-initiative co-creativity (MI-CC) framework. In the Tanagra model of Smith et al. (2010), the machine proposes and the human edits. In the present study, the machine procedurally places the figures and the artist is able to calibrate parametrically, but is not obliged to do so. A key distinction is that calibration operates at the behavior level rather than the object level, so that intervention shifts authorship granularity from per-element repair to scene-wide parameter adjustment rather than merely reducing correction frequency. The fact that parametric calibration in this project required no per-element correction suggests that the design parameters (horizon parameter, category ramps) captured artist intent sufficiently within the context of this case. Egocentric distance perception in virtual environments is known to exhibit systematic bias (Renner et al., 2013), and this phenomenon supports the view that the MI-CC structure established here goes beyond technical production reliability and serves an aesthetically functional role. Specifically, it allows the artist to compensate for perceptual bias

in line with the artist's own intent. The availability of intervention capacity preserves this flexibility across different scenes and viewing conditions (Liapis et al., 2013). In particular, even when inverse-compensation scaling is applied automatically, deviations from the compositional intent were observed in figures placed at greater distances after cylindrical wrapping. The parametric calibration mechanism provides a means to correct these deviations in bulk, on a per-category basis.

Cylindrical Mapping

The panoramic illustration represents a 360° environment, and a flat-plane placement cannot be carried over directly to the dome. Converting the panoramic illustration into a cylindrical scene layout — that is, turning horizontal position into an angular coordinate — is the core design requirement of this step. A cylindrical intermediate was preferred because it preserves the horizontally continuous panoramic organization while avoiding the vertical deformation that would complicate ground-referenced billboard placement in a dome-native

parameterization. It provides 360° circumferential wrapping along the horizontal axis while preserving the figures' relation to the ground plane and their upright billboard orientation along the vertical axis. The horizontal pixel position of each figure is converted into an angular coordinate through the relation (Eq. 3).

$$\theta = X_w / R \quad (\text{Eq. 3})$$

and then wrapped onto the cylindrical surface through arc-length parameterization. Here, X_w denotes the horizontal position in world coordinates and R denotes the wrapping radius. The point $\theta = 0$ corresponds to the front center of the scene, whereas $\theta = \pm\pi$ corresponds to the back center. The wrapping ratio is adjustable by the artist and provides a continuous transition between the flat plane and a full cylindrical layout. The original illustration is also mapped onto a reference cylinder via its UV coordinates. This construct serves as a production-time guide for the visual comparison of figure placements with the original composition. The reference cylinder is a production aid, not part of the final output. The cylindrical layout (θ) applies to the 360° scene arrangement, whereas dome projection (180° fisheye) is applied separately within the rendering engine.

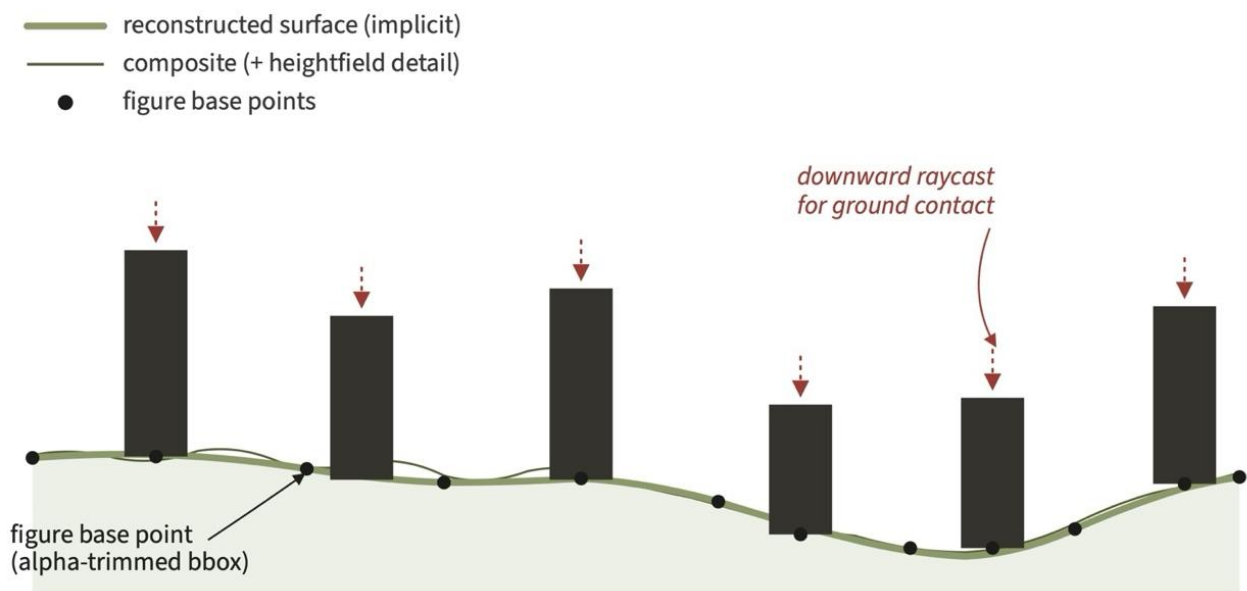
Figures initially placed on the flat plane are wrapped directly into the cylindrical form through a procedural bending operation. This step is modular: when cylindrical wrapping is disabled, a flat scene can also be produced for non-dome use cases. Terrain fitting is applied after this step, so the base points are in cylindrical coordinates at that stage. This step preserves the 360° continuity of the panoramic composition and provides the geometric foundation for the dome-specific spatial layout.

Terrain Fitting

During the export of each PSD layer, the minimum bounding rectangle of the non-transparent pixels — the alpha-trimmed

Figure 3.
Figure-derived terrain reconstruction and raycast seating.

Figure-derived terrain reconstruction and raycast seating



Note: Sparse figure base points (alpha-trimmed bounding box references) form the input from which a continuous coarse surface is reconstructed through implicit surface reconstruction (Kazhdan et al., 2006). A pre-eroded heightfield detail layer is overlaid to produce the composite surface. Each terrain-grounded figure is then seated onto the composite surface through a downward raycast, with a small lift offset applied to prevent z-fighting and sub-pixel sinking.

bounding box — is computed using the alpha channel, and its lower edge is stored as the base reference for that layer. At the scene stage, each figure card is then shifted vertically so that this lower edge aligns with the world-space ground Y coordinate. This approach captures the representative contact geometry of each figure. For example, the hem of a figure in a long dress and the base point of a tree correspond to different base geometries. Figures that must be excluded from terrain fitting (birds, clouds, airborne objects) are filtered directly from the CSV through the categories marked with Photoshop's color-tag system.

The terrain mesh is not a pre-prepared input but is instead derived from figure base points. A continuous, smooth coarse base mesh is generated from these points through implicit surface reconstruction (Kazhdan et al., 2006). After smoothing via surface resampling, the mesh is clipped to the scene boundaries and consolidated into a solid terrain body. This approach derives a manifold-like continuous surface directly from a sparse point set and preserves topological consistency across transitions between sparse and dense regions.

This figure-derived coarse terrain is then overlaid with an independently prepared, pre-eroded heightfield layer that adds high-frequency terrain details such as rock surfaces, flow marks, and micro-depressions. In this two-layer approach, the figure-derived surface defines the macro-form of the terrain while the heightfield contributes only micro-topographic detail. The large-scale landform is derived procedurally from the scene, whereas details that require artistic control are drawn from a pre-prepared source.

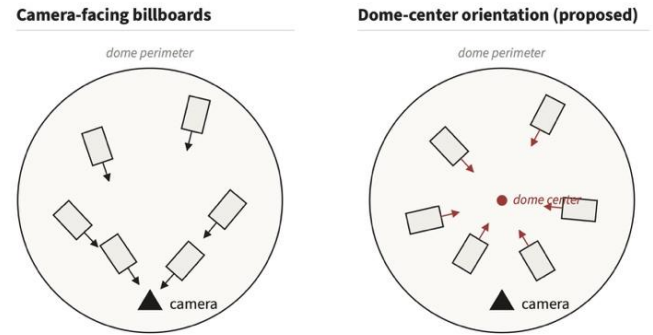
Finally, each figure is repositioned onto this composite surface through a downward raycast. A small lift offset, derived from the surface normal at the ray intersection, prevents z-fighting with the depth buffer and sub-pixel sinking (Figure 3).

Full dome Integration

The scene data produced in the procedural stage are transferred to the rendering engine through the Houdini Engine plugin. An automated script prepares each figure's texture reference in a form compatible with the material system, so that material assignment for large numbers of figure textures is handled automatically. Each figure is represented as a billboard whose world-space placement and orientation follow the placement-and-orientation scheme defined in the Vertical Position-Based Depth Mapping Strategy: positions and billboard orientations remain stable under the fixed dome camera, and changes in camera position remain outside the scope of this study. The use of billboards is a long-established optimization strategy in the image-based modeling and rendering (IBMR) literature (Szeliski et al., 1999). On the rendering-engine side, the billboard materials are configured with the Masked blend mode and alpha testing, which keeps the transparent pixels of the figures out of the depth buffer and prevents the occlusion artifacts that would otherwise be caused by those transparent pixels. As a result, the system produces a 2.5D billboard representation positioned in 3D space rather than a full geometric reconstruction: figures preserve their original texture fidelity as billboards, while their positions, scales, and occlusion orderings are defined in 3D space (Figure 4). A fixed camera is used for dome rendering. It is placed at eye level (1.70 m) near the base of the dome, with a 180° field of view and equidistant projection, producing output in the domemaster format. As noted in the Cylindrical Mapping section, depth computation and projection are independent modules: the cylindrical layout governs the 360° scene arrangement, while the

180° fisheye projection is handled separately by the rendering engine.

Figure 4.
Billboard orientation under the fixed dome camera.



Note: The left panel shows conventional camera-facing billboards, which re-orient as the camera rotates. The right panel shows the proposed geometric-center orientation used in this workflow, in which each billboard faces the fixed center of the dome rather than the camera. Under the latter scheme, world-space placement and billboard orientation remain stable across camera rotations.

Algorithmic Summary

The full procedure is summarized below (Algorithm 1). Variable definitions follow the preceding subsections and the depth, scale, and angle equations. Parameter values are project-specific and are left to artist calibration (Table 4).

Algorithm 1.

Procedural full dome scene assembly from layered 2D source.

```

Input: layer records (CSV); per-layer alpha-trimmed PNG textures
       artist parameters: y_horizon, ramp, d_max, α, w_scale, R
Output: per-figure world position P', scale S, and billboard orientation
1 for each layer record r in the CSV
2   read pixel box (xpx, ypx, wpx, hpx), category label, and texture
3   place r on the flat plane from its pixel position (pixel-to-meter scale)
4   Y ← lower edge of the alpha-trimmed bounding box           # ground-contact point
5   θ ← X_w / R                                                 # horizontal position → angle
6   wrap r onto the cylinder by arc-length parameterization
7   push ← d_max · ramp(normalize(Y))                          # vertical position → depth
8   P' ← P + N · push                                           # displace along radial normal N
9   S ← 1 - α·w_scale + α·w_scale·(z1/z0); apply S           # inverse scale compensation
10 build composite terrain from the base points of eligible figures
11 for each terrain-grounded figure: raycast downward onto the terrain; seat base point (+ lift offset)
12 orient each billboard toward the fixed dome centre; export to the rendering engine

```

Technical Assessment Protocol

The technical assessment protocol operationalizes the five criteria introduced above. This subsection details the occlusion-order audit, the principal procedure requiring a separate reproducibility protocol. The remaining measures are reported together with their corresponding results in the Results section.

Occlusion consistency is assessed through an automatic, alpha-mask overlap-based pairwise depth-ordering audit that does not rely on manually selected pairs. Each source figure is represented by its transparent layer export. A pair of figures is treated as overlapping only when their alpha masks intersect at the pixel level at their source positions, so that empty or transparent regions of the bounding boxes do not produce spurious overlaps. The audit is restricted to such pixel-level

overlapping pairs because the source-layer order does not encode a meaningful front-back relation for non-overlapping or compositionally unrelated figures. For every overlapping pair, the front-back relation given by the source-layer order is compared with the final procedural depth ordering, where the figure with the smaller radial distance from the fixed camera origin is taken as the nearer one. A source pixel is counted as occupied where its mask alpha exceeds 10 of 255, and the reported audit uses a minimal overlap-area threshold of one intersecting pixel; the result remains stable up to a 500-pixel threshold. Pairs whose radial-distance difference is below $\epsilon = 0.05$ m are treated as ties and excluded. Because the source-layer order is the reference, the audit reports consistency with the source composition's front-back relations rather than an external evaluation. The audit covers the 308 discrete figure layers for which alpha-mask exports

and procedural depth values were available. The remaining production elements are background, terrain, atmospheric, or non-discrete components without separate pairwise occlusion relations.

Results

The technical assessment was conducted on the production case described in the Methods (Table 2). This section reports the occlusion-order consistency audit, the observational perspective-scale consistency, the ground-contact correction, the qualitative MDE observation, the production metrics, and the structural assessment of determinism.

Technical Assessment Results

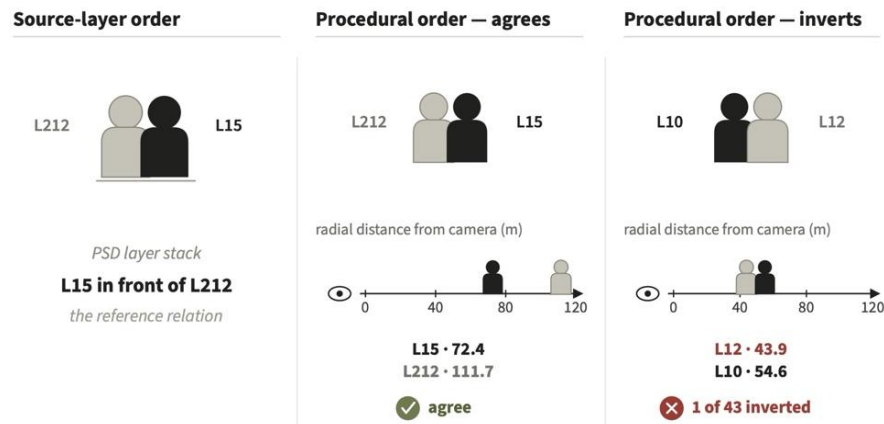
Occlusion consistency: The alignment between the procedural depth ordering and the source composition was examined through one visual production check and one automatic numerical audit, each reported as a consistency indicator rather than as a perceptual measure of visual correctness.

First, the original panoramic illustration was mapped onto the reference cylinder and placed as a background in the 3D scene, and the final depth order of the figures was compared with the apparent front-back hierarchy of the original composition. This

check also served as a production tool that provided feedback to the artist during the parametric calibration iterations.

Second, the alpha-mask overlap-based pairwise depth-ordering audit (described in the Technical Assessment Protocol) identified 400 overlapping figure pairs through pixel-level intersection of the source alpha masks, without any manual pair selection. Across these pairs, the final procedural depth ordering agreed with the source-layer front-back relation in 357 of 400 pairs (89.25%). For comparison, a coarser bounding-box criterion would have admitted 659 pairs, of which 259 had no actual alpha-level overlap. Restricting the audit to genuine pixel-level overlap removes these spurious pairs, indicating that several apparent disagreements under the bounding-box criterion arose from pairs that do not in fact occlude one another. The result was robust to the overlap-area threshold, changing by less than half a percentage point as the threshold increased from one to 500 intersecting pixels. The 43 non-agreeing pairs are concentrated where the vertical-position-driven push term locally outweighs the layer-order-based base depth, adjusting the ordering in response to the non-uniform perspective of the source illustration (Figure 5). This value is a consistency indicator with respect to the source composition's front-back relations and does not constitute a complete perceptual evaluation of occlusion correctness (Table 5).

Figure 5.
Alpha-mask overlap-based occlusion-order audit.



Note: Overlapping figure pairs are identified automatically by pixel-level intersection of the source alpha masks, without manual selection. Left: for a representative pair, the front-back relation given by the PSD layer order (L15 in front of L212). Center: the same pair in the procedural system output, ordered by radial distance from the fixed dome camera (L15 nearer, 72.4 m vs. 111.7 m); the source-layer order is preserved and the pair is counted as agreeing. Right: one of the 43 non-agreeing pairs: the procedural depth ordering places L12 nearer (43.9 m vs. 54.6 m), inverting the source-layer order in which L10 is in front, because the vertical-position-driven push locally outweighs the layer-order base depth.

Table 5. <i>Alpha-mask overlap-based occlusion-order consistency audit.</i>	
Overlapping pairs (alpha-mask, pixel-level)	400
Pair selection	Automatic; no manual selection
Ordering compared	Source-layer front-back relation vs. final procedural depth ordering (radial distance from the fixed camera origin)
Agreement	357/400 (89.25%)
Tie threshold	$\epsilon = 0.05$ m (0 ties)
Robustness	< 0.5 percentage points across overlap-area thresholds (1-500 intersecting pixels)
Status	Source-layer-order consistency indicator; not a complete perceptual evaluation of occlusion correctness

Note: Overlapping pairs are determined automatically from pixel-level intersection of the source alpha masks, so transparent regions of the bounding boxes do not create spurious pairs. A coarser bounding-box criterion would have admitted 659 pairs, of which 259 had no alpha-level overlap.

Observational assessment of perspective scale consistency: On the same test set, perspective scale behavior was evaluated observationally by the author. Through the inverse-compensation scaling mechanism, figures did not shrink disproportionately in camera space after the depth push, and the scale relationships of the original illustration composition were largely preserved. This assessment is not an independent perceptual test but a visual verification that the system's inverse-compensation mechanism produces the expected behavior.

Ground-contact agreement: For layered 2D figures, "ground contact" cannot be reduced to a single point. Because of walking and running poses, tilted stances, and perspective effects, the

two feet may appear at different heights, or the lower edge of the alpha-trimmed bounding box may not fully represent the true contact point. For this reason, while the initial procedural placement relies on the representative base point determined at the export stage, some figures can exhibit micro-scale hovering or sinking artifacts. In the proposed method, this uncertainty is resolved through a raycast on the composite terrain surface derived from the scene: for each figure, a downward ray is cast from the base point to find the ground intersection, and the figure's world Y position is corrected accordingly.

The magnitude of the ground-contact correction was measured as the vertical difference of the base point before and after raycasting: $\Delta Y_i = |Y_{pre} - Y_{post}|$. This value represents the correction amount applied by the terrain-fitting step for each figure. Of the 400 scene elements, 372 contributed to the terrain mesh, and 28 were filtered out (empty alpha-trimmed bounding box, below the height threshold, or outside the scene frame). The 400/400 procedural placement ratio indicates that every scene element received a position, scale, and orientation transformation, whereas the 372/400 terrain contribution covers only the figures that participated in deriving the ground mesh. The 28 filtered figures (airborne objects and out-of-scene elements) were procedurally placed but excluded from the terrain generation process. Note that the post-raycast residual reported in Table 6 is a structural consequence of deriving the terrain surface from the same base points and is not an independent validation metric. It is reported here as a geometric-consistency check on the raycast operation itself. Table 6 summarizes the correction statistics. The mean correction magnitude, computed over the 372 terrain-grounded elements, was 0.62 m and the median 0.48 m, and 78% of the ΔY correction magnitudes fell within a ± 1 m range. These values do not represent the system's final error margin. Rather, they indicate the magnitude of the correction applied during the raycast-based seating step used to align the representative base point with the composite terrain surface. The ground-contact correction metric should therefore be read as a measure of adjustment demand rather than as final positional error. The right-skewed shape of the distribution suggests that a small number of figures – those with greater pose or perspective ambiguity – require larger corrections. The residual after raycasting is expected by construction to approach zero, since the terrain surface is itself derived from the figure base points. This result should therefore not be interpreted as an independent empirical validation metric but as a structural outcome of the design (Table 6).

Table 6. <i>Raycast ground-contact correction statistics for terrain-grounded elements (N = 372; 28 of 400 scene elements filtered from terrain contribution).</i>				
Metric	Mean	Median	P95	Max
ΔY correction magnitude (m)	0.62	0.48	1.92	3.84
Residual after raycasting (m)	≈ 0	≈ 0	≈ 0	≈ 0

Terrain mesh: produced through implicit surface reconstruction (Kazhdan et al., 2006) and, after resampling, consisting of ~4,200 triangles with a target edge length of ~2.8 m. Note: the ≈ 0 residual after raycasting is not an independent validation metric but a structural consequence of deriving the terrain surface from the figure base points.

Comparative observation against AI depth estimation in a production context: The MDE limitations discussed in the Depth Estimation and 2D-to-3D Scene Assembly section were also observed on the panoramic inputs of this project. MiDaS (Ranftl

et al., 2022) processed the input but produced a near-uniform depth gradient that did not distinguish figures, architecture, or ground from the sky region. Depth Anything (Yang et al., 2024) did not process the panoramic input at all: the model's standard inference pipeline was not operationally compatible with the input resolution. This dual observation suggests that current MDE pipelines are unsuitable for stylized panoramic content not only in output quality but also in operational feasibility. This is a qualitative comparison, not a systematic benchmark, included to clarify the production rationale for the proposed rule-based approach. The comparison was not designed as a controlled benchmark: the panoramic input was evaluated in its production resolution and layout, and no model-specific retraining or dataset adaptation was performed. Each output was assessed qualitatively by whether it preserved figure-ground separation, edge fidelity, and depth differentiation usable by the workflow.

Production metrics: The production efficiency of the system was evaluated along two dimensions. First, placement coverage: all 400 scene elements were procedurally placed and no per-element manual correction was required. Here, "400/400 procedural placement" indicates that an automatic position, scale, and orientation transformation was produced for every scene element. This should not be interpreted as a claim of correctness, but rather as a scalability indicator suggesting that control in this workflow is shifted into the parameter space rather than remaining tied to per-element revision. The calibration process was carried out within the hierarchical parametric structure: the global ramp was used to set the overall behavior of the entire scene, and the category ramps were then used for element-type-specific fine tuning. This calibration was completed in three iterations (Table 7).

Table 7. <i>Hierarchical calibration summary (three iterations, total calibration time ~26 min).</i>				
Iteration	Stage	Parameters adjusted	Scope	Observation
1	Global	Depth ramp profile, maximum push distance	400 scene elements	Overall depth range and horizon transition setting
2	Category	Human scale compensation, tree push distance, architectural wrap ratio	320 + 38 + 22 elements	Type-based fine tuning
3	Category	Animal pivot correction, atmospheric element terrain exclusion	10 + 4 elements	Final ground-contact and filter adjustments

Placement coverage: full coverage across the eligible scene elements. Per-element manual correction: 0. The object category (6 elements) reached sufficient results with the global ramp settings alone and did not require a separate category calibration. The calibration time reported here is included in the procedural-system setup time reported in Table 8.

Second, production efficiency: placing the 400 scene elements individually through manual scene assembly is estimated to require approximately 27 hours (extrapolated from a pilot measurement; see the footnote to Table 8). In the procedural system, infrastructure setup was completed in approximately 55

minutes, and every subsequent parameter change propagated through all scene elements within 30-120 seconds (~30× speedup). The system was developed in the project's early stages with approximately 100 figures and scaled to 400 scene elements without requiring additional infrastructure changes (Table 8).

Table 8. <i>Scene assembly and revision time comparison (excluding asset preparation and illustration time).</i>			
Workflow	Setup time	Iteration time	Total (estimated)
Manual baseline	~27 hours	per-element re-adjustment	~27+ hours
Procedural system	~55 min	30-120 s (scene-wide update)	~1 hour

The manual baseline estimate was obtained from pilot trials conducted in the early stages of the project with a small subset of figures: positioning, scaling, and orientation adjustment of each scene element in 3D space took an average of ~4 minutes ($400 \times \sim 4 \text{ min} \approx 27 \text{ hours}$). This value covers only the scene-assembly duration and does not include asset preparation, iteration, or revision time. In practice, a fully manual production is projected to take more than a working week.

Structural assessment of determinism: Figure positions responded systematically and predictably to changes in the ramp curves, indicating that the parametric control layer behaved consistently. Determinism itself, however, is treated here as a structural property of the rule-based system: repeated runs with the same CSV input and parameter values produced identical output. This property corresponds directly to the determinism criterion defined in the Methods section and should be understood as an architectural guarantee rather than as an independent empirical metric.

Discussion

The primary finding of this study is that all 400 scene elements were placed procedurally without per-element manual correction, with revision absorbed into category-level parameter calibration over three iterations. Interpreted through the mixed-initiative co-creativity (MI-CC) literature, this result locates the workflow at the parametric-control end of the sliding scale between full manual authoring and fully autonomous generation. In the systems that established this paradigm – Tanagra (Smith et al., 2010), Sentient Sketchbook (Liapis et al., 2013), and collaborative design spaces (Talton et al., 2009) – the machine proposes and the designer refines at the level of individual edits. The present case extends the same division of labor to full-dome heritage assembly, a domain the reviewed MI-CC literature had not addressed, and shows that the artist's authorial granularity shifts from per-element repair to scene-wide parameter adjustment. That calibration converged in three iterations without per-element intervention suggests the chosen parameters (horizon, ramp profiles, category scale and offset) encoded compositional intent adequately for this production. This outcome is therefore evidence of controlled automation rather than of automation displacing design judgment.

Framing this contribution for immersive heritage requires care about what the workflow does and does not claim. The system is a tool for heritage communication and spatial storytelling – assembling an interpretive scene for a dome audience – rather than an instrument of authoritative reconstruction. This distinction matters because the source content rests on artistic interpretation of subjects that cannot be photographed, and the workflow neither generates nor adjudicates historical evidence. Two reference frameworks clarify the responsibilities involved.

The London Charter (Denard, 2009) calls for intellectual transparency and documentation of the process by which a visualization is produced. Because every placement here derives from explicit, logged parameters rather than an opaque statistical model, the pipeline is inherently more auditable against that requirement than learned alternatives. The determinism discussed below is, in this sense, also a heritage-transparency property. The Seville Principles (López-Menchero Bendicho & Grande, 2011; ratified by ICOMOS, 2017) further distinguish historical evidence from interpretation and call for scientific transparency in virtual archaeology. Read against these frameworks, the workflow's contribution to design and planning is a reproducible, parameter-documented route from layered 2D interpretation to immersive spatial experience, with the explicit caveat that the representational accuracy of the depicted content remains the responsibility of the source authorship and falls outside the scope of the assembly method.

The occlusion-order result is best read as a statement about depth ordering rather than depth reconstruction. Occlusion and height-in-picture-plane are among the strongest ordinal depth cues in human vision (Cutting & Vishton, 1995), and the system encodes both: layer order as base depth and vertical position as a calibrated push. Image-based 2D-to-3D methods such as Tour Into the Picture (Horry et al., 1997) and Layered Depth Images (Shade et al., 1998) reconstruct parallax from stored pixel layers, but they yield only as much parallax as the stored data allows and degrade on the wide-angle, stylized panoramas a dome requires. The object-based approach taken here treats each artist-prepared layer as the source-layer reference and reconstructs ordinal depth deterministically. An objective, automatic consistency audit based on pixel-level alpha-mask overlap (described in the Technical Assessment Protocol) found that the procedural ordering agreed with the source-layer front-back relation in 357 of 400 overlapping pairs (89.25%). This value is a consistency indicator with respect to the source composition's ordinal relations, not a perceptual proof of occlusion correctness, and the 43 non-agreeing pairs are concentrated where the height-driven push locally outweighs the layer-order base depth, in response to the non-uniform perspective of the original illustration. The tradeoff is explicit: the method preserves ordinal structure and artist fidelity at the cost of intra-figure parallax, since each element remains a billboard rather than a reconstructed volume.

The contrast with AI-based monocular depth estimation (MDE) is epistemic and production-contextual rather than a benchmark comparison. No systematic quantitative evaluation against MDE models was conducted, and the observations reported here are qualitative. MDE models such as MiDaS (Ranftl et al., 2022) and Depth Anything (Yang et al., 2024) infer depth from statistical regularities in natural photographs, so an artist's deliberate compositional choices may be treated as noise and no direct control over the output is available. Maruani et al. (2025) similarly note that illustrations and printed visual elements are often treated as flat media by MDE pipelines. This observation is drawn from recent preprint work and reported here as consistent with, rather than confirmation of, the present observations. Notably, van Dijk and de Croon (2019) found that MDE networks rely on the vertical image position of the ground-contact point rather than apparent object size – the same cue this workflow encodes explicitly. The rule-based system therefore draws on the depth cue that learned models use implicitly, but expresses it in a transparent, editable, and deterministic form. Determinism here is a structural property of the rule-based architecture – identical inputs yield identical outputs – and not a performance metric. Its

value in this setting is reproducibility and artist control, which the heritage-transparency requirements above render a substantive rather than merely technical advantage.

These findings should be read within clear limits. The technical assessment rests on a single production project, so the numerical results are case-specific and carry no claim of statistical generalization. The two-layer parameter space comprises scene-dependent parameters (horizon, cylinder radius, inverse-compensation ratio, scale-falloff curve, ramp profiles) that are re-tuned per project, and system-dependent components (cylindrical transformation, terrain algorithm, billboard orientation, base-point detection) that are held fixed. This structure offers a structural pathway to other projects, but evaluation across multiple projects and operators remains future work. The occlusion audit is an algorithmic consistency check against the source layer order: because the base depth is itself the layer order, agreement is partly structural, and the metric measures how often the calibrated push preserves the artist's source ordering rather than providing an independent perceptual validation. The estimated production-time gain – approximately 30-fold relative to a manual baseline – is extrapolated from small-subset pilot trials by a single experienced operator and should be treated as a production-context estimate, not a general performance result. The ground-contact correction magnitude (mean 0.62 m) reflects adjustment demand during raycast seating rather than final positional error. Finally, the reference figure height of 2.4 m is a stylized upper bound that conditions metric scaling and perceived depth, and user-experience evaluation in the dome was kept outside the scope of this system-focused study.

Conclusion and Recommendations

Given that the study was developed within single-case and fixed-camera constraints, future research may proceed in four directions in particular. First, hybrid AI approaches: methods such as the layer-index prediction proposed by Maruani et al. (2025) could be combined with the rule-based depth mapping heuristic to increase automation. Second, support for the Universal Scene Description (OpenUSD) format could extend transfer to different rendering engines. Third, multi-horizon support could address more complex scene geometries. Fourth, the material system on the rendering-engine side, instancing optimization, and dynamic environmental elements will be detailed in a future study.

Overall, the findings demonstrate that the proposed workflow positions scene assembly from layered 2D illustrations into fulldome environments as a form of controlled automation. In this production case, that position was supported by full procedural placement coverage across the eligible scene elements, 357/400 (89.25%) agreement in an automatic alpha-mask overlap-based occlusion-order audit, and an estimated, pilot-extrapolated production speedup of approximately 30-fold relative to the manual workflow. The workflow shifts revision from per-element correction into the parameter space while keeping artistic control operationally central.

Ethics Committee Approval Certificate: The author declared that an ethics committee approval certificate is not required.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author has no conflicts of interest to declare.

Data Availability: The underlying production imagery and full project files cannot be shared owing to rights-holder, anonymity, and production constraints. The workflow parameters, data structures, and algorithmic procedure are documented in Tables 3 and 4 and

Algorithm 1, and the consistency-audit procedure is described in the Methods. Anonymized tabular data and audit scripts may be made available by the author upon reasonable request, subject to rights and project constraints.

Financial Disclosure: The author declared that this study has received no financial support.

Use of Artificial Intelligence: AI-assisted tools were used for language editing only. The author takes full responsibility for the content, originality, and accuracy of the manuscript.

References

- Bourke, P. (2005). Spherical mirror: A new approach to hemispherical dome projection. *Planetarian*, 34(4), 6-9.
- Candy, L., & Edmonds, E. (2018). Practice-based research in the creative arts: Foundations and futures from the front line. *Leonardo*, 51(1), 63-69. https://doi.org/10.1162/LEON_a_01471
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69-117). Academic Press. <https://doi.org/10.1016/B978-012240530-3/50005-5>
- Denard, H. (Ed.). (2009). *The London Charter for the computer-based visualisation of cultural heritage* (Version 2.1). <https://www.londoncharter.org>
- Gardner, J. S., Austerweil, J. L., & Palmer, S. E. (2010). Vertical position as a cue to pictorial depth: Height in the picture plane versus distance to the horizon. *Attention, Perception, & Psychophysics*, 72(2), 445-453. <https://doi.org/10.3758/APP.72.2.445>
- Horry, Y., Anjyo, K., & Arai, K. (1997). Tour into the picture: Using a spidery mesh interface to make animation from a single image. *SIGGRAPH '97*, 225-232. <https://doi.org/10.1145/258734.258854>
- Karadağ, D. (2025). AI in architectural education: Rethinking studio culture. *PLANARCH - Design and Planning Research*, 9(2), 243-253. <https://doi.org/10.54864/planarch.1749891>
- Kazhdan, M., Bolitho, M., & Hoppe, H. (2006). Poisson surface reconstruction. *Eurographics Symposium on Geometry Processing*, 61-70. <https://doi.org/10.2312/SGP/SGP06/061-070>
- Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), Article 139. <https://doi.org/10.1145/3592433>
- Koskinen, I., Zimmerman, J., Binder, T., Redström, J., & Wensveen, S. (2012). *Design research through practice: From the lab, field, and showroom*. Morgan Kaufmann. <https://doi.org/10.1016/C2010-0-65896-2>
- Lantz, E. (2007). A survey of large-scale immersive displays. In *Proceedings of the 2007 Workshop on Emerging Displays Technologies: Images and Beyond*. ACM. <https://doi.org/10.1145/1278240.1278241>
- Liapis, A., Yannakakis, G. N., & Togelius, J. (2013). Sentient Sketchbook: Computer-aided game level authoring. *FDG '13*, 213-220.
- López-Menchero Bendicho, V. M., & Grande, A. (2011). *International principles of virtual archaeology (the Seville Principles)*. Spanish Society of Virtual Archaeology (SEAV) / International Forum of Virtual Archaeology. (Ratified by the 19th ICOMOS General Assembly, New Delhi, 2017.)
- Maruani, N., Zhang, P., Chaudhuri, S., Fisher, M., Zhao, N., Kim, V. G., Alliez, P., Desbrun, M., & Wang, Y. (2025). Illustrator's depth: Monocular layer index prediction for image decomposition. *arXiv:2511.17454* [Preprint]. <https://doi.org/10.48550/arXiv.2511.17454>
- Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2022). Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3), 1623-

1637. <https://doi.org/10.1109/TPAMI.2020.3019967>
- Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments: A review. *ACM Computing Surveys*, 46(2), Article 23. <https://doi.org/10.1145/2543581.2543590>
- Shade, J., Gortler, S., He, L., & Szeliski, R. (1998). Layered depth images. *SIGGRAPH '98*, 231-242. <https://doi.org/10.1145/280814.280882>
- Shaker, N., Togelius, J., & Nelson, M. J. (2016). *Procedural content generation in games*. Springer. <https://doi.org/10.1007/978-3-319-42716-4>
- Smith, G., Whitehead, J., & Mateas, M. (2010). Tanagra: A mixed-initiative level design tool. *FDG '10*, 209-216. <https://doi.org/10.1145/1822348.1822376>
- Szeliski, R., & Shum, H.-Y. (1997). Creating full view panoramic image mosaics and environment maps. *SIGGRAPH '97*, 251-258. <https://doi.org/10.1145/258734.258861>
- Szeliski, R., Anandan, P., & Baker, S. (1999). From 2D images to 2.5D sprites: A layered approach to modeling 3D scenes. *IEEE International Conference on Multimedia Computing and Systems*, 1, 44-50. <https://doi.org/10.1109/MMCS.1999.779118>
- Szymanowicz, S., Insaftudinov, E., Zheng, C., Campbell, D., Henriques, J. F., Rupprecht, C., & Vedaldi, A. (2024). Flash3D: Feed-forward generalisable 3D scene reconstruction from a single image. *arXiv:2406.04343* [Preprint]. <https://doi.org/10.48550/arXiv.2406.04343>
- Talton, J. O., Gibson, D., Yang, L., Hanrahan, P., & Koltun, V. (2009). Exploratory modeling with collaborative design spaces. *ACM Transactions on Graphics*, 28(5), Article 167. <https://doi.org/10.1145/1618452.1618513>
- van Dijk, T., & de Croon, G. (2019). How do neural networks see depth in single images? *Proceedings of the IEEE/CVF International Conference on Computer Vision* (ICCV), 2183-2191. <https://doi.org/10.1109/ICCV.2019.00227>
- Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., & Zhao, H. (2024). Depth anything: Unleashing the power of large-scale unlabeled data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 10371-10381. <https://doi.org/10.1109/CVPR52733.2024.00987>
- Young, M. (2018). Drawn to 360°: How can the aesthetics and qualities of traditional 2D animation storytelling add to the immersive VR projection paradigm? *Animation Practice, Process & Production*, 7(1), 11-40. https://doi.org/10.1386/ap3.7.1.11_1
- Yu, H.-X., Duan, H., Herrmann, C., Freeman, W. T., & Wu, J. (2025). WonderWorld: Interactive 3D scene generation from a single image. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 5916-5926. <https://doi.org/10.1109/CVPR52734.2025.00555>
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. *CHI '07*, 493-502. <https://doi.org/10.1145/1240624.1240704>