Olbedo: An Albedo and Shading Aerial Dataset for Large-Scale Outdoor Environments

Supplementary Material

A. Image Processing Details

This section expands the description of our aerial image preprocessing pipeline in the main paper. We detail how RAW images are decoded, converted to linear RGB EXRs, georegistered, and downsampled before albedo–shading decomposition.

Acquisition setup. All aerial images are captured using a DJI Phantom 4 Pro V2.0 with the built-in 1-inch CMOS camera. We record *only* RAW Digital Negative (DNG) files; in-camera JPEGs are not used anywhere in our pipeline. First, JPEGs apply vendor-specific tone curves, camera response functions (CRFs), and other non-linear image signal processing that destroy the linear relationship between scene radiance and pixel values. Second, the JPEG images have been undistorted with the built-in coefficients, which can conflict with the radial–tangential distortion model used in photogrammetry software. Third, RAW DNGs provide a higher bit depth (12 or 14 bits per channel) compared to JPEGs (8 bits per channel), which helps preserve subtle details in shadows and highlights that are crucial for accurate albedo–shading decomposition, as illustrated in Fig. A.1.

(a) In-camera JPEG (b) RAW DNG

Figure A.1. Comparison between in-camera JPEG (a) and RAW DNG (b) from the same scene. The JPEG image exhibits visible artifacts such as color clipping in highlights, loss of shadow details, and posterization effects due to aggressive compression and tone mapping. In contrast, the RAW DNG preserves a wider dynamic range and more accurate color representation, which are essential for reliable albedo–shading decomposition.

RAW decoding and linearization. We decode all DNGs using OpenImageIO (OIIO) configured to return values in the camera's linear RAW color space:

- `oiio:RawColor = 1` to request RAW sensor-space values without applying DNG color transforms.
- `raw:ColorSpace = Linear` to ensure the output is linear with respect to sensor radiance.

No in-camera JPEG pipeline, tone mapping, gamma curve, or camera response function (CRF) is applied at any stage of our preprocessing. We use OIIO's demosaicing and white-balance handling as provided by its RAW interface, while keeping the output in a linear radiometric scale.

Conversion to EXR. All RGB, albedo, and shading images that we release are stored as OpenEXR (EXR) files in linear color space:

- The RGB EXRs are obtained directly from the RAW decoding described above.
- The albedo and shading EXRs are generated by running the intrinsic decomposition method [2] on the downsampled linear RGB images.
- No gamma correction, tone mapping, or sRGB transfer function is baked into these EXRs.

sRGB is used only for visualization in the paper (figures) and for low-resolution preview thumbnails. Any user wishing to work in display space should explicitly convert the provided EXRs using a standard sRGB transfer function.

Camera calibration and geometry. Camera intrinsics, poses, and 3D geometry are estimated using the commercial photogrammetry software Bentley iTwin Capture[1]. To handle multiple flights of the same area, we use virtual ground control points to align different captures into a single, georegistered coordinate system. The virtual GCPs are created by manually selecting keypoints on prominent, static features visible across different flights (e.g., building corners, road intersections). iTwin Capture uses these points to optimize camera poses and align the reconstructed models into a common world coordinate system with real-world scale and orientation. A uniform coordinate system allows us to reproject from one image to another using depth maps for cross-view consistency checks and change detection. We provide the calibrated camera intrinsics and extrinsics for each image in the dataset. The reconstructed 3D geometry is exported as a textured mesh in OBJ format.

B. Image Decomposition Method

We adopt an outdoor illumination model proposed by Song et al. [2] to describe the shading conditions in outdoor images. We start from the rendering equation below. The illumination consists of a directional component from solar radiation and a spherical component from the sky.

[1] https://www.bentley.com/software/itwin-capture/

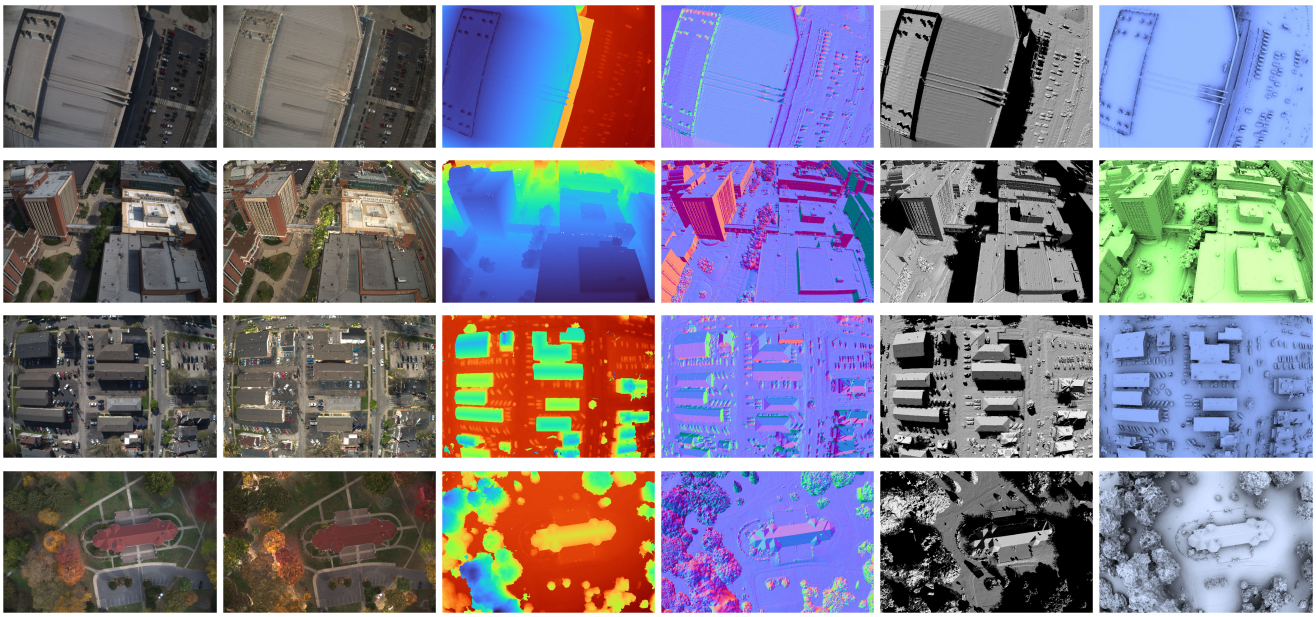


Figure A.2. Additional examples from the Olbedo dataset. Rows from top to bottom: arena, office, residential, and park scenes.

$$L_o(\omega_o) = L_e(\omega_o) + \int_\Omega L_i(\omega_i) f_r(\omega_i, \omega_o) \langle \omega_i \cdot \mathbf{n} \rangle^+ d\omega_i, \tag{B.1}$$

where $L_o(\omega_o)$ is the outgoing radiance in direction ω_o, $L_e(\omega_o)$ is the emitted radiance, $L_i(\omega_i)$ is the incident radiance from direction ω_i, $f_r(\omega_i, \omega_o)$ is the bidirectional reflectance distribution function (BRDF), $\mathbf{n}$ is the surface normal, and $(\cdot)^+$ denotes the positive-part function, i.e., $\max(0, \cdot)$.

For diffuse, non-emissive surfaces, the rendering equation simplifies to

$$L_o(\omega_o) = \int_\Omega L_i(\omega_i) \langle \omega_i \cdot \mathbf{n} \rangle^+ d\omega_i, \tag{B.2}$$

$$\text{where} \quad L_i(\omega_i) = L_{sun}(\omega_i) + L_{sky}(\omega_i). \tag{B.3}$$

The incident radiance $L_i(\omega_i)$ can be further decomposed into sun and sky components as Eqs. (B.4) and (B.5).

$$L_{sun}(\omega_i) = \psi_{sun} \cdot V_{sun} * \delta(\omega_i - \theta_{sun}), \tag{B.4}$$

where ψ_{sun} is the solar radiance constant, V_{sun} is the visibility function for the sun (1 when the sun is visible, 0 otherwise), and δ is the Dirac delta function centered at the sun direction θ_{sun}.

$$L_{sky}(\omega_i) = \psi_{sky} \cdot \mathcal{G}(\omega_i - \theta_{sun}) \cdot V_{sky}(\omega_i), \tag{B.5}$$

where ψ_{sky} is the sky radiance constant, and $\mathcal{G}$ is a spherical function modeling the sky light distribution, with its lobe centered at the sun direction θ_{sun}. To simplify the model in practice, we use a uniform sky distribution, i.e., $\mathcal{G}$ is constant over the visible sky. V_{sky} is the sky visibility function; it is defined on the hemisphere above the horizontal plane, matching the support of the sky light distribution.

$$\phi = \frac{\psi_{sun}}{\psi_{sky}}. \tag{B.6}$$

We can integrate the sun and sky components separately and rearrange them as normalized shading components corresponding to the image formation model described in the main paper:

$$\begin{aligned} S_{sun} &= V_{sun} \cdot \langle \mathbf{n}, \theta_{sun} \rangle^+, \\ S_{sky} &= \int_\Omega V_{sky}(\omega_i) \cdot \mathcal{G}(\omega_i - \theta_{sun}) \cdot \langle \mathbf{n}, \omega_i \rangle^+ d\omega_i. \end{aligned} \tag{B.7}$$

In Fig. A.2, we present additional data samples spanning multiple scenes and modalities from the *Olbedo* dataset.

C. Failure Cases in Pseudo-Ground-Truth Generation

During data production, we observe several recurring failure modes of the pseudo-ground-truth generation process, shown in Fig. C.1. These include geometry holes, sharp shadow-boundary noise, imperfect tree geometry, glass reflection, and glass transmission. To improve the training data quality, we manually screen the generated results and identify a

subset of 2.49k relatively clean images with fewer artifacts such as incorrect shadows, reflections, and transmissions.

D. Confidence Masks

The binary confidence mask (shown in Fig. D.1) is automatically generated from two cues only: geometric boundaries from the reconstructed scene geometry and shadow-projection boundaries from the estimated cast-shadow map. We dilate and combine these boundaries to suppress supervision near geometry holes, depth discontinuities, and sharp shadow transitions, where the albedo–shading decomposition is most unstable. The resulting mask mainly removes artifacts around holes and shadow boundaries.

During fine-tuning, we apply this confidence mask to restrict supervision to valid regions, as described in Section 4 of the main paper. This helps prevent unreliable pseudo-ground-truth near geometry and shadow boundaries from dominating training.

E. Multi-view Consistent Retexturing

We apply the predicted albedo obtained from our fine-tuned RGB↔X model to retexture the reconstructed 3D models. Experiments are conducted on both a self-collected UAV dataset featuring buildings with fine structural details, and the publicly available BlendedMVS dataset [3] containing a variety of landscape types. As shown in Fig. E.1 and Fig. E.2, the retextured models exhibit a consistent appearance across all scenes, with most shading effects effectively removed compared to the original RGB-textured models. Notably, the retextured models preserve fine texture details—for example, building roof patterns and road markings remain clearly visible. These visualizations highlight the advantage of albedo recovery: data acquisition becomes less constrained by lighting conditions, and the albedo-based textures further enable flexible downstream applications such as relighting and material editing.

F. Relighting Renders

We provide visual comparisons in Fig. F.1 across a diverse set of outdoor scenes. Each scene is rendered using both the original RGB texture and the recovered albedo texture. The examples span a wide range of structural geometries, surface materials, vegetation densities, and illumination conditions, including strong directional sunlight, partially occluded skylight, overcast weather, rainy environments, and nighttime lighting. Across all scenarios, the RGB-textured models exhibit baked-in shading effects. In contrast, the albedo-textured models consistently suppress these shading artifacts and produce spatially coherent, radiometrically stable appearances while preserving high-frequency material details. These results demonstrate that our approach generates albedo textures substantially more suitable for downstream rendering, relighting, and weather-agnostic asset creation.

G. Additional Segmentation Results

We provide additional qualitative segmentation examples using SAM [1] on both RGB and albedo images. As shown in Fig. G.1, the segmentation performance is consistently better on albedo images across a broad range of outdoor scenes, particularly those containing complex cast shadows.

References

[1] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. SAM 2: Segment Anything in Images and Videos, 2024.

[2] Shuang Song and Rongjun Qin. A general albedo recovery approach for aerial photogrammetric images through inverse rendering. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218:101–119, 2024.

[3] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. BlendedMVS: A Large-scale Dataset for Generalized Multi-view Stereo Networks. *Computer Vision and Pattern Recognition (CVPR)*, 2020.

(a) (b) (c) (d) (e)

Figure C.1. Representative failure cases in pseudo-ground-truth generation. The top row shows the raw inputs and the bottom row shows the corresponding recovered albedo. Columns (a)–(e) illustrate geometry holes, sharp shadow-boundary noise, imperfect tree geometry, glass reflection, and glass transmission, respectively.

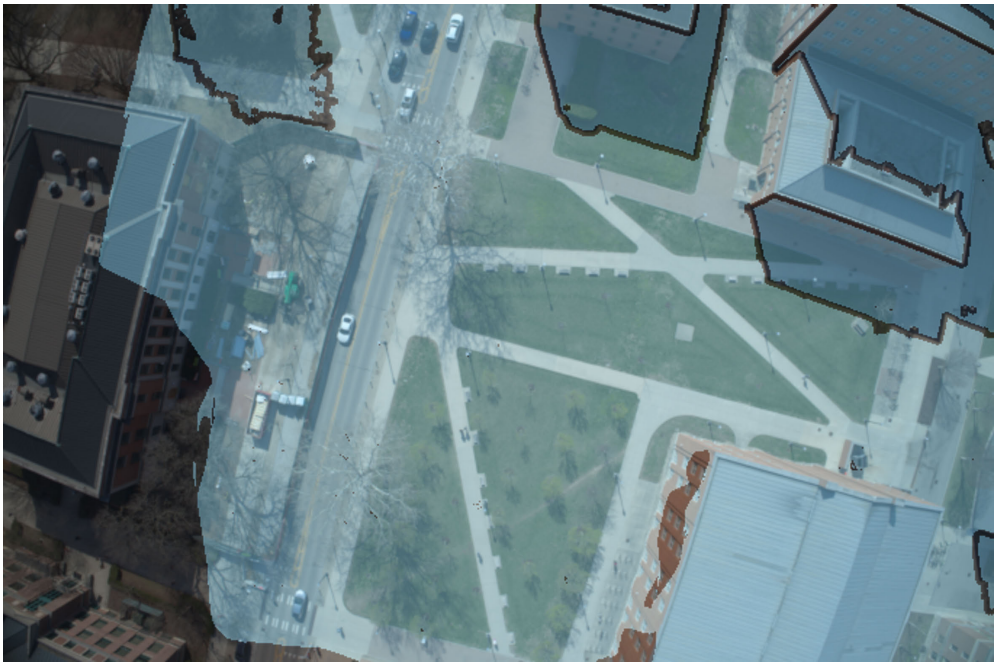
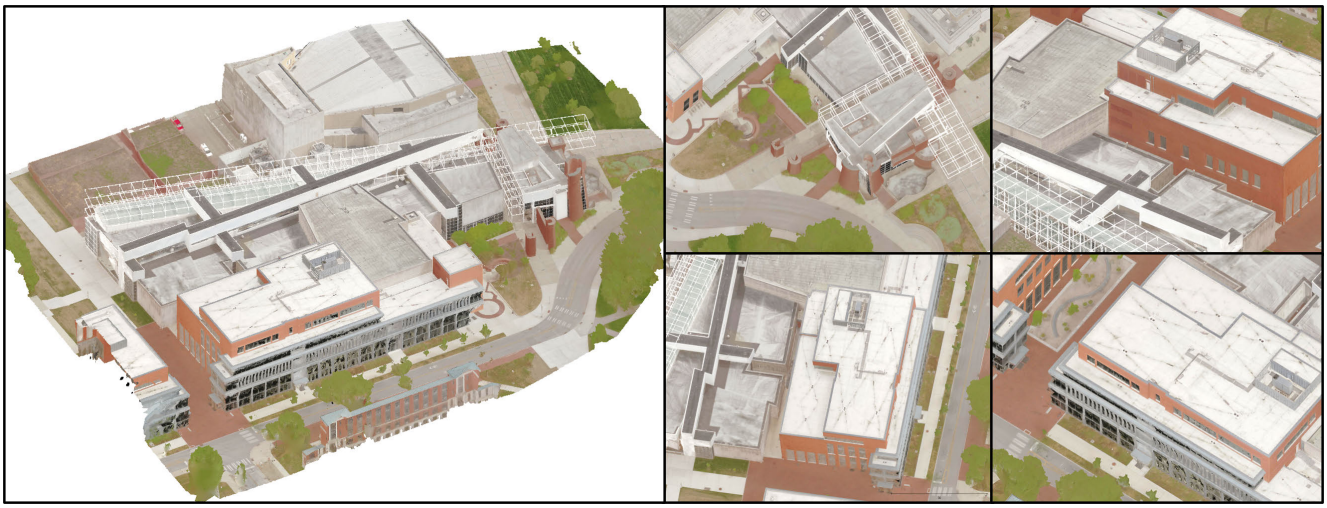


Figure D.1. Example of the binary confidence mask used during fine-tuning. The mask is constructed by combining geometry boundaries and shadow-projection boundaries. Blue regions indicate high-confidence supervision, while dimmer regions are excluded from training.



(a) Campus scene.

(b) BlendedMVS scenes.

Figure E.1. Retextured models using recovered albedo images from the fine-tuned RGB↔X.

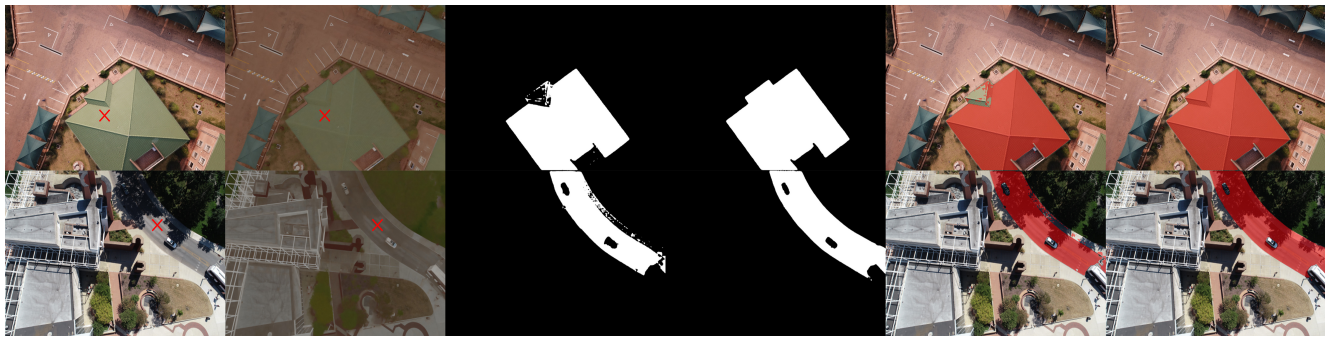
(a) RGB-textured models.

(b) Albedo-retextured models.

Figure E.2. Comparison between RGB-textured models and albedo-retextured models. Shading is largely removed while texture details are well preserved.

RGB Albedo(Ours) RGB Albedo(Ours) RGB Albedo(Ours)

Figure F.1. Visual comparisons between the RGB textures (left column of each pair) and our recovered albedo textures (right column of each pair) under novel lighting conditions. The scenes encompass diverse architectural styles, materials, and environmental conditions. Original textures contain baked-in lighting such as cast shadows, shading gradients, and environment-induced color shifts, whereas our albedo textures remove these illumination effects and retain only the intrinsic diffuse reflectance. These results highlight the stability and consistency of our method across varied outdoor environments.



Original Albedo Original Mask Albedo Mask Original Overlay Albedo Overlay

Figure G.1. Additional segmentation results using SAM on RGB and albedo images. Each group of six images shows, from left to right: the RGB image, the corresponding albedo, the SAM mask predicted on RGB, the SAM mask predicted on albedo, and the two masks overlaid on the RGB image. Across diverse scenes, masks obtained from albedo remain more stable in shadowed regions and better preserve object boundaries than those obtained from RGB.