Photometric Mesh Optimization for Video-Aligned 3D Object Reconstruction (Supplementary Material)

1. Architectural and Pretraining Details

We use AtlasNet [2] as the base network architecture for our experiments. Following Groueix *et al*. [2], the image encoder is the ResNet-18 [3] architecture where the last fully-connected layer is replaced with one with an output dimension of 1024, which is the size of the latent code. We use the 25-patch version of the AtlasNet mesh decoder, where each deformable patch is an open triangular mesh with $5^2 \times 2 = 50$ triangles on a 5×5 regular grid. We redirect the readers to Groueix *et al*. [2] for more details.

In the stage of pretraining AtlasNet on ShapeNet [1] with textured background from SUN360 [6], we train all networks using the Adam optimizer [4] with a constant learning rate of 10^{-4}. We set the batch size for all experiments to be 32. We initialize the AtlasNet encoder with the pretrained ResNet-18 on ImageNet [5] except for the last modified layer (before the latent code), and we initialize the decoder with that pretrained from a point cloud autoencoder from Groueix *et al*. [2].

2. Warp Parameterization Details

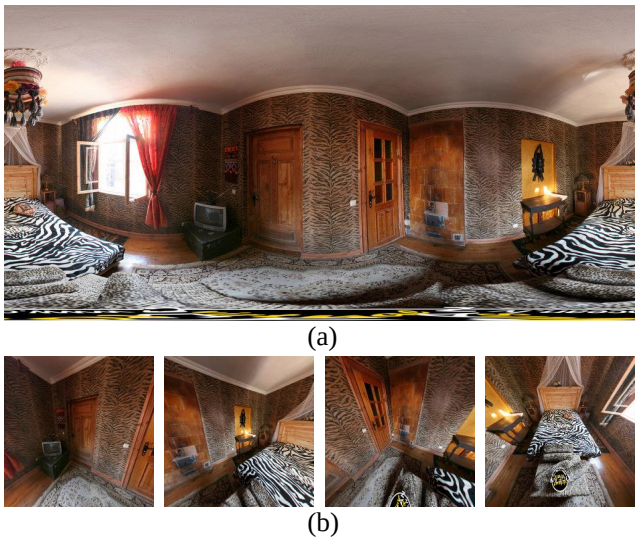
We parameterize the rotation component of 3D similarity transformations with the $\mathfrak{so}(3)$ Lie algebra. Given a warp parameter vector $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^\top \in \mathfrak{so}(3)$, the rotation matrix $\mathcal{R}(\boldsymbol{\omega}) \in \mathbb{SO}(3)$ can be written as

$$\mathcal{R}(\boldsymbol{\omega}) = \exp\left(\begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}\right) , \qquad (1)$$

where exp is the exponential map (*i.e*. matrix exponential). $\mathcal{R}$ is the identity transformation when $\boldsymbol{\omega}$ is an all-zeros vector. The exponential map is Taylor-expandable as

$$\mathcal{R}(\boldsymbol{\omega}) = \exp(\boldsymbol{\omega}_\times) = \lim_{K\to\infty} \sum_{k=0}^{K} \frac{\boldsymbol{\omega}_\times^k}{k!} . \qquad (2)$$

We implement the $\mathfrak{so}(3)$ parameterization using the Taylor approximation expression with $K = 20$. We have also tried parametrizing the 3D similarity transformations with the self-contained Lie group $\mathrm{Sim}(3)$, where the scale is incorporated into the exponential map; we find it to yield almost identical results. We also take the exponential on the scale s to ensure positivityl; the resulting scale does not change when $s = 0$.



(a) (b)

Figure 1: (a) Example panoramic (spherical) image and (b) sample cropped images at different camera viewpoints.

3. SUN360 Background Data Generation

The background images from SUN360 [6] are cropped from spherical images with a resolution of 1024×512, using a field of view of 90°. Fig. 1 illustrates an example of the original spherical image and its generated crops.

References

[1] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 1

[2] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry. Atlasnet: A papier-mâché approach to learning 3d surface generation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference*

on computer vision and pattern recognition, pages 770–778, 2016. 1

[4] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 1

[6] J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba. Recognizing scene viewpoint using panoramic place representation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2695–2702. IEEE, 2012. 1