

# Hand Shadow Art: A Differentiable Rendering Perspective

Aalok Gangopadhyay, Prajwal Singh, Ashish Tiwari & Shanmuganathan Raman

CVIG Lab, Indian Institute of Technology Gandhinagar

## Abstract

Shadow art is an exciting form of sculptural art that produces captivating artistic effects through the 2D shadows cast by 3D shapes. Hand shadows, also known as shadow puppetry or shadowgraphy, involve creating various shapes and figures using your hands and fingers to cast meaningful shadows on a wall. In this work, we propose a differentiable rendering-based approach to deform hand models such that they cast a shadow consistent with a desired target image and the associated lighting configuration. We showcase the results of shadows cast by a pair of two hands and the interpolation of hand poses between two desired shadow images. We believe that this work will be a useful tool for the graphics community.

## CCS Concepts

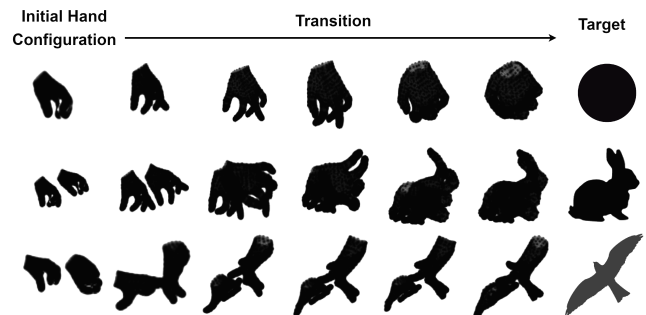
• **Computing methodologies** → Rendering; Optimization; Neural Nets;

## 1. Introduction

"Will he not fancy that the shadows which he formerly saw are truer than the objects which are now shown to him?" - Plato, the Republic.

Shadows are crucial in how we perceive the world around us. They are the silent storytellers of our world, revealing secrets that light alone cannot. For a long time, it has been gathering the attention of artists, stage performers, and filmmakers. Hand shadows and other forms of storytelling, such as ancient Chinese shadow puppetry, have been instrumental in unleashing the capacity of the human brain to recognize 3D objects through shadows. Hand shadows offer a versatile and imaginative form of shadow art that can be used for entertainment, education, artistic expression, and various cultural and creative expression forms. In this work, we attempt to answer how we can deform our hands such that the shadow thus cast matches the desired image using differentiable rendering [RRN\*20] (see Figure 2).

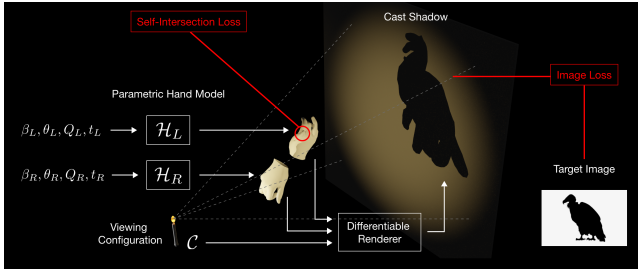
While a few recent methods have addressed shadow art using either optimization [MP09] or differentiable rendering [STR22], none of them have addressed hand shadow art. Mitra *et al.* [MP09] described shadow art more formally by introducing a voxel-based optimization framework to recover the 3D shape from arbitrary input (shadow) images by deforming them and handling inherent image inconsistencies. Sadekar *et al.* [STR22] demonstrated the potential of differentiable rendering (mesh and voxel-based) in generating 3D shadow sculptures all from arbitrary shadow images without any explicit input image deformation. In this work, we explicitly focus on developing a differential rendering-based optimization framework to create hand shadow art by deforming the hand mesh models and show the dynamics involved in interpolating between two shadow images. To the best of our knowledge, this is the first such attempt to generate hand shadow art progressively using differentiable rendering.



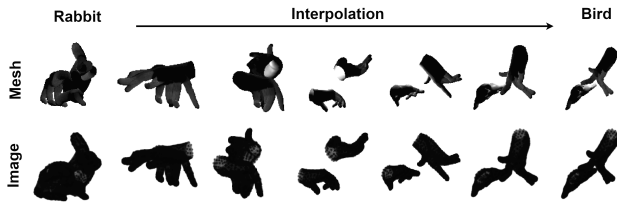
**Figure 1:** Different hand configurations to render the target shadow image during the optimization process from one hand (first row) and two hands (second and third row). The images are rendered in  $256 \times 256$  resolution.

## 2. Method

The design of the proposed framework is described in Figure 2. Let  $\mathcal{I}^* = [0, 1]^{H \times W}$  denote the space of all grayscale images defined on a grid of size  $H \times W$ . Let  $I \in \mathcal{I}^*$  be a given target image and  $\mathcal{M} = \mathcal{H}(\beta, \theta, Q, t)$  be the parametric hand model [RTB22]. Here,  $\beta$ ,  $\theta$ ,  $Q$ , and  $t$  are the shape parameter, pose parameter, rotation matrix, and translation vector, with  $\mathcal{M}$  being the mesh model of the hand for a given set of parameters. Please note that we use the parametric models for both the left and right hand given by  $\mathcal{M}_L = \mathcal{H}_L(\beta_L, \theta_L, Q_L, t_L)$  and  $\mathcal{M}_R = \mathcal{H}_R(\beta_R, \theta_R, Q_R, t_R)$ . Consider  $\mathcal{C}$  as the camera with fixed parameters placed at the origin. The lighting and camera parameters together are termed as *viewing configuration* ( $\mathcal{V}$ ). Let  $\mathcal{R}(\mathcal{C}, \mathcal{M}_L, \mathcal{M}_R)$  denote the silhouette image rendering of both the hands as seen from the camera. The objective is to find the values of parameters  $\theta_L$ ,  $\theta_R$ ,  $Q_L$ ,  $Q_R$ ,  $t_L$ , and  $t_R$  that minimize  $\|I - \mathcal{R}(\mathcal{C}, \mathcal{M}_L, \mathcal{M}_R)\|_2$ , keeping the shape param-



**Figure 2:** The outline of the proposed differentiable rendering framework to simulate hand shadow art. The framework is optimized over 5000 iterations to reach convergence.



**Figure 3:** Different hand configurations (top: hand meshes, bottom: rendered image) while interpolating from a Rabbit to a Bird silhouette. Here is a [link to a dynamic visualization](#).

ters ( $\beta_L$  and  $\beta_R$ ) fixed. The non-rigid nature of hands and fingers poses an additional challenge apart from rigid transformations *i.e.* rotations and translations and mesh intersections towards learning optimal hand configurations. To avoid self-intersections and cross-intersections among a pair of hand meshes, we penalize such intersections according to pen loss [Kar12, TBS\*16]. Further, we restrict the angular movement across 15 different joints per hand in the MANO hand models [RTB22] to simulate realistic and plausible human hand movements (see Table 1).

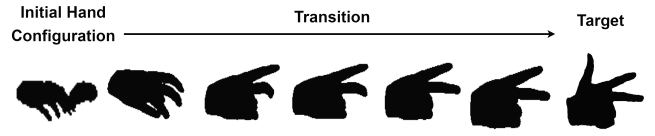
### 3. Results and Discussion

Figure 1 shows a few transitional hand configurations to cast the shadow resembling a target image and also transitions between one image to another.

**Hand shadow art.** We start with a random initial configuration of hand models to reach the final pose, creating a shadow similar to the given target image when viewed from a fixed configuration.

**Interpolation** between a pair of shadow images cast by a pair of hands is shown in Figure 3. Given a pair of images  $I_A$  and  $I_B$ , we find a sequence of hand mesh models ( $\mathcal{M}_L^t, \mathcal{M}_R^t$ ) such that  $\mathcal{R}(\mathcal{C}, \mathcal{M}_L^0, \mathcal{M}_R^0) = I_A$  and  $\mathcal{R}(\mathcal{C}, \mathcal{M}_L^T, \mathcal{M}_R^T) = I_B$ , where  $t \in [0, T]$ .

We observe that a good initial hand configuration makes the convergence faster and easier, otherwise it leads to incorrect results. For example, the model cannot capture a larger transition of the thumb to recreate the target hand configuration from a specific initial configuration (see Figure 4). In the near future, we wish to work towards adapting the algorithm to be free of specific initialization and extend the current single-view setting to a multi-view setting



**Figure 4:** Failure cases over large and abrupt transitions and sensitivity to the initial hand configuration.

Joint Name/ Locations	RH Index	RH Middle	RH Little	RH Ring	RH Thumb	LH Index	LH Middle	LH Little	LH Ring	LH Thumb
J.1 ( $\theta, \phi, \psi$ )	0.0,	0.0,	0.0,	0.0,	-50.0,	0.0,	0.0,	0.0,	0.0,	20.0,
	-15.0,	-25.0,	-40.0,	-15.0,	-10.0,	25.0,	10.0,	25.0,	15.0,	30.0,
	-35.0,	-15.0,	-35.0,	-35.0,	-30.0,	75.0,	75.0,	75.0,	75.0,	40.0,
J.2 ( $\theta, \phi, \psi$ )	0.0,	0.0,	0.0,	0.0,	-10.0,	0.0,	0.0,	0.0,	0.0,	10.0,
	0.0,	0.0,	0.0,	0.0,	-20.0,	0.0,	0.0,	0.0,	0.0,	20.0,
	-45.0,	-15.0,	-45.0,	-45.0,	-10.0,	45.0,	55.0,	55.0,	55.0,	20.0,
J.3 ( $\theta, \phi, \psi$ )	0.0,	0.0,	0.0,	0.0,	-5.0,	0.0,	0.0,	0.0,	0.0,	5.0,
	0.0,	0.0,	0.0,	0.0,	-50.0,	0.0,	0.0,	0.0,	0.0,	25.0,
	-15.0,	-45.0,	-5.0,	-15.0,	0.0,	75.0,	75.0,	65.0,	75.0,	0.0,

**Table 1:** ( $\theta, \phi, \psi$ ) represents rotation about ( $x, y, z$ ) axis and J.1, J.2, J.3 represents joints along each finger of the left (LH) and right (RH) hand.

where a single-hand configuration can cast different meaningful shadows when viewed and lit from multiple directions.

### 4. Conclusion

In this work, we used a differentiable rendering-based approach to simulate hand shadow art through optimization. While we could create a variety of shadows through hand models, including interpolation, we observed that the framework is sensitive to initial hand configuration and susceptible to errors under large and abrupt transitions, which we aim to address in the near future, including dynamic shadow art. This versatile medium can be applied to various contexts, from traditional storytelling and entertainment to contemporary art installations and experimental exhibitions, and we hope it will be helpful for the graphics community.

### References

- [Kar12] KARRAS T.: Maximizing parallelism in the construction of bvhs, octrees, and k-d trees. In *Proceedings of the Fourth ACM SIGGRAPH / Eurographics Conference on High-Performance Graphics* (2012), Eurographics Association, pp. 33–37. 2
- [MP09] MITRA N. J., PAULY M.: Shadow art. *ACM Transactions on Graphics* 28, CONF (2009), 156–1. 1
- [RRN\*20] RAVI N., REIZENSTEIN J., NOVOTNY D., GORDON T., LO W.-Y., JOHNSON J., GKIOXARI G.: Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501* (2020). 1
- [RTB22] ROMERO J., TZIONAS D., BLACK M. J.: Embodied hands: Modeling and capturing hands and bodies together. *arXiv preprint arXiv:2201.02610* (2022). 1, 2
- [STR22] SADEKAR K., TIWARI A., RAMAN S.: Shadow art revisited: a differentiable rendering based approach. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2022), pp. 29–37. 1
- [TBS\*16] TZIONAS D., BALLAN L., SRIKANTHA A., APONTE P., POLLEFEYS M., GALL J.: Capturing hands in action using discriminative salient points and physics simulation. *International Journal of Computer Vision (IJCV)* 118, 2 (June 2016), 172–193. 2