Shadow Art Revisited: A Differentiable Rendering Based Approach

Kaustubh Sadekar
CVIG Lab, IIT Gandhinagar
sadelkar.k@iitgn.ac.in

Ashish Tiwari
CVIG Lab, IIT Gandhinagar
ashish.tiwari@iitgn.ac.in

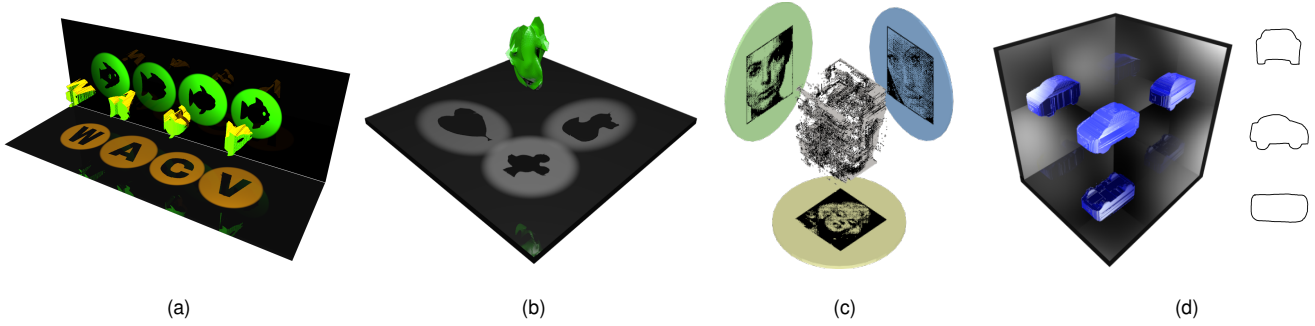Shanmuganathan Raman
CVIG Lab, IIT Gandhinagar
shanmuga@iitgn.ac.in

Figure 1: Shadow art sculptures generated using differentiable rendering casting the shadows of (a) WACV acronym on one plane and fishes on the other resembling an aquarium of floating objects, (b) dropping *Heart*, *Duck*, and *Mickey* (all on the same plane), and (c) face sketches using half-toned images. (d) 3D reconstruction of a car from hand drawn sketches.

Abstract

While recent learning based methods have been observed to be superior for several vision-related applications, their potential in generating artistic effects has not been explored much. One such interesting application is Shadow Art - a unique form of sculptural art where 2D shadows cast by a 3D sculpture produce artistic effects. In this work, we revisit shadow art using differentiable rendering based optimization frameworks to obtain the 3D sculpture from a set of shadow (binary) images and their corresponding projection information. Specifically, we discuss shape optimization through voxel as well as mesh-based differentiable renderers. Our choice of using differentiable rendering for generating shadow art sculptures can be attributed to its ability to learn the underlying 3D geometry solely from image data, thus reducing the dependence on 3D ground truth. The qualitative and quantitative results demonstrate the potential of the proposed framework in generating complex 3D sculptures that go beyond those seen in contemporary art pieces using just a set of shadow images as input. Further, we demonstrate the generation of 3D sculptures to cast shadows of faces, animated movie characters, and applicability of the framework to sketch-based 3D reconstruction of underlying shapes.

1. Introduction

According to an ancient Roman author, Pliny the Elder, *the very art of painting originates from trailing the edge of shadow*. If art can be defined as creating visual or auditory elements that express the author's imaginative or technical skill, then shadow art represents those skills in play with shadows. Most of us have seen or atleast heard of "someone" making "something" interesting out of shadows. However, it is usually limited to people playing around with their fingers around a lamp, making shadows of rabbits or horses on the wall. In this work, we show how differentiable rendering can be used to generate some amazing 3D sculptures which cast mind-boggling shadows when lit from different directions.

Figure 2 (a) shows the cover of the book *Gödel, Escher, Bach* by *Douglas Hofstadter* that features blocks casting shadows of different letters when seen from different sides. *Kumi Yamashita* - one of the most prominent contemporary artists - demonstrated that seemingly simple objects arranged in a certain pattern cast startling silhouettes when lit from just the right direction. An exclamation point becomes a question mark when lit from its side (Figure 2 (b)) and a bunch of aluminum numbers placed in a 30-story office add up to an image of a girl overlooking the crowd below (Figure 2 (c)). All of these, and other pieces made by *Kumi Yamashita* not only please our eyes, but also inspire emo-

1

tion and pose intriguing questions. *Tim Noble* and *Sue Webster* have been presenting this type of artwork since 1997, creating projected shadows of people in various positions (Figure 2 (d)). This specifically arranged ensemble shows how readily available objects can cast the clearest of illusions of clearly recognizable scenes (Figure 2(e)). Figure 2 (f) shows the aquarium of floating characters by *Shigeo Fukuda* where the shadows of the fish reveal their names in kanji characters. Even after such fascinating effects, the current state of shadow art seems to be well described by Louisa May Alcott, who says *"Some people seemed to get all sunshine, and some all shadow..."*. Shadow art was first introduced to the vision and graphics community by [9] where they formally addressed the problem in an optimization framework. Since then, no significant progress has been observed in this direction.

(a) Godel, Esher, Bach   (c) City view   (e) Sunset over Manhattan

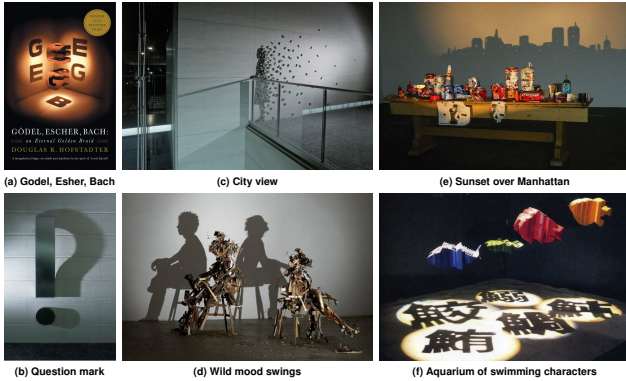(b) Question mark   (d) Wild mood swings   (f) Aquarium of swimming characters

Figure 2: Examples of shadow art sculptures by (a) *Douglas Hofstadter*, (b, c) *Kumi Yamashita*, (d, e) *Tim Noble* and *Sue Webster*, and (f) *Shigeo Fukuda*.

The question: Can we develop a method that learns to create or optimize to 3D sculptures that can generate such artistic effects through their shadows? In this work, we attempt to answer this question through the use of *Differentiable Rendering*. Here, instead of trying to realistically render a scene of our creation, we try to reconstruct a representation of a scene from one or more images of a scene [2]. Our work is mostly inspired by examples of shadow art shown in Figure 2. Specifically, our objective is to generate 3D shadow art sculptures that cast different shadows (of some recognizable objects) when lit from different directions using a differentiable renderer.

Why differentiable rendering? Most learning-based methods for 3D reconstruction require accurate 3D ground truths as supervision for training. However, all we have is a set of desired shadow images in our case. Differentiable rendering based methods require only 2D supervision in the form of single or multi-view images to estimate the underlying 3D shape, thus, eliminating the need for any 3D data collection and annotation.

Contributions. The following are the major contributions of this work.

- We introduce the creation for 3D shadow art sculptures that cast different shadows when lit from different directions using differentiable rendering just from the input shadow images and the corresponding projection information.

- We demonstrate the efficacy of deploying differentiable rendering pipeline over voxel and mesh based representations to generate shadow art sculptures.

- We show that the proposed framework can create artistic effects that go beyond those seen in contemporary art pieces by generating 3D sculptures using halftoned face images and its sketches drawn from multiple viewpoints.

- To the best of our knowledge, ours is the first work to address shadow art using differentiable rendering.

Organization. We start by describing the literature covering the relevant related works in Section 2. We discuss the problem statement more formally and describe both voxel and mesh based differentiable rendering pipelines in Section 3. Section 3.4.1 describes the loss functions deployed for optimization. In Section 4, we perform qualitative and quantitative analysis of results and compare the performance of the proposed framework with that of the state-of-the-art method. Section 5 describes interesting artistic effects and applications of shadow art before we conclude the work in Section 6.

2. Related Work

Shadows play an essential role in the way we perceive the world and have been central in capturing the imagination of many artists including stage performers. Several artists have typically used manual, trial-and-error style approaches to create 3D shadow sculptures. However, with the advent of digital design technology, the need of an automated framework is inevitable.

Shadow Art. Shadows in many computer graphics and computer vision applications have been studied from both perceptual (artist's) and mathematical (programmer's) point of views. It started with studying the effect of shadow quality on perception of spatial relationships in a computer generated image [21, 22]. Pellacini *et al.* developed an interface for interactive cinematic shadow design that allows the user to modify the positions of light sources and shadow blockers by specifying constraints on the desired shadows [16]. The idea of computing the shadow volume from a set of shadow images evolved after that. This is similar

to the construction of a visual hull used for 3D reconstruction. Visual hull is the intersection volume of a set of generalized cones constructed from silhouette images and the corresponding camera locations [3]. Sinha and Polleyfeys [18] studied the reconstruction of closed continuous surfaces from multiple calibrated images using min-cuts with strict silhouette constraints.

Relation with the state-of-the-art method. The work closest to ours is by Mitra *et al.* [9]. They described shadow art more formally by introducing a voxel-based optimization framework to recover the 3D shape from arbitrary input images by deforming the input shadow images and handled inherent image inconsistencies. In this work, we demonstrate the potential of differentiable rendering in generating 3D shadow sculptures all from the arbitrary shadow images without any explicit input image deformation. Although the associated 3D object might not exist in the real-world, but the method still creates shadow sculptures that go beyond those seen in contemporary art pieces casting the physically realizable shadows when lit from appropriate directions.

Differentiable Rendering. We briefly review methods that learn the 3D geometry via differentiable rendering. These methods are categorized based on the underlying representation of 3D geometry: point clouds, voxels, meshes, or neural implicit representation. In this work, we primarily focus on voxel and mesh based representations.

Several methods operate on voxel grids [7, 12, 15, 20]. Paschalidou *et al.* [15] and Tulsiani *et al.* [20] propose a probabilistic ray potential formulation. Although they provide a solid mathematical framework, all intermediate evaluations need to be saved for backpropagation. This limits these approaches to relatively small-resolution voxel grids. On one hand, Sitzmann *et al.* [19] have inferred implicit scene representations from RGB images via an LSTM-based differentiable renderer and Liu *et al.* [6] perform max-pooling over the intersections of rays with a sparse number of supporting regions from multi-view silhouettes. On the other hand, [13] show that volumetric rendering is inherently differentiable for implicit representations and hence, no intermediate results need to be saved for the backward pass. OpenDR [8] roughly approximates the backward pass of the traditional mesh-based graphics pipeline. Liu *et al.* [5] proposed Soft Rasterizer to replace the rasterization step with a soft version to make it differentiable using a deformable template mesh for training and yields compelling results in reconstruction tasks. We deploy this in our mesh-based differentiable rendering pipeline for rasterization.

Both voxel and mesh based representations have their own strength and weaknesses. In this work, we describe the differentiable rendering optimization framework for both these 3D representation and discuss which model fits the best for different scenarios to create plausible shadow art sculptures

3. Method

3.1. Problem Formulation

The key idea of our work is to generate an artistic 3D sculpture $\mathcal{S}$ that casts N different shadows when lit from N different directions using differentiable rendering based optimization pipeline. The prime focus here is to create interesting shadow art effects using the 3D sculpture $\mathcal{S}$. The input to the pipeline is a set $\mathcal{X} = \{X_1, X_2, ..., X_N\}$ of shadow configuration $X_i = (I_i, P_i)$. I_i represents the target shadow image and P_i is the corresponding projection information.

The shadow of an object can be regarded as its projection on a planar surface. Assuming directional lighting, this projection is an *orthographic projection* when the surface is perpendicular to the lighting direction and a *perspective projection*, otherwise [1]. Obtaining shadow of an object is equivalent to finding the corresponding silhouette captured by a camera pointing in the same direction as the light source. Therefore, I_i the shadow image, is essentially a silhouette. From here on, we shall use the term silhouette images and shadow images, interchangeably.

The shadow art problem is similar to a multi-view 3D reconstruction problem [4, 10], where we try to estimate the 3D structure of an object given its N silhouette views. However, the key differences in shadow art are: (i) the N views can correspond to arbitrary silhouettes (not necessarily of the same object) and (ii) the learned 3D sculpture may bear no resemblance with any real-world object and just be an abstract art that casts the desired shadows when lit from appropriate directions. Undoubtedly, there exist multiple 3D shapes that can cast the same set of shadows. However, our concern is just to learn one such 3D sculpture that can create the desired artistic effects through its shadows.

3.2. System Overview

By providing shadow configuration $\mathcal{X} = \{X_i = (I_i, P_i)|i = 1, 2, ..., N\}$ as input to the pipeline, the objective is to learn the underlying 3D sculpture $\mathcal{S}$, as described earlier. The projection information P_i corresponds to the camera position (and hence, the light source position) associated with i^{th} shadow image I_i such that $P_i = (\mathbf{R}_i, \mathbf{t}_i)$. Here, $\mathbf{R}_i$ and $\mathbf{t}_i$ are the 3D rotation and translation of the camera, respectively. We start by initialising $\mathcal{S}$ with a standard geometry which is further optimized by minimizing image-based losses, such that the rendered silhouette images $\widetilde{I}_i = I_i$ for all $i = 1, 2, ..., N$. The prime reason for using differentiable rendering is that it allows gradient flow directly from images back to parameters of $\mathcal{S}$ to optimize it in an iterative fashion. In other words, it does not require any explicit 3D supervision and optimizes the 3D shape solely from image based losses. For further simplicity, let the set of target shadow images and the associated projec-
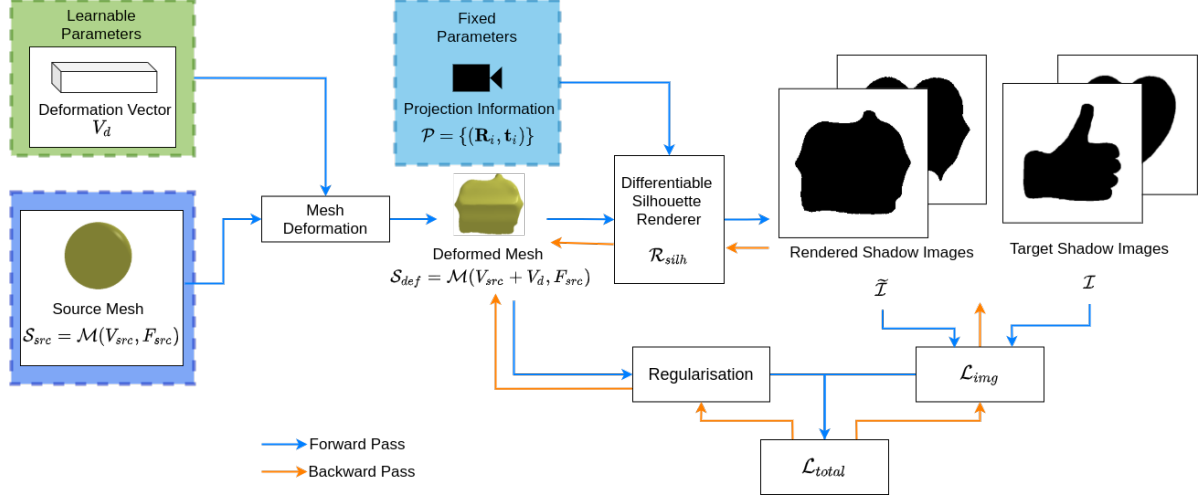
Figure 3: Information flow in the proposed mesh-based differentiable rendering pipeline.

tion information be denoted as $\mathcal{I}$ and $\mathcal{P}$, respectively, such that $\mathcal{I} = \{I_1, I_2, ..., I_N\}$ and $\mathcal{P} = \{P_1, P_2, ..., P_N\}$. Further, let $\widetilde{\mathcal{I}} = \{\widetilde{I}_1, \widetilde{I}_2, ..., \widetilde{I}_N\}$ be the set of shadow images obtained from learned 3D sculpture $\mathcal{S}$ as per projections $\mathcal{P}$.

In this work, we consider two common representations for 3D shapes i.e. *voxel* and *mesh* based representations. In the following section, we elaborate the optimization pipelines for voxel and mesh based representations of the 3D object to create visually plausible shadow art using differentiable rendering.

3.3. Voxel Based Optimization

In this section, we look at a differentiable rendering pipeline that uses voxels to represent the 3D geometry. A voxel is a unit cube representation of a 3D space. The 3D space is quantized to a grid of such unit cubes. It is parameterized by an N-dimensional vector containing information about the volume occupied in 3D space. Additionally, it encodes occupancy, transparency, color, and material information. Even though occupancy and transparency probabilities (in the range $[0, 1]$) are different, they can be interpreted in the same way in order to maintain differentiability during the ray marching operation [2]. A typical rendering process involves collecting and aggregating the voxels located along a ray and assigning a specific color to each pixel based on the transparency or the density value. All the voxels that are located along a ray projecting to a pixel are taken into account when rendering that pixel. However, our objective is to do the inverse, i.e., to find the 3D geometry associated with silhouettes corresponding to different directions.

We assume that the 3D object $\mathcal{S}$ is enclosed in a 3D cube of known size centered at the origin. Hence, $\mathcal{S}$ can be defined by a learnable 3D tensor V that stores the density values for each voxel. We initialize V with all ones. The

color value for each voxel is set to 1 and is kept fixed in the form of a color tensor C. Next, we render $\mathcal{S}$ using a differentiable volumetric rendering method described in [17]. To restrict the voxel density values to the range $[0, 1]$, V is passed through a sigmoid activation function (σ) to obtain $\widetilde{V}$, as described in Equation 1.

$$\widetilde{V} = \sigma(V) \tag{1}$$

We then pass $\widetilde{V}$ through the differentiable volume renderer $\mathcal{R}_{vol}$ along with the fixed color tensor C and the associated projection information $\mathcal{P}$ to obtain the set of corresponding rendered images $\widetilde{I}$, as described in Equation 2.

$$\widetilde{\mathcal{I}} = \mathcal{R}_{vol}(\widetilde{V}, C, \mathcal{P}) \tag{2}$$

The voxel densities V are optimized by minimizing the image level loss between a set of rendered shadow images $\widetilde{\mathcal{I}}$ and the corresponding target shadows in $\mathcal{I}$. The image level loss $\mathcal{L}_{img}$ is a weighted combination of L_1 and L_2 losses, as described in Equation 3.

$$\mathcal{L}_{img} = \lambda_1 \mathcal{L}_{L_1} + \lambda_2 \mathcal{L}_{L_2} \tag{3}$$

Here, $\lambda_1 = 10.0$ and $\lambda_2 = 10.0$ are the weights associated with L_1 and L_2 losses, respectively. The resulting voxel based representation of $\mathcal{S}$ can finally be converted to a 3D mesh making it suitable for 3D printing. One simple way to achieve this is by creating faces around each voxel having density greater than a certain threshold value (as described in [17]).

3.4. Mesh Based Optimization

In this section, we also propose to use mesh based differentiable rendering to meet our objective. The entire work-

flow is described in Figure 3. The 3D object $\mathcal{S}$ can be represented as a mesh $\mathcal{M}(V, F)$. Here, V is a set of vertices connected by a set of triangular faces F that define the surface of $\mathcal{S}$.

We start by initializing a source mesh $\mathcal{S}_{src} = \mathcal{M}(V_{src}, F_{src})$ with an icosphere consisting of $|V_{src}|$ vertices and $|F_{src}|$ faces. The idea is to learn the per-vertex displacements V_d to deform $\mathcal{S}_{src}$ to the final desired mesh that casts desired shadows (silhouettes), when lit from appropriate directions. This is achieved by rendering the deformed mesh $\mathcal{S}_{def} = \mathcal{M}(V_{def}, F_{def})$ through a mesh-based differentiable silhouette renderer $\mathcal{R}_{silh}$ (as described in [17]) from the associated projection $\mathcal{P}$ such that,

$$\begin{aligned} V_{def} &= V_{src} + V_d \\ F_{def} &= F_{src} \\ \widetilde{\mathcal{I}} &= \mathcal{R}_{silh}(\mathcal{S}_{def}, \mathcal{P}) \end{aligned} \qquad (4)$$

3.4.1 Loss Function

The source mesh is optimized by minimizing image level loss $\mathcal{L}_{img}$ (described in Equation 3), normal consistency loss, and imposing Laplacian and edge length regularisation.

Normal consistency. We use normal consistency loss to ensure smoothness in the resulting 3D sculpture. For a mesh $\mathcal{M}(V, F)$, let $e = (\mathbf{v}_x, \mathbf{v}_y)$ be the connecting edge of two neighboring faces $f_x = (\mathbf{v}_x, \mathbf{v}_y, \mathbf{a})$ and $f_y = (\mathbf{v}_x, \mathbf{v}_y, \mathbf{b})$, such that $f_x, f_y \in F$ with normal vectors $\mathbf{n}_x$ and $\mathbf{n}_y$, respectively. If $\widetilde{\mathcal{E}}$ is the set of all such connecting edges e and $|F|$ is the total number of faces in mesh, the normal consistency over all such neighbouring faces f_x and f_y is given as per Equation 5.

$$\mathcal{L}_{norm} = \frac{1}{|F|} \sum_{e \in \widetilde{\mathcal{E}}} (1 - cos(\mathbf{n}_x, \mathbf{n}_y)) \qquad (5)$$

where,

$$\begin{aligned} \mathbf{n}_x &= (\mathbf{v}_y - \mathbf{v}_x) \times (\mathbf{a} - \mathbf{v}_x) \\ \mathbf{n}_y &= (\mathbf{b} - \mathbf{v}_x) \times (\mathbf{v}_y - \mathbf{v}_x). \end{aligned}$$

Laplacian regularisation. In order to prevent the model from generating large deformations, we impose uniform Laplacian smoothing [11], as described by Equation 6.

$$\mathcal{L}_{lap} = \frac{1}{|V|} \sum_{i=1}^{|V|} \left(\left\| \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} w_{ij} \mathbf{v}_j - \mathbf{v}_i \right\|_1 \right) \qquad (6)$$

Here, $|V|$ is the number of vertices in the mesh $\mathcal{M}$ and $\mathcal{N}(\mathbf{v}_i)$ is the neighbourhood of vertex $\mathbf{v}_i$.

$$w_{ij} = \frac{\omega_{ij}}{\sum_{k \in \mathcal{N}(i)} \omega_{ik}}$$

For uniform Laplacian smoothing, $\omega_{ij} = 1$, if $(\mathbf{v}_i, \mathbf{v}_j)$ form an edge, $\omega_{ij} = -1$ if $i = j$, and $\omega_{ij} = 0$, otherwise.

Edge length regularisation. Edge-length regularisation is included to prevent the model from generating flying vertices and is given by Equation 7.

$$\mathcal{L}_{edge} = \sum_{i=1}^{|V|} \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} \| \mathbf{v}_i - \mathbf{v}_j \|_2^2 \qquad (7)$$

Finally, the overall loss function is as described in Equation 8.

$$\mathcal{L}_{total} = \lambda_a \mathcal{L}_{img} + \lambda_b \mathcal{L}_{norm} + \lambda_c \mathcal{L}_{lap} + \lambda_c \mathcal{L}_{edge} \qquad (8)$$

Here, $\lambda_a = 1.6$, $\lambda_b = 2.1$, $\lambda_c = 0.9$, and $\lambda_d = 1.8$ are the weights associated with the losses $\mathcal{L}_{img}$, $\mathcal{L}_{norm}$, $\mathcal{L}_{lap}$, and $\mathcal{L}_{edge}$, respectively.

3.5. Implementation Details

The aforementioned differentiable rendering pipelines are implemented using Pytorch3D [17]. As an initialisation for mesh, we use a level 4 icosphere composed of 2,562 vertices and 5,120 faces. For the voxel based rendering pipeline, we assume that the object is inside a cube (a grid of $128 \times 128 \times 128$ voxels) centered at origin with side of length 1.7 world units. We train the optimization pipeline with custom silhouette images of size 128×128 for 2000 epochs. We keep the learning rate to 1×10^{-4}. We keep the learning rate to 1×10^{-2} and train the optimization pipeline for 500 epochs. The training is performed on NVIDIA Quadro RTX 5000 with 16 GB memory.

4. Experimental Analysis

In this section, we perform an extensive analysis over the results obtained using voxel and mesh based differentiable rendering pipelines to create plausible shadow art effects. We start by discussing the evaluation metrics and perform ablation studies to understand the effect of various loss terms in the design.

4.1. Evaluation Metrics

Following our discussion in Section 3.1, we assess the quality of silhouettes (shadow images) obtained through the 3D sculpture $\mathcal{S}$ as per projections $\mathcal{P}$. To compare the rendered silhouette images with the target silhouette images (representing shadows), we use Intersection over Union (IoU) and Dice Score (DS). Additionally, we need to quantify the quality of the 3D sculpture $\mathcal{S}$ obtained after optimization. While we do not have any ground truth for 3D shape, and this is an optimization framework, we need a "no reference" quality metric. Therefore, we decided to use normal consistency evaluated over $\mathcal{S}$ to assess the quality of the mesh.

Mesh 3D Sculpture | Voxel 3D Sculpture | Mesh 3D Sculpture | Voxel 3D Sculpture | Mesh 3D Sculpture | Voxel 3D Sculpture | Mesh 3D Sculpture | Voxel 3D Sculpture

Normal Consistency: 0.9629 | Normal Consistency: 0.7548 | Normal Consistency: 0.9007 | Normal Consistency: 0.7095 | Normal Consistency: 0.9642 | Normal Consistency: 0.7495 | Normal Consistency: 0.9634 | Normal Consistency: 0.6551

Target Shadows

Mitra et al.

(a)
- IOU = 0.8827, DS = 0.9396 | IOU = 0.8975, DS = 0.9460

(b)
- IOU = 0.8495, DS = 0.9186 | IOU = 0.8526, DS = 0.9205 | IOU = 0.8933, DS = 0.9436

(c)
- IOU = 0.9281, DS = 0.9627 | IOU = 0.9216, DS = 0.9592 | IOU = 0.9304, DS = 0.9639

(d)
- IOU = 0.9146, DS = 0.9554 | IOU = 0.8711, DS = 0.9311 | IOU = 0.8155, DS = 0.9149

Rendered Shadows — Mesh based
- (a) IOU = 0.9787, DS = 0.9882 | IOU = 0.9855, DS = 0.9940
- (b) IOU = 0.9028, DS = 0.9489 | IOU = 0.9284, DS = 0.9629 | IOU = 0.9570, DS = 0.9780
- (c) IOU = 0.9790, DS = 0.9893 | IOU = 0.9788, DS = 0.9893 | IOU = 0.9766, DS = 0.9881
- (d) IOU = 0.9669, DS = 0.9832 | IOU = 0.9120, DS = 0.9540 | IOU = 0.9738, DS = 0.9867

Voxel based
- (a) IOU = 0.9613, DS = 0.9803 | IOU = 0.9819, DS = 0.9909
- (b) IOU = 0.9484, DS = 0.9735 | IOU = 0.9636, DS = 0.9814 | IOU = 0.9508, DS = 0.9748
- (c) IOU = 0.9603, DS = 0.9797 | IOU = 0.9636, DS = 0.9814 | IOU = 0.9508, DS = 0.9748
- (d) IOU = 0.9743, DS = 0.9870 | IOU = 0.9505, DS = 0.9746 | IOU = 0.9727, DS = 0.9862
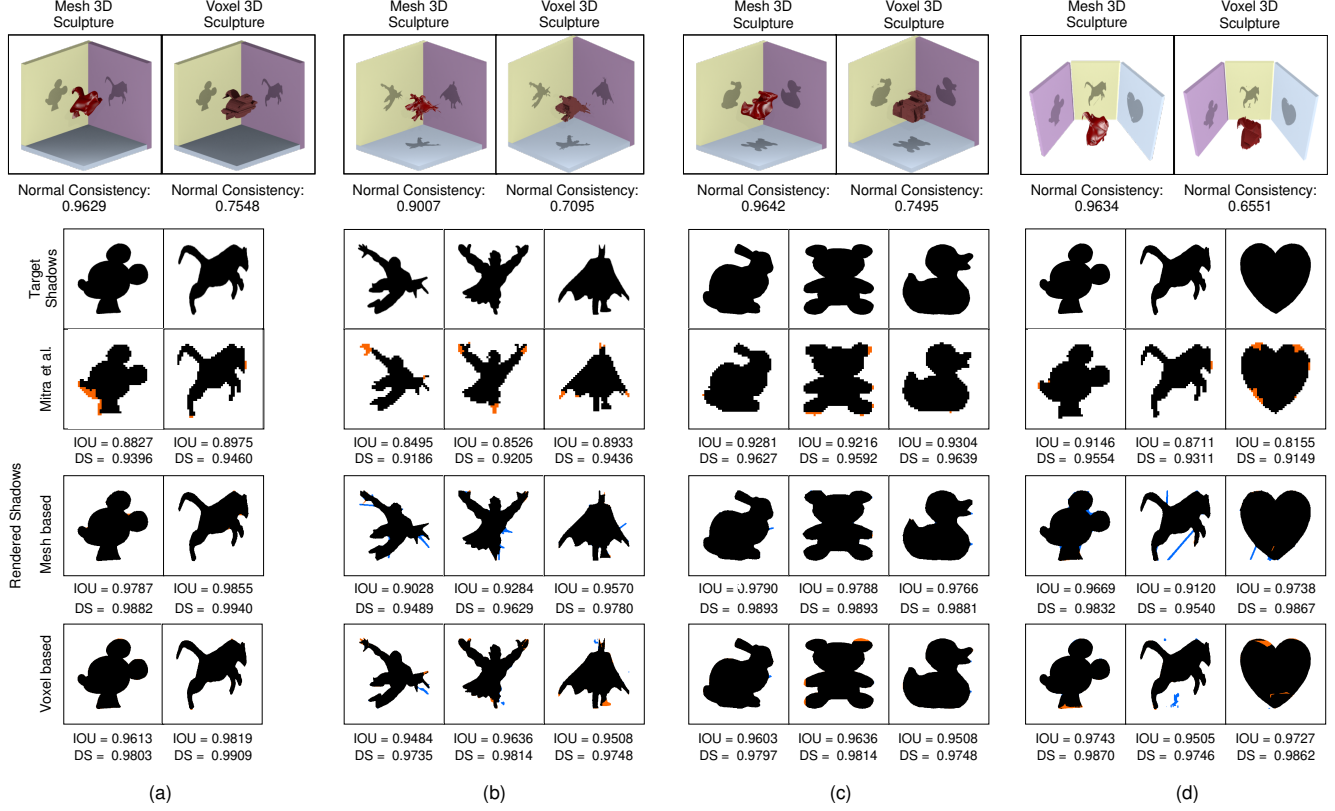
(a)   (b)   (c)   (d)

Figure 4: Qualitative and quantitative results on (a) two views (b,c) three orthogonal views, and (d, e) three non-orthogonal views using voxel and mesh-based rendering for shadow art.

4.2. Ablation Study

Figure 5 depicts the qualitative effect of different loss terms used in the optimization pipeline. The underlying mesh in this figure corresponds to the arrangement shown in Figure 4 (c). The image based loss $\mathcal{L}_{img}$ alone is not sufficient for generating plausible 3D sculptures as they are expected to suffer from distortions due to flying vertices (spike-like structures in Figure 5 (a)) or large deformations. Since we do not have any ground truth for explicit 3D supervision, we examine the effect of including regularisation in the objective function. Figure 5 (b) shows that the spikes are reduced by introducing edge-length regularisation. Further, as shown in Figure 5 (c), Laplacian smoothing prevents the sculpture from experiencing super large deformations. Finally, normal consistency loss ensures further smoothness in the optimized surface. Figure 5 (d) shows the result obtained by applying all the aforementioned regularization terms along with the image based loss. The resulting quality of the mesh validates our choice of loss terms.

4.3. Qualitative and Quantitative Analysis

In this section, we perform the qualitative and quantitative evaluation on a wide variety of shadow images includ-
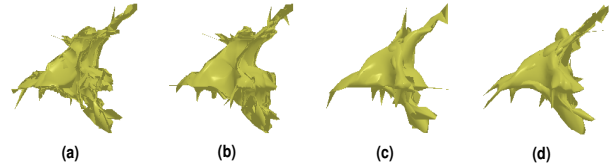
(a)   **(b)**   **(c)**   **(d)**

Figure 5: Qualitative analysis of effect of various loss terms. (a) $\mathcal{L}_{img}$, (b) $\mathcal{L}_{img} + \mathcal{L}_{edge}$, (c) $\mathcal{L}_{img} + \mathcal{L}_{edge} + \mathcal{L}_{lap}$, and (d) $\mathcal{L}_{img} + \mathcal{L}_{edge} + \mathcal{L}_{lap} + \mathcal{L}_{norm}$.

ing those used in [9] to illustrate the versatility of our approach in generating 3D shadow art sculptures represented using both voxels and mesh. For every result in Figure 4 (a)-(d), we show the learned 3D sculptures (voxel and mesh based) along with the respective shadows casted from different specified directions. We could not include the optimized 3D sculpture from [9] as the associated object file was not downloadable through their optimization tool. We have been able to incorporate both orthogonal (Figure 4 (a, b, c)) and non-orthogonal views (Figure 4 (d) and Figure 1 (b)) to obtain the shadows that are consistent with the desired target shadow images. For a quantitative comparison,
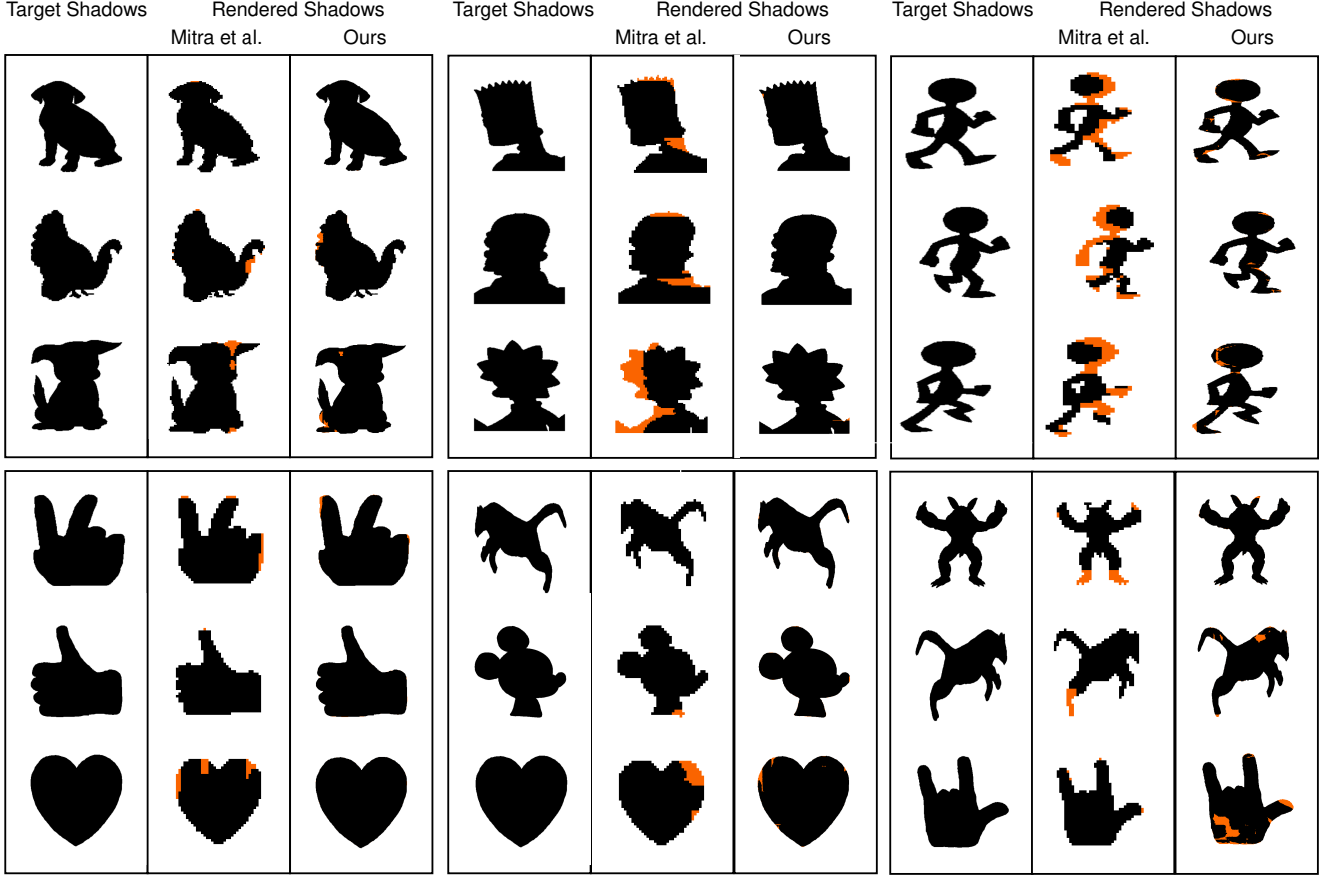
Figure 6: Qualitative evaluation of results obtained through (A) shadow art tool in [9] and (B) our voxel based rendering pipeline. The inconsistencies are highlighted in orange color.

we also report IoU and Dice score. As depicted in Figure 4, the IoU and Dice Score are comparable for both voxel and mesh based renderings. However, the corresponding voxel based 3D sculptures are not that smooth (low normal consistency value) when compared to those of mesh based 3D sculptures. It is important to note that the underlying voxel representation has been converted to a mesh representation to compute normal consistency values. While [9] have focused only on foreground inconsistencies (marked in orange color), we also show the background inconsistencies (marked in blue color) that appear in some of rendered shadow images. Ours is an end-to-end optimization approach without any additional editing tool to prune the generated 3D sculpture. In some cases, the mesh based approach is found to produce certain discontinuities near non-convex regions (Figure 4 (b,d)) for atleast one view. This is mainly attributed to the inability of icosphere to handle sharp discontinuities in the desired shape, especially when regularisation has been imposed (Equation 8). The voxel based approaches may contain a few outliers (voxels outside the desired 3D shape, as marked in blue in Figure 4 (d))

which is generally not the case with mesh based approaches. However, the mesh based differentiable rendering method lags in handling sharp discontinuities and holes present in the shadow images. While these shortcomings are handled effectively by voxel based methods, they tend to generate discretized 3D sculpture and are often associated with high memory and computational requirements. Overall, the differentiable rendering based optimization for both the approaches has been able to generate plausible 3D shadow art sculptures and is observed to have outperformed [9] in handling shadow inconsistencies by a large extent without having to explicitly deform the desired shadow images.

4.4. Comparison with the State-of-the-art method

We show the qualitative comparison of the results obtained using our voxel based differentiable rendering pipeline and the voxel based optimization tool presented in [9] without any deformation to the target shadow image. In Figure 6, we observe that the shadows rendered using the proposed pipeline are highly consistent with that of the desired target shadows when compared to those produced
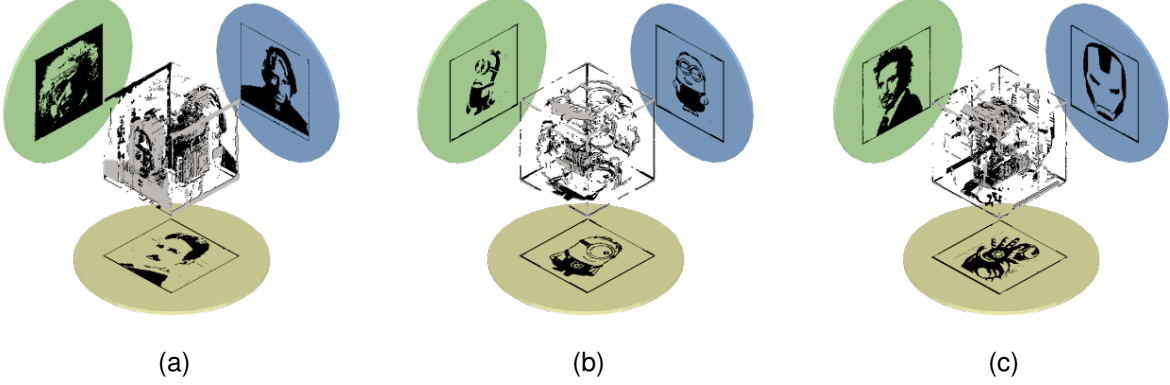
Figure 7: A seemingly random voxel soup creates three distinct shadow images of (a) *Albert Einstein*, *Nikola Tesla*, and *APJ Abdul Kalam*, (b) *Minions*, and (c) *Ironman*.
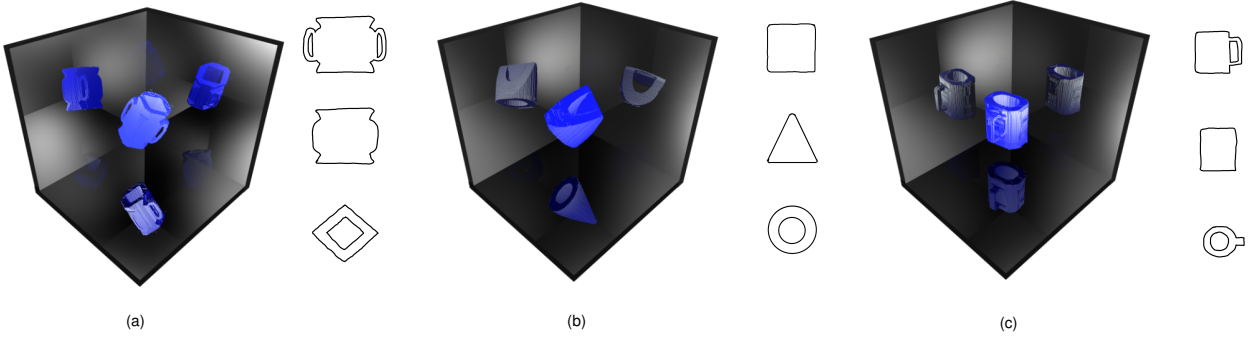

Figure 8: 3D reconstruction of (a) flower vase, (b) pen-stand, and (c) coffee mug using the associated hand drawn sketches from three different views.

by [9]. The authors of [9] argue that finding a consistent configuration for a given choice of input images might be impossible and hence, propose to introduce deformation in the input image so as to achieve consistency of the rendered shadow images with the desired ones. However, the differentiable rendering based optimization can handle inconsistencies without causing any explicit change in the target shadow images.

5. Applications

In this section, we show some additional artistic shadow art creations and an extension to yet another application that can also benefit from our optimization approach. Figure 7 depicts the creation of faces of well known scientists around the world and movie characters like *Minions* and *Ironman*, demonstrating the strength of differentiable rendering based optimization approach to handle complex objects or scenes with consistency. In addition to the binary silhouette images, half-toned images can also be used to generate 3D shadow art sculptures, as shown in Figure 1. Another interesting extension is towards sketch-based modeling [14] where we use hand-drawn sketches of a shape from differ-

ent viewpoints to automatically create the underlying 3D object. We demonstrate the creation of a flower vase (Figure 8 (a)), pen-stand (Figure 8 (b)), and a coffee mug (Figure 1 (c)) solely from hand-drawn sketches from three different views.

6. Conclusion

We have introduced an optimization framework for generating 3D shadow art sculptures from a set of shadow images and the associated projection information. The key idea is to explore the strength of differentiable rendering in creating visually plausible and consistent shadows of rigid objects, faces, and animated movie characters by generating the associated 3D sculpture. We have discussed both voxel and mesh-based rendering pipelines and have identified the benefits of each of them for the task at hand. Additionally, we have demonstrated the applicability of the proposed framework in reconstructing 3D shapes using their sketches drawn from three different viewpoints. At present, we have primarily considered the shadows that are associated with static sculptures and hence, themselves are static in nature. Dynamic shadow art can also be explored in near future.

References

[1] W. Abbott. *The Projection of Shadows*, pages 336–342. Springer Netherlands, Dordrecht, 1971.

[2] Hiroharu Kato, Deniz Beker, Mihai Morariu, Takahiro Ando, Toru Matsuoka, Wadim Kehl, and Adrien Gaidon. Differentiable rendering: A survey. *arXiv preprint arXiv:2006.12057*, 2020.

[3] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on pattern analysis and machine intelligence*, 16(2):150–162, 1994.

[4] Jinho Lee, Baback Moghaddam, Hanspeter Pfister, and Raghu Machiraju. Silhouette-based 3d face shape recovery. 2003.

[5] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7708–7717, 2019.

[6] Shichen Liu, Shunsuke Saito, Weikai Chen, and Hao Li. Learning to infer implicit surfaces without 3d supervision. *arXiv preprint arXiv:1911.00767*, 2019.

[7] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *arXiv preprint arXiv:1906.07751*, 2019.

[8] Matthew M Loper and Michael J Black. Opendr: An approximate differentiable renderer. In *European Conference on Computer Vision*, pages 154–169. Springer, 2014.

[9] Niloy J Mitra and Mark Pauly. Shadow art. *ACM Transactions on Graphics*, 28(CONF):156–1, 2009.

[10] Adem Yasar Mulayim, Ulas Yilmaz, and Volkan Atalay. Silhouette-based 3-d model reconstruction from multiple images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 33(4):582–591, 2003.

[11] Andrew Nealen, Takeo Igarashi, Olga Sorkine, and Marc Alexa. Laplacian mesh optimization. In *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, pages 381–389, 2006.

[12] Thu Nguyen-Phuoc, Chuan Li, Stephen Balaban, and Yong-Liang Yang. Rendernet: A deep convolutional network for differentiable rendering from 3d shapes. *arXiv preprint arXiv:1806.06575*, 2018.

[13] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3504–3515, 2020.

[14] Luke Olsen, Faramarz F Samavati, Mario Costa Sousa, and Joaquim A Jorge. Sketch-based modeling: A survey. *Computers & Graphics*, 33(1):85–103, 2009.

[15] Despoina Paschalidou, Ali Osman Ulusoy, and Andreas Geiger. Superquadrics revisited: Learning 3d shape parsing beyond cuboids. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10344–10353, 2019.

[16] Fabio Pellacini, Parag Tole, and Donald P Greenberg. A user interface for interactive cinematic shadow design. *ACM Transactions on Graphics (TOG)*, 21(3):563–566, 2002.

[17] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020.

[18] Sudipta N Sinha and Marc Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 349–356. IEEE, 2005.

[19] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *arXiv preprint arXiv:1906.01618*, 2019.

[20] Shubham Tulsiani, Tinghui Zhou, Alexei A Efros, and Jitendra Malik. Multi-view supervision for single-view reconstruction via differentiable ray consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2626–2634, 2017.

[21] Leonard Wanger. The effect of shadow quality on the perception of spatial relationships in computer generated imagery. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 39–42, 1992.

[22] Leonard R Wanger, James A Ferwerda, Donald P Greenberg, et al. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications*, 12(3):44–58, 1992.