Neural Kaleidoscopic Space Sculpting: Supplementary Material

Byeongjoo Ahn, Michael De Zeeuw, Ioannis Gkioulekas, Aswin C. Sankaranarayanan
Carnegie Mellon University

1. Introduction

In this supplementary material, we cover the following topics:

1. In Section 2, we describe the hardware setup used for the experiments in the paper.
2. In Section 3, we provide the implementations details of the proposed method, including the mask processing for visual hull baseline implementation, the effect of manual mask refinement, and the effect of initial shape. We also provide the algorithm for the implementation and discuss the limitation regarding transparent objects.
3. In Section 4, we show supplemantary results, including full results with visual hull, and the comparison to Nerfies.

2. Hardware

Figure 1 shows our hardware prototype used for the the experiments in the paper, comprising an RGB camera (FLIR Blackfly S BFS-U3-200S6C) and four triangular mirrors (Edmund Optics 46-656 customized). The mirrors are coated with metal, with the size of $200\,\mathrm{mm} \times 307\,\mathrm{mm}$ and the surface flatness of $4 - 6\lambda$.

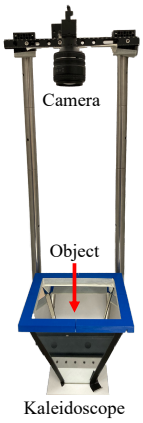


Figure 1. **Prototype.**

3. Implementation Details

Mask processing for visual hull. For the implementation of the baseline visual hull [4], we dilate the input mask to handle the false background pixels caused by the incorrect masking or the gap between mirrors. Figure 2 shows the visual hull results with different dilations. We use 8 pixels of dilation for the baseline results used in comparisons, and conservatively use 16 pixels of dilation ignoring mirror boundary for obtaining $\mathcal{H}_{\mathbf{x}}^{\mathrm{vh}}$ for the visual hull constraint. The voxel resolution of the visual hull is 128 on each axis.

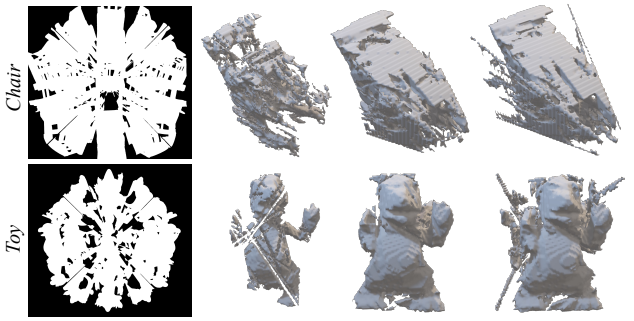


Figure 2. **Mask processing for the baseline method (kaleidoscopic visual hull [4]).** We dilate the input mask to compute the visual hull to handle the incorrect masking and the gap between mirrors.

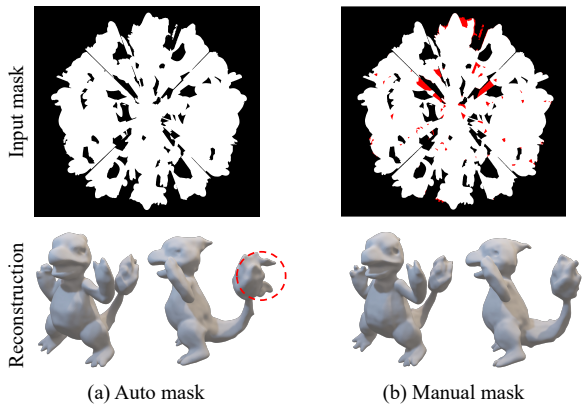
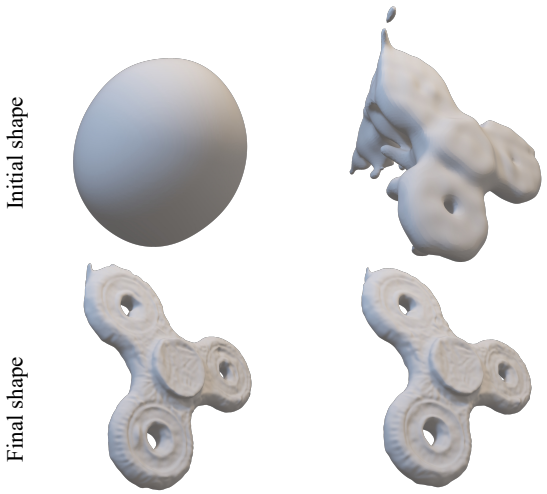


Figure 3. **Effect of manual mask refinement.** (a) Mask automatically generated from the difference between the images with and without the object. (b) Mask additionally refined using Adobe Photoshop manually.

Manual mask refinement. We obtain the input mask first by computing the difference between the images with and without the object, and then manually refining the mask using Adobe Photoshop for correcting some pixels. Figure 3 shows the effect of the manual mask refinement. Although they provide similar results overall, the manual refinement



(a) Sphere initialization (b) Visual hull initialization

Figure 4. **Effect of initial shape.** We observe the initial shape (*i.e.*, approximate unit sphere or visual hull) does not affect to the final shape significantly.

provides better results in the regions where the difference between the images with and without the object is not clear (*e.g.*, the tail of *Toy*).

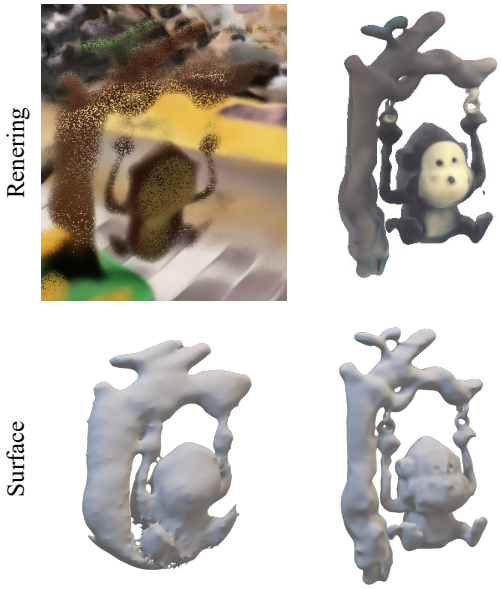
Initial shape. We use an approximate unit sphere as an initial shape [1] in the experiments. Figure 4 shows the effect of the initial shape by comparing the result with another initialization obtained by fitting the SDF to the visual hull using IGR [2]. We observe the initial shape does not affect to the final shape significantly.

Algorithm. Algorithm 1 shows how the points are selected and used for the optimization in the proposed neural kaleidoscopic space sculpting.

Limitations regarding transparent objects. We discuss the challenges in reconstructing transparent objects or objects with refraction and reflection effects. Reconstruction of such objects is challenging for many 3D reconstruction techniques that need to establish correspondences across viewpoints, and not just ours. Compared to non-kaleidoscopic techniques, our kaleidoscopic method further needs the assumption that the object is opaque in order to be able to establish unique correspondences between each pixel and a single virtual camera (i.e., a unique label for each pixel).

Algorithm 1: Neural kaleidoscopic space sculpting

Input: Kaleidoscopic image $\mathbf{I}$ and mask $\mathbf{M}$
Output: Neural network θ, ϕ for shape and texture
Initialize $\mathcal{P}_{\text{carve}} = \emptyset$, $\mathcal{P}_{\text{model}} = \emptyset$;
for $b = 0 : B_{\mathbf{x}}^{\emptyset}$ **do**
  for $\mathbf{x} \in \mathcal{X}_{\text{batch}}$ **do**
    Compute intersection b/w $\mathbf{r}_{\mathbf{x}}^{b}(t)$ and $f(\cdot;\theta)$;
    if $\mathbf{M}(\mathbf{x}) = 0$ **then**
      `// BG (Eq.(14))`
      $\mathcal{P}_{\text{carve}} \leftarrow \mathcal{P}_{\text{carve}} \cup \mathcal{P}_{\mathbf{x}}^{\text{bg}}$;
    else
      `// FG ((Eqs.(17-19)))`
      $\mathcal{P}_{\text{model}} \leftarrow \mathcal{P}_{\text{model}} \cup \mathcal{P}_{\mathbf{x}}^{\text{fg,nh}} - \mathcal{P}_{\mathbf{x}}^{\text{ovh}}$;
      $\mathcal{P}_{\text{carve}} \leftarrow \mathcal{P}_{\text{carve}} \cup \mathcal{P}_{\mathbf{x}}^{\text{ovh}}$;
      $\mathcal{P}_{\text{tex}} \leftarrow \mathcal{P}_{\text{tex}} \cup \mathcal{P}_{\mathbf{x}}^{\text{fg,h}}$;
    end
  end
end
Update θ and ϕ with loss$(\theta, \phi; \mathcal{P}_{\text{carve}}, \mathcal{P}_{\text{model}}, \mathcal{P}_{\text{tex}})$;
Repeat with different batches;

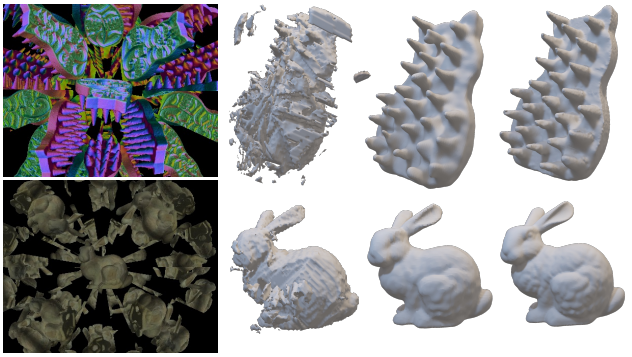


(a) Nerfies (b) Ours

Figure 5. **Comparison to Nerfies [3].** The surface extracted from Nerfies is distorted as the deformation of the dynamic object is not correctly estimated.

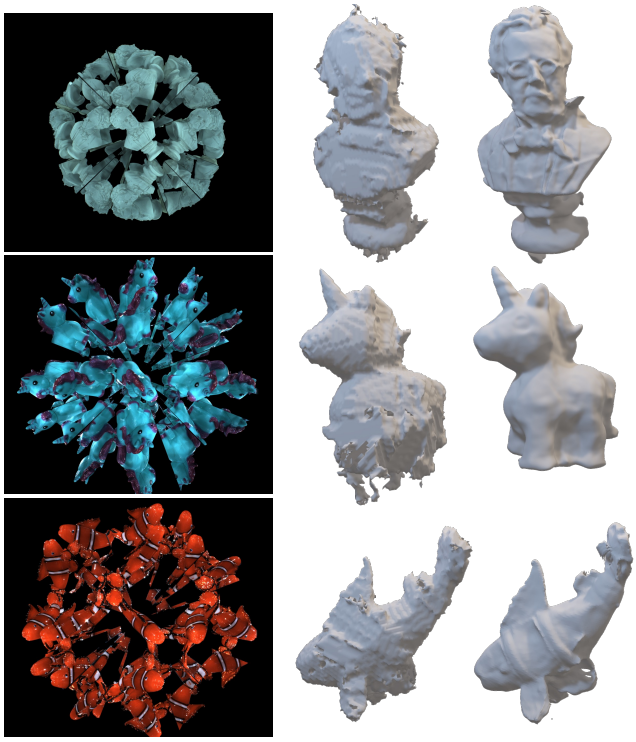
4. Supplementary results

Nerfies surface extraction. Figure 5 provides a comparison to Nerfies [3] with surface results as well as the rendering results. The Nerfies surface is obtained by creating an isosurface from the multi-view depth map results. We



(a) Kaleidoscopic image (b) Visual hull (c) Ours (d) Ground truth

Figure 6. **Comparison to visual hull [4] on synthetic data.**



(a) Kaleidoscopic image (b) Visual hull (c) Ours

Figure 7. **Comparison to visual hull [4] on real data.**

observe the shape of the monkey and the branch are distorted, which is because the Nerfies could not capture the deformation of this dynamic object correctly.

Comparison to kaleidoscopic visual hull. Figures 6 and 7 show the comparison to the baseline method of kaleidoscopic visual hull [4] for the synthetic and real data, respectively. We observe the proposed method produces superior results than the baseline method.

Full results. Figure 8 shows our full results with the comparison to kaleidoscopic visual hull [4] for all target objects. The visual hull produces an approximate shape of the object but cannot capture the details of the object (*e.g.*, the face of the bust *Venus*). It also cannot produce a reasonable shape when the input image does not contain enough background pixels (*e.g.*, *Chair*). Please zoom in to see the differences.

References

[1] Matan Atzmon and Yaron Lipman. Sal: Sign agnostic learning of shapes from raw data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[2] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *International Conference on Machine Learning*, 2020. 2

[3] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *IEEE International Conference on Computer Vision (ICCV)*, 2021. 2

[4] Ilya Reshetouski, Alkhazur Manakov, Hans-Peter Seidel, and Ivo Ihrke. Three-dimensional kaleidoscopic imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 1, 3

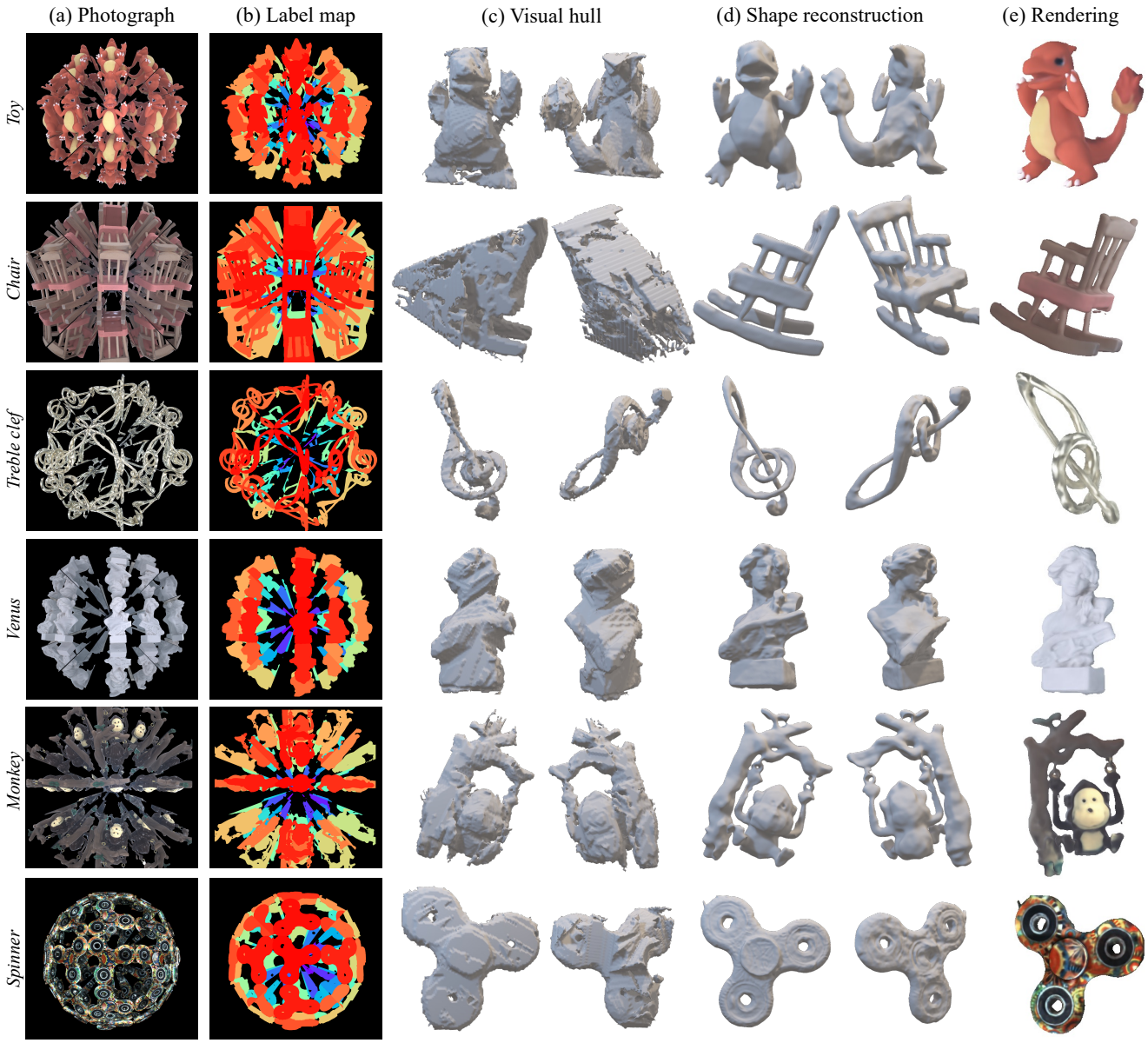


Figure 8. **Full results.** Please zoom in to see the differences.