Humans as Light Bulbs: 3D Human Reconstruction from Thermal Reflection

Ruoshi Liu and Carl Vondrick
Columbia University
thermal.cs.columbia.edu

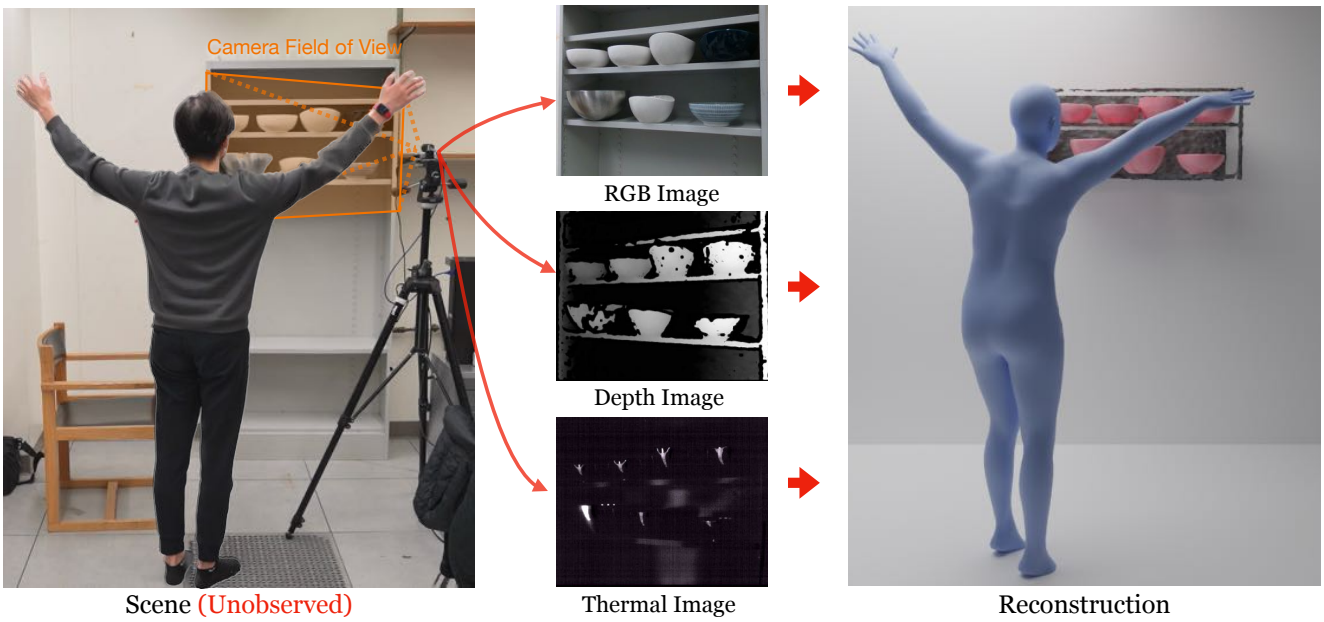


Figure 1. We introduce a method to reconstruct the 3D position and pose of a person from their thermal reflections in everyday objects with non-planar surfaces. Given an RGBD image and a thermal image in the middle of the figure, our method is able to recover the 3D mesh of the person (blue) as well as the objects (pink), even though they are not within the field of view of the camera system. Our system **never sees the scene on the left**, which is only shown for visualization purposes.

Abstract

The relatively hot temperature of the human body causes people to turn into long-wave infrared light sources. Since this emitted light has a larger wavelength than visible light, many surfaces in typical scenes act as infrared mirrors with strong specular reflections. We exploit the thermal reflections of a person onto objects in order to locate their position and reconstruct their pose, even if they are not visible to a normal camera. We propose an analysis-by-synthesis framework that jointly models the objects, people, and their thermal reflections, which combines generative models with differentiable rendering of reflections. Quantitative and qualitative experiments show our approach works in highly challenging cases, such as with curved mirrors or when the person is completely unseen by a normal camera.

1. Introduction

One of the major goals of the computer vision community is to locate people and reconstruct their poses in everyday environments. What makes thermal cameras particularly interesting for this task is the fact that humans are often the hottest objects in indoor environments, thus becoming infrared light sources. Humans have a relatively stable body temperature of 37 degrees Celcius, which according to the Stefan-Boltzmann law, turns people into a light source with constant brightness under long-wave infrared (LWIR). This makes LWIR images a robust source of signals of human activities under many different light and camera conditions.

Since infrared light on the LWIR spectrum has a wavelength that is much longer than visible light (8μm-14μm vs. 0.38μm-0.7μm), the objects in typical scenes look qualitatively very different from human vision. Many surfaces of

objects in our daily life – such as a ceramic bowl, a stainless steel fridge, or a polished wooden table top – have stronger specular reflections than in the visible light spectrum [7,58]. Figure 1 shows the reflection of a person with the surface of salad bowls, which is barely visible to the naked eye, if at all, but clearly salient in the LWIR spectrum.

In cluttered environments, a visible light camera may not always be able to capture the person, such as due to a limited field of view or occlusions. In such scenes, the ideal scene for locating and reconstructing a person would be an environment full of mirrors. This is what the world looks like under the LWIR spectrum. Infrared mirrors are abundant in the thermal modality, and reflections reveal significant non-line-of-sight information about the surrounding world.

In this paper, we introduce a method that uses the image of a thermal reflection in order to reconstruct the position and pose of a person in a scene. We develop an analysis-by-synthesis framework to model objects, people, and their thermal reflections in order to reconstruct people and objects. Our approach combines generative models with differentiable rendering to infer the possible 3D scenes that are compatible with the observations. Given a thermal image, our approach optimizes for the latent variables of generative models such that light emitting from the person will reflect off the object and arrive at the thermal camera plane.

Our approach works in highly challenging cases where the object acts as a curved mirror. Even when a person is completely unseen by a normal visible light camera, our approach is able to localize and reconstruct their 3D pose from just their thermal reflection. Traditionally, the increased specularity of surfaces has posed a challenge to thermography, making it extremely difficult to measure the surface temperature of a thermally specular surface, which brings out a line of active research aiming to remove the specular reflection for more accurate surface temperature measurement [4, 5, 40, 80]. We instead exploit these "difficulties" of LWIR to tackle the problem of 3D human reconstruction from a single view of thermal reflection image.

The primary contribution of the paper is a method to use the thermal reflection of the human body on everyday objects to infer their location in a scene and its 3D structure. The rest of the paper will analyze this approach in detail. Section 2 provides a brief overview of related work for 3D reconstruction and differentiable rendering. Section 3 formulates an integrated generative model of humans and objects in a scene, then discusses how to perform differentiable rendering of reflection, which we are able to invert to reconstruct the 3D scene. Section 4 analyzes the capabilities of this approach in the real world. We believe thermal cameras are powerful tools to study human activities in daily environments, extending computer vision systems' ability to function more robustly even under extreme light conditions.

2. Related Work

Differentiable Rendering. Differentiable rendering is a differentiable process of rendering 2D images given 3D scenes. The gradient obtained from the image space w.r.t. the scene parameters can be calculated and used to perform optimization. Recent advances in implicit 3D representations, especially Neural Radiance Field (NeRF) [2,3,52,54, 60, 69], have made impressive results on rendering photo-realistic images for the view-synthesis problems.

Another line of work focuses on differentiable rasterization [32, 42, 46, 47, 63, 73]. These works aim to replace the traditional rasterization process in computer graphics based on 2D projections of primitives such as polygons with z-buffering, with a differentiable rasterization process.

While differentiable, these methods are limited by the intrinsic difficulty of modeling single or multiple bounces of light in a scene, which can be modeled with physics-based differentiable ray tracing [25,30,41,57,73,81]. In our problem, because humans are light sources and we need to perform differentiable rendering of one-bounce reflection, we extended Soft Rasterizer [46].

Single-View 3D Reconstruction. From a practical point of view, obtaining 3D ground truth for supervision is often difficult and expensive [28]. In terms of the quantity of data available, the unlabeled 3D data is not comparable to the 2D data on the internet. This spurs a long-standing interest from the general computer vision community to pursue 3D reconstruction with as little information as a single-view [19, 31, 43, 45, 46, 75, 77].

In addition to general 3D object reconstruction, another line of research focus on the 3D reconstruction of human body from single-view images and videos [35, 36, 44, 53, 61, 64]. Representatively, SMPL-X [61] is an expressive whole-body model with details around hands and faces, represented as a triangle mesh with 10,475 vertices. In the same paper, SMPLify-X was proposed to estimate an SMPL-X model from just a single RGB image. This is done by first detecting human keypoints from the image with an off-the-shelf keypoint detector [6, 10, 10, 14, 17, 21, 38, 70]. Then the parameters of an SMPL-X model is optimized to fit the keypoints which serve as the observation of human in the 2D image.

3D Generative Model. Our system utilizes generative models for both objects and humans. For 3D objects, generative models are usually trained with synthetic datasets composed of CAD models [11]. Different generative architectures including VAE [9, 20, 20, 76], GAN [26, 62, 62, 74], normalizing flow [33, 34], and diffusion models [48] were proposed to generate objects in meshes, point clouds, or voxels. More recently, implicit 3D representation, or coordinate-based models, become a popular choice of modality to perform generative tasks [16, 23, 27, 51, 56, 59].

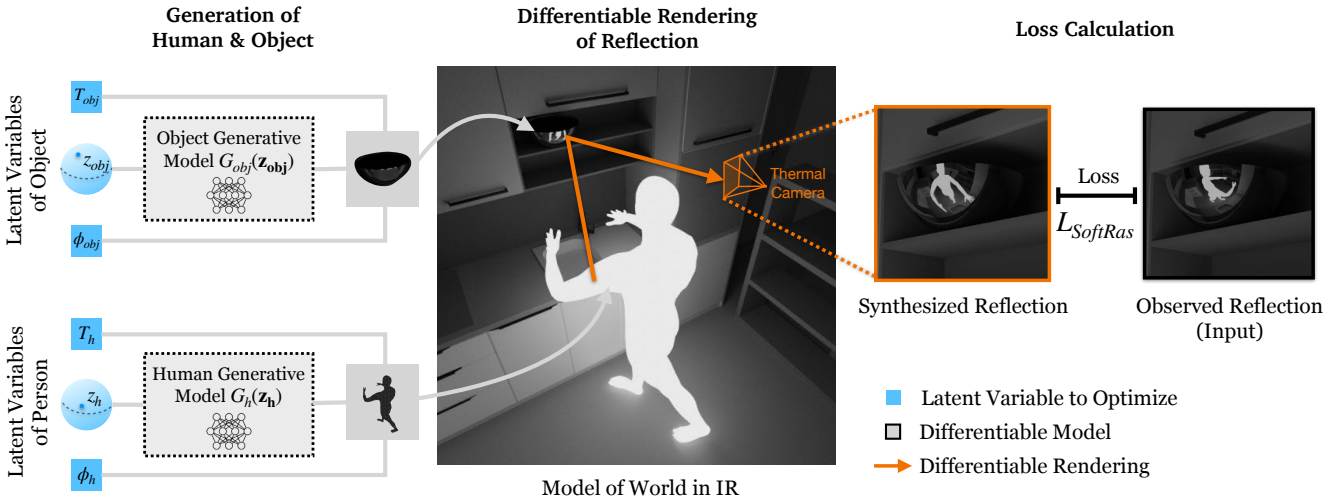


Figure 2. High-level overview of our analysis-by-synthesis framework. We sample random initializations from the latent space of pretrained generative models of humans and objects in 3D. Through a differentiable rendering process, we synthesize a reflection image of a human body on object surfaces. This synthesized reflection is compared with the observed reflection with an L_1 loss. Gradients are backpropagated through differentiable rendering and generative models to the latent variables.

For humans, [61, 71] proposed generative models for 3D humans, represented as SMPL-X models. In [61], a VAE is trained to generate human poses from 4 datasets including Human3.6M, LSP, CMU Panoptic, and PosePriors [1, 22, 55, 78]. The VAE samples a latent vector from a high-dimensional Gaussian distribution and generates a human pose vector. This pose vector is applied with a sparse linear regressor and a linear blend skinning function to generate a triangle mesh in a fully differentiable manner.

Thermal Computer Vision. Previous work has applied computer vision to thermal images for various problems [13, 15, 18, 24, 37, 65, 72]. ContactDB [8] used thermal imaging to obtain accurate human grasps of everyday objects for robotics applications. [49] studied the problem of thermal non-line-of-sight imaging. In comparison, this work focuses on the 3D reconstruction of people from their thermal reflections in non-planar objects. Other work pursues 3D reconstruction of objects from thermal images [12, 50, 66, 67]. To our knowledge, we are the first to perform 3D reconstruction of humans from their thermal reflection.

3. Methods

Our system takes an RGBD image and a thermal image of everyday objects with thermally reflective surfaces and performs a 2-stage optimization to estimate 3D objects and a human not in sight from both cameras' perspectives. In the first stage, a 6 DoF pose, scale, and a neural signed distance function [59] are jointly estimated for each object present in the scene. In the second stage, the location, orientation, and pose of the human are jointly estimated to reconstruct the observed thermal reflection.

Section 3.1 formulates the problem we aim to solve. Section 3.2 gives an overview of the approach. Section 3.3 describes the generative models we used in our approach in detail. Section 3.4 lays out a differentiable rendering algorithm of human thermal reflection. Section 3.5 formulates the optimization process and the objective functions.

3.1. Problem Formulation

We decompose a scene into 3 components: a human body, objects with specular surfaces in LWIR spectrum, and environmental heat sources. We first obtain a segmentation mask from each object in the scene from the RGBD image. To obtain the thermal reflection image, we perform ray tracing starting from the camera sensor to the light source – the human body, under Helmholtz reciprocity. Assuming a pinhole camera model, let $\mathbf{n}$ be the surface normal of the object at point $\mathbf{p}$, $\mathbf{r}$ be the vector from the camera sensor to $\mathbf{p}$ and $\mathbf{r}'$ the reflected ray vector. We model the intensity of each pixel $I_{\mathbf{x}}$ in the thermal camera as a binary value:

$$I_{\mathbf{x}} = \begin{cases} 1, & \mathbf{r}' \text{ intersects with } \mathcal{T}_{\phi,T}(M_h) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where M_h represents the human shape in the form of a triangle mesh, and $\mathcal{T}_{\phi,T}$ represents an SE(3) transformation matrix parameterized by rotation, translation, and scale. With background subtraction, the noise coming from environmental heat sources can be mitigated.

As described in figure 1, the calibrated thermal camera and RGBD camera with known intrinsic matrix and unknown extrinsic matrix capture an RGB image, a depth map, and a thermal image. Given these images as our observation containing N objects, we solve for the following 7

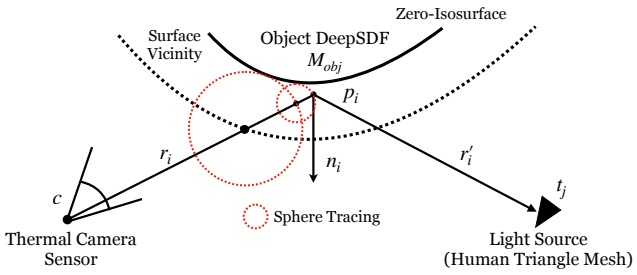


Figure 3. Differentiable Rendering of Reflection. Ray direction shown reverses the physical propagation direction of light by Helmholtz reciprocity.

Figure 4. 3D Object Reconstruction from RGBD

variables via optimization: locations $\{\mathbf{T}_{obj}\}_{i=0}^{N}$, rotations $\{\phi_{obj}\}_{i=0}^{N}$, scales $\{s_{obj}\}_{i=0}^{N}$, and the shape $\{M_{obj}\}_{i=0}^{N}$ of the objects, location $\mathbf{T}_h$, rotation ϕ_h, and the shape M_h of the human, all in camera's perspective.

3.2. Overview of Approach

The optimization problem we are solving is severely under-constrained, so we choose to leverage the priors provided by pretrained generative models. As described in figure 2, we first randomly sample the aforementioned 7 variables as initial input to the generative models to generate a 3D human and objects in the scene. Then we perform a differentiable rendering of human thermal reflection. For every ray from the camera sensor that intersects with an object, we can analytically calculate the reflected ray vector, given that the surface normals of the objects are defined by the output of the object generative model. With these reflected ray vectors, we can render a binary reflection image based on whether the reflected ray vectors intersect with humans, whose exact 3D shape and location are defined by the output of the human generative model. The optimization objective is to maximize the similarity between the rendered reflection image and the observed image captured by the thermal camera.

In order for such a pipeline to be differentiable, we need both the generative models of humans and objects, as well as the rendering algorithm, to be differentiable. In the following sections, we will describe how we achieve this.

3.3. Generative Models

Object: DeepSDF. We decided to use DeepSDF [59] as our generative models for objects. SDF, or signed distance function, is a function between a point in space and its orthogonal distance to the closest surface. In essence, DeepSDF is an SDF parameterized by a neural network G_{obj} whose input is a 3D coordinate $\mathbf{p}$ and output is a signed distance s. Following [59], we condition a DeepSDF model on a latent vector $\mathbf{z}_{obj}$ from a probabilistic latent space to make them generative model:

$$G_{obj}(\mathbf{p}, \mathbf{z}_{obj}) = s : \mathbf{p} \in \mathbb{R}^3, s \in \mathbb{R} \tag{2}$$

Human: SMPL-X. We adopted SMPL-X [61] as our generative models of 3D humans. Broadly, SMPL-X is composed of 2 components. The first is a variational autoencoder (VAE) that projects a latent vector $\mathbf{z}_h$ sampled from a probabilistic latent space with Gaussin prior to the human pose space, in the form of rotations of human body joints. The generated human body pose is then applied with a differentiable sparse linear regressor to generate vertices and triangle meshes representing the surface skins of a human body. Because both the VAE and the linear vertex regressor are differentiable, the location of each vertex is differentiable w.r.t. the latent vector $\mathbf{z}_h$.

3.4. Differentiable Rendering of Reflection

The information we have from the thermal image of objects is a reflected human silhouette. Soft rasterizer (SoftRas) [46] is a method of choice to perform differentiable rendering from 2D silhouette images. However, SoftRas is a differentiable rasterization algorithm, which does not directly apply to reflection, especially when the reflective surface is a curved surface defined by a DeepSDF. To overcome this limitation, we extended SoftRas to ray tracing under non-planar reflection off the zero-isosurface of a DeepSDF. This process is visualized in figure 3.

DeepSDF Depth Estimation. The complex geometry of an everyday object prevents us from projecting all triangles to the 2D image plane as in [46]. Thus, we need to march rays $\{\mathbf{r}_i\}$ from camera sensor $\mathbf{c}$, through the reflection point on the surface $\{\mathbf{p}_i\}$ with a surface normal $\{\mathbf{n}_i\}$, to the reflected rays $\{\mathbf{r}_i'\}$. To obtain the intersection point with the surface $\{\mathbf{p}_i\}$ given an SDF representation of an object, we need a differentiable method to extract the zero-isosurface and calculate the depth of the surface along the incoming ray $\mathbf{r}_i$. Previously, [79] proposed to perform surface projection by first grid-searching for a point close to the

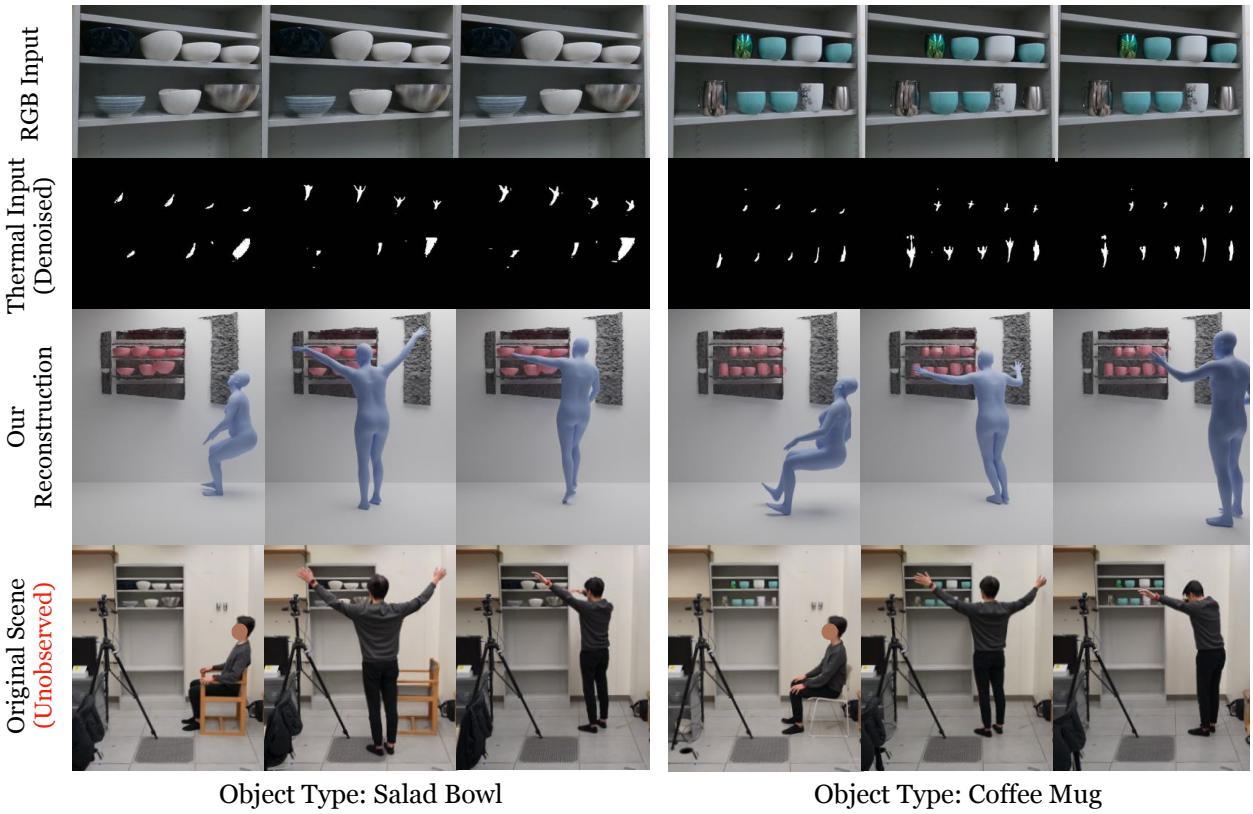


Figure 5. 3D Human Reconstruction (visualized from another camera view). From an RGBD image, we recover the 3D location $\mathbf{T}_{obj}$, pose ϕ_{obj}, and shape $\mathbf{z}_{obj}$ of each object. A marching cube visualization of the reconstructed 3D objects is shown in pink. With reconstructed objects, we recover 3D location $\mathbf{T}_h$, pose ϕ_h, and shape $\mathbf{z}_h$ of the human from a denoised thermal input showing reflections of the human on object surfaces, which we visualize in blue. We also include the original scene and our reconstruction from a calibrated third-camera view for comparison. This image is **not seen** by our system during reconstruction. The black mesh where the objects are located is the depth pointclouds captured by the RGBD camera.

zero-isosurface, then projecting along gradient direction $\frac{\partial G}{\partial \mathbf{p}}$ with the predicted distance. However, because the gradient direction is not in the same direction as the incoming ray, performing such an operation could yield a point far from the intersection point between the incoming ray and zero-isosurface, especially when the attack angle is small. To mitigate this error, we perform finite steps of sphere tracing along the ray to estimate the intersection point $\{\mathbf{p}_i\}$ as shown in figure 3.

DeepSDF Surface Normal. With the estimated intersection point $\{\mathbf{p}_i\}$ between $\{\mathbf{r}_i\}$ and the surface of the object, we calculate the surface normal of the object at $\{\mathbf{p}_i\}$:

$$\mathbf{n}_i = \frac{\partial G_{obj}(\mathbf{p}_i, \mathbf{z}_{obj})}{\partial \mathbf{p}_i} \tag{3}$$

We can then calculate the reflected ray vector as:

$$\mathbf{r}'_i = \mathbf{r}_i + 2 \cdot \mathbf{r}_i \cdot \frac{\mathbf{n}_i}{\|\mathbf{n}_i\|_2} \tag{4}$$

3D Ray-Triangle Distance. We can then calculate the pairwise distance matrix, denoted as $\mathcal{D}_{i,j}$ between each reflected ray $\mathbf{r}'_i$ and each triangle $t_j \in \{\mathbf{M}_h\}$, where $\mathbf{M}_h$ represents human body mesh. Each element in the distance matrix $d_{i,j} \in \mathcal{D}$ can be expressed as a differentiable function of vertices of t_j and the reflected ray vector $\mathbf{r}'_i$. We can also obtain a ray-triangle intersection matrix Λ with the same dimension as the distance matrix. Since the value of the ray-triangle intersection is binary, this calculation is not required to be differentiable.

Differentiable Ray Occupancy. Following SoftRas [46], we define the influence of each triangle t_j on each ray r'_i where the influence is expressed as a function of distance $d_{i,j}$:

$$d'_{i,j} = \text{sigmoid}\left(\lambda_{i,j}\frac{d^2_{i,j}}{\sigma}\right), \quad \lambda_{i,j} \in \Lambda, \quad d_{i,j} \in \mathcal{D} \tag{5}$$

where $\lambda_{i,j} = 1$ if reflected ray $\mathbf{r}'_i$ intersects with triangle $\mathbf{t}_j$, otherwise -1. $d_{i,j}$ denotes the distance between ray $\mathbf{r}'_i$ and triangle $\mathbf{t}_j$, σ is a hyperparameter that controls the

Figure 6. Real-world 3D human reconstruction from thermal reflections of cars. A diverse set of human poses can be reconstructed by using the surfaces of different types of cars as infrared mirrors. RGB input (1st row) and thermal input (2nd row) captured by a depth camera and a thermal camera are used as input to our method. Our reconstruction (3rd row) is compared with the original scene (4th row), both rendered/captured from another camera viewpoint.

"softness" of the influence. We then aggregate the influence of each triangle for a ray reflected $\mathbf{r}'_i$ to obtain the estimated binary occupancy of the ray by human body mesh $\mathbf{M}_h$:

$$\hat{I}_i = \mathcal{A}(\{\mathcal{D}\}_j) = 1 - \Pi_j(1 - d_{i,j}) \tag{6}$$

The estimated binary occupancy of ray $\hat{I}_i$ is a value between 0 and 1 and is compared with the ground truth binary thermal image defined in Eq. 1.

3.5. Optimization for Inference

3D Object Reconstruction. We first estimate the 6 DoF pose, scale, and shape of the objects present in the scene following a similar method as in [29]. We optimize the locations $\{\mathbf{T}_{obj}\}_{i=0}^{N}$, rotations $\{\phi_{obj}\}_{i=0}^{N}$, scale $\{\mathbf{s}_{obj}\}_{i=0}^{N}$, and the shape of the objects $\{\mathbf{z}_{obj}\}_{i=0}^{N}$, where $\{\mathbf{z}_{obj}\}_{i=0}^{N}$ are latent variables sampled from the probabilistic latent space of DeepSDF $\mathbf{G}_{obj}$ s.t. $\mathbf{M}_{obj} = G_{obj}(\mathbf{z}_{obj})$. For each object, we minimize the objective:

$$\mathcal{L}_{obj} = \mathcal{L}_{depth} + \mathcal{L}_{mask} + \mathcal{L}_{prior} \tag{7}$$

where $\mathcal{L}_{depth}$ is the L_1 loss between the estimated depth map and the measured depth map, $\mathcal{L}_{mask}$ denotes a pixel-wise L_2 loss between the estimated segmentation mask and the observed segmentation mask obtained from RGB observation, and $\mathcal{L}_{prior}$ is a shape prior regularization term.

3D Human Reconstruction Given the estimated translations, rotations, scales, and shape latent vectors from 3D object reconstruction, we optimize translation $\mathbf{T}_h$, rotation ϕ_h, and shape $\mathbf{z}_h$ of the human where $\mathbf{z}_h$ is the latent vector sampled from the pose VAE in SMPL-X s.t. $\mathbf{M}_h = G_h(\mathbf{z}_h)$. Upon obtaining the estimated reflection image $\hat{\mathbf{I}}$ and observed thermal silhouette image $\mathbf{I}$, we minimize the objective:

$$\mathcal{L}_{human} = \mathcal{L}_{silhouette} + \mathcal{L}_{prior} \tag{8}$$

where

$$\mathcal{L}_{silhouette} = 1 - \frac{\|\hat{\mathbf{I}} \otimes \mathbf{I}\|_1}{\|\hat{\mathbf{I}} \oplus \mathbf{I} - \hat{\mathbf{I}} \otimes \mathbf{I}\|_1} \tag{9}$$

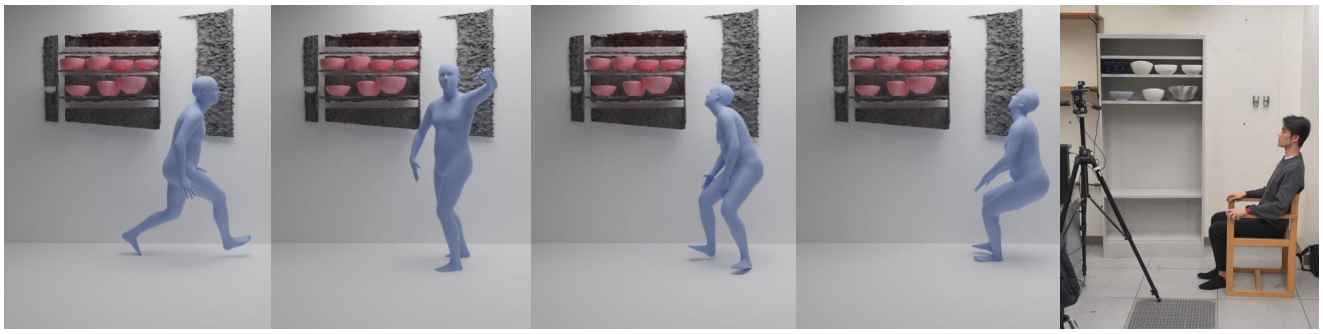
and $\mathcal{L}_{prior}$ is an L_2 regularization term on the human latent vector $\mathbf{z}_h$

4. Experiments

The goal of our experiments is to validate our hypothesis that LWIR thermal reflection on everyday objects provides sufficient information to perform accurate 3D human reconstruction in the real world. In section 4.1, we first demonstrate the accurate 3D reconstruction of objects from a single RGBD image, which serves as a foundation for 3D human reconstruction from reflection. We showcase our results on 3D human reconstruction with different poses and object types with everyday objects (section 4.2) and cars (section 4.3). Lastly, in section 4.4 we perform quantitative and qualitative ablation studies to evaluate the effectiveness of our technical approach.

4.1. 3D Object Reconstruction

Real-world depth sensors are subject to often significant measurement errors and are sensitive to lighting conditions (assuming an active stereo sensor). The surface depth estimated is often noisy, non-smooth, and full of "holes", as shown in figure 4. Performing differentiable rendering of reflection using the direct output of the depth sensor will



w/o Edge Sampling | w/o Sphere Tracing | w/o Surface Smoothing | Full Model | Original Scene (Unobserved)

Figure 7. Visualization of reconstruction obtained from ablated variations of our full model. While all variations can still find the 3D location of humans relatively accurately, the fine-grained details of human poses are significantly improved in our full model.

necessarily introduce an excessive amount of noise, given the reflected ray direction is calculated from the surface normal. Therefore, we opted to perform 3D object reconstruction from RGBD input first, then use the reconstructed surfaces for differentiable ray tracing.

In figure 4, we visualize the reconstructed objects from the RGBD input. Because our 3D representation of objects is an implicit function – DeepSDF, we perform marching cubes to extract the zero-isosurface of each object generated from the latent vector $\mathbf{z}_{obj}$. We then applied the SE(3) transformation matrix which is calculated from the estimated location T_{obj} and pose ϕ_{obj}.

As shown in figure 4, the location, pose, and shape of 3D objects can be faithfully reconstructed. Most importantly, we are able to obtain a high-fidelity, smooth, and accurate object surface without an explicit regularization on surface smoothness, which sets the foundation for the differentiable rendering of reflection. The successful reconstruction even when the depth input is noisy can be largely attributed to the object priors provided by searching in the latent space of a pretrained generative model. Generative priors such as a bowl is usually symmetric, the outside surface of a mug is often smooth, are enforced during the optimization process.

4.2. 3D Human Reconstruction

Given the reconstructed objects represented as individual DeepSDF models and their locations, we perform joint optimization of human location $\mathbf{T}_h$, orientation ϕ_h, and shape $\mathbf{z}_h$. The input to the differentiable rendering algorithm is a single binary thermal reflection image, representing a mask of human silhouette on each reflective object, as shown in figure 5. The binary mask of reflection is obtained from the thermal camera pointing towards the reflective objects, with simple denoising and thresholding. In addition to RGBD and thermal cameras, we put a third calibrated camera in the scene to capture the scene from another angle for evaluation and visualization. Note that any images from this camera are not used as input to our system.

We render the reconstruction from the third camera's perspective for comparison with the original scene at the exact time input data was captured. As shown in figure 5, the output of our method very accurately reconstructs the original scene. Note that the subject in the original scene is wearing normal clothing and the data is collected in a normal office environment without special lab environmental control. Besides, the objects used to reflect human thermal radiation are everyday objects with a variety of textures and materials that we purchased from supermarkets. This indicates the robustness of our system and its practical applicability to various settings.

4.3. Cars as Infrared Mirrors

Non-line-of-sight information of human activity plays a crucial role in the safe deployment of autonomous driving systems. Therefore, we showcase an experiment where we use cars as infrared mirrors to reconstruct the 3D location, orientation, and shape of a pedestrian that's not in the line-of-site of a camera system. In figure 6, we show the results in a similar fashion as figure 5. 3D reconstruction from thermal imaging could allow new opportunities for autonomous vehicles to sense and safely avoid occluded pedestrians.

4.4. Ablation Studies

To solve the extremely under-constrained and challenging problem, we made a lot of design decisions that turned out to be crucial to the quality of reconstruction. To evaluate the effectiveness of our technical approach, we perform ablation studies and compare our reconstruction with a baseline. We have included both quantitative evaluations as well as qualitative visualizations. Here we described some representative design decisions in detail.

Edge Sampling. As pointed out by [41], edge sampling plays an important role in differentiable ray tracing. This is even more significant for human reflection silhouettes. In addition, unless a person is standing right in front of the reflector, the reflection silhouette usually occupies a small

Evaluation Method	Object Type	w/o Edge Sampling	w/o Sphere Tracing	w/o Surface Smoothing	Full Model	Random
2D Keypoints [17]	Bowl	0.231	0.224	0.145	**0.116**	0.346
2D Keypoints [17]	Mug	0.101	0.209	0.109	**0.094**	0.371
3D Skeleton [39]	Bowl	0.309	0.272	0.212	**0.152**	0.322
3D Skeleton [39]	Mug	0.223	0.215	0.202	**0.126**	0.317

Table 1. Quantitative evaluation of our reconstructed 3D human. We used two evaluation methods by comparing the extracted 2D keypoints and 3D skeleton with a calibrated 3rd camera view. Object type indicates the type of objects serving as reflectors. Columns 3-5 are three variations of our full model with some parts ablated. Random shows the corresponding metric if a random sample were to be drawn from the HumanEva [68] dataset, which includes diverse poses in daily human activities. Numbers show the average normalized Euclidean distance between reconstruction and ground truth.

region of the thermal image. We therefore perform edge detection on the reflection image to extract edges of human silhouette and sampling ray with a probability distribution concentrated at the vicinity of these edges and increasing the concentration as training progresses as a type of curriculum training.

Sphere Tracing. As we've described in 3.4, direct surface projection from the vicinity of an SDF will yield a point far from the real intersection between the incoming ray and the zero-isosurface of the SDF. Therefore, we perform 3 steps of sphere tracing to estimate the intersection point on the object.

Surface Smoothing. From experiments, we discovered that even after we perform sphere tracing, the reflection surface normals are still noisy, causing the differentiable rendering algorithm to produce a noisy reflection. This effectively injects noise into the gradients, making the optimization more challenging. We discovered that this is caused by the reconstructed DeepSDF having a locally non-smooth zero iso-surface. In figure 8, we visualize the surface normals calculated from a small region of zero-isosurface which shows the non-smoothness. To mitigate this error, we perform surface smoothing during differentiable rendering by sampling 8 neighboring rays surrounding the main ray and averaging all estimated surface normals for reflection calculation.

Evaluation. We evaluate our reconstruction as well as the 3 aforementioned ablated methods by comparing the 2D keypoints and 3D skeleton estimated from synchronized images captured by a calibrated third camera. We used [17] for 2D keypoints detection and [39] for 3D skeleton estimation. For comparison, we compared the reconstruction to 200 randomly sampled 2D human keypoints and 3D skeletons from the HumanEva dataset [68].

Both the quantitative experiments and qualitative visualizations have shown the effectiveness of our technical approach as well as the design decisions. Particularly, we believe our findings regarding differentiable rendering of reflections on implicit surfaces will provide insights to other computer vision researchers working with reflections.

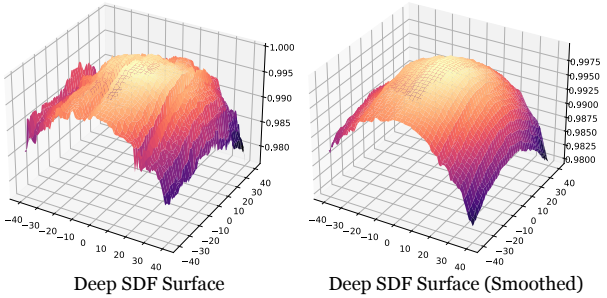


Figure 8. Visualization of DeepSDF Surface Normals within a 1.6 cm$\times$1.6 cm area. From the visualization, we can clearly see an improvement in surface smoothness at a small scale, which is beneficial to the differentiable rendering process. The X-Y plane (horizontal) indicates the location on a surface with a step size of 0.2mm. Given the unit surface normal vector at a point on the grid (x, y), we compute its dot product with the unit surface normal vector at $(0, 0)$, and plot this value on the Z-axis. This shows the curvature of the surface as well as its level of smoothness.

5. Conclusion

This paper shows that 3D position and pose of a human can be reconstructed from a single thermal image of everyday objects reflecting human thermal radiations. We approach this problem by combining the priors learned by pretrained 3D generative models and differentiable rendering of reflections. By formulating the problem as an optimization problem, we perform analysis by synthesis to explain the observations. We believe thermal cameras are powerful tools to study human activities in daily environments and integrating them with modern computer vision models will bring out many downstream applications in robotics, graphics, and 3D perception.

Acknowledgements: This research is based on work partially supported by the Toyota Research Institute, the NSF NRI Award #1925157, and the NSF CAREER Award #2046910. We acknowledge Shree Nayar, Shuran Song, Runlin Xu, James Tompkin, Mark Sheinin, Mia Chiquier, Jeremy Klotz for helpful feedback, and Su Li, Dylan Chen, Sophia Su for helping with data collection.

References

[1] Ijaz Akhter and Michael J Black. Pose-conditioned joint angle limits for 3d human pose reconstruction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1446–1455, 2015. 3

[2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. 2

[3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 2

[4] Ganbayar Batchuluun, Na Rae Baek, Dat Tien Nguyen, Tuyen Danh Pham, and Kang Ryoung Park. Region-based removal of thermal reflection using pruned fully convolutional network. *IEEE Access*, 8:75741–75760, 2020. 2

[5] Ganbayar Batchuluun, Hyo Sik Yoon, Dat Tien Nguyen, Tuyen Danh Pham, and Kang Ryoung Park. A study on the elimination of thermal reflections. *IEEE Access*, 7:174597–174611, 2019. 2

[6] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. Blazepose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204*, 2020. 2

[7] HE Bennett and JO119764 Porteus. Relation between surface roughness and specular reflectance at normal incidence. *JOSA*, 51(2):123–129, 1961. 2

[8] Samarth Brahmbhatt, Cusuh Ham, Charles C Kemp, and James Hays. Contactdb: Analyzing and predicting grasp contact via thermal imaging. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8709–8719, 2019. 3

[9] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. *arXiv preprint arXiv:1608.04236*, 2016. 2

[10] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017. 2

[11] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 2

[12] Chia-Yen Chen, Chia-Hung Yeh, Bao Rong Chang, and Jun-Ming Pan. 3d reconstruction from ir thermal images and reprojective evaluations. *Mathematical Problems in Engineering*, 2015, 2015. 3

[13] I-Chien Chen, Chang-Jen Wang, Chao-Kai Wen, and Shiow-Jyu Tzou. Multi-person pose estimation using thermal images. *IEEE Access*, 8:174964–174971, 2020. 3

[14] Bowen Cheng, Bin Xiao, Jingdong Wang, Honghui Shi, Thomas S Huang, and Lei Zhang. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5386–5395, 2020. 2

[15] James W Davis and Mark A Keck. A two-stage template approach to person detection in thermal imagery. In *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1*, volume 1, pages 364–369. IEEE, 2005. 3

[16] Yu Deng, Jiaolong Yang, and Xin Tong. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10286–10296, 2021. 2

[17] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE international conference on computer vision*, pages 2334–2343, 2017. 2, 8

[18] Lu Gan, Connor Lee, and Soon-Jo Chung. Unsupervised rgb-to-thermal domain adaptation via multi-domain attention network. *arXiv preprint arXiv:2210.04367*, 2022. 3

[19] Shubham Goel, Angjoo Kanazawa, , and Jitendra Malik. Shape and viewpoints without keypoints. In *ECCV*, 2020. 2

[20] Yanran Guan, Tansin Jahan, and Oliver van Kaick. Generalized autoencoder for volumetric shape generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 268–269, 2020. 2

[21] Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos. Densepose: Dense human pose estimation in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7297–7306, 2018. 2

[22] Fei Han, Brian Reily, William Hoff, and Hao Zhang. Space-time representation of people based on 3d skeletal data: A review. *Computer Vision and Image Understanding*, 158:85–105, 2017. 3

[23] Zekun Hao, Hadar Averbuch-Elor, Noah Snavely, and Serge Belongie. Dualsdf: Semantic shape manipulation using a two-level representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7631–7641, 2020. 2

[24] Jingu Heo, Seong G Kong, Besma R Abidi, and Mongi A Abidi. Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 122–122. IEEE, 2004. 3

[25] Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. Difftaichi: Differentiable programming for physical simulation. *arXiv preprint arXiv:1910.00935*, 2019. 2

[26] Le Hui, Rui Xu, Jin Xie, Jianjun Qian, and Jian Yang. Progressive point cloud deconvolution generation network. In *European Conference on Computer Vision*, pages 397–413. Springer, 2020. 2

[27] Moritz Ibing, Isaak Lim, and Leif Kobbelt. 3d shape generation with grid-based implicit functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13559–13568, 2021. 2

[28] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, jul 2014. 2

[29] Muhammad Zubair Irshad, Sergey Zakharov, Rares Ambrus, Thomas Kollar, Zsolt Kira, and Adrien Gaidon. Shapo: Implicit representations for multi-object shape, appearance, and pose optimization. *arXiv preprint arXiv:2207.13691*, 2022. 6

[30] Yue Jiang, Dantong Ji, Zhizhong Han, and Matthias Zwicker. Sdfdiff: Differentiable rendering of signed distance fields for 3d shape optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1251–1261, 2020. 2

[31] Angjoo Kanazawa, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In *ECCV*, 2018. 2

[32] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3d mesh renderer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3907–3916, 2018. 2

[33] Hyeongju Kim, Hyeonseung Lee, Woo Hyun Kang, Joun Yeop Lee, and Nam Soo Kim. Softflow: Probabilistic framework for normalizing flow on manifolds. *Advances in Neural Information Processing Systems*, 33:16388–16397, 2020. 2

[34] Roman Klokov, Edmond Boyer, and Jakob Verbeek. Discrete point flow networks for efficient point cloud generation. In *European Conference on Computer Vision*, pages 694–710. Springer, 2020. 2

[35] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5253–5263, 2020. 2

[36] Nikos Kolotouros, Georgios Pavlakos, Michael J Black, and Kostas Daniilidis. Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2252–2261, 2019. 2

[37] Zülfiye Kütük and Görkem Algan. Semantic segmentation for thermal images: A comparative survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 286–295, 2022. 3

[38] Jia Li, Wen Su, and Zengfu Wang. Simple pose: Rethinking and improving a bottom-up approach for multi-person pose estimation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11354–11361, 2020. 2

[39] Jiefeng Li, Chao Xu, Zhicun Chen, Siyuan Bian, Lixin Yang, and Cewu Lu. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3383–3393, 2021. 8

[40] Ning Li, Yongqiang Zhao, Quan Pan, and Seong G Kong. Removal of reflections in lwir image with polarization characteristics. *Optics express*, 26(13):16488–16504, 2018. 2

[41] Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. Differentiable monte carlo ray tracing through edge sampling. *ACM Transactions on Graphics (TOG)*, 37(6):1–11, 2018. 2, 7

[42] Tzu-Mao Li, Michal Lukáვc, Michaël Gharbi, and Jonathan Ragan-Kelley. Differentiable vector graphics rasterization for editing and learning. *ACM Transactions on Graphics (TOG)*, 39(6):1–15, 2020. 2

[43] Xueting Li, Sifei Liu, Kihwan Kim, Shalini De Mello, Varun Jampani, Ming-Hsuan Yang, and Jan Kautz. Self-supervised single-view 3d reconstruction via semantic consistency. In *European Conference on Computer Vision*, pages 677–693. Springer, 2020. 2

[44] Kevin Lin, Lijuan Wang, and Zicheng Liu. End-to-end human pose and mesh reconstruction with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1954–1963, 2021. 2

[45] Ruoshi Liu, Sachit Menon, Chengzhi Mao, Dennis Park, Simon Stent, and Carl Vondrick. Shadows shed light on 3d objects. *arXiv preprint arXiv:2206.08990*, 2022. 2

[46] 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. 2, 4, 5

[47] Matthew M Loper and Michael J Black. Opendr: An approximate differentiable renderer. In *European Conference on Computer Vision*, pages 154–169. Springer, 2014. 2

[48] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021. 2

[49] Tomohiro Maeda, Yiqin Wang, Ramesh Raskar, and Achuta Kadambi. Thermal non-line-of-sight imaging. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2019. 3

[50] E Maset, A Fusiello, F Crosilla, R Toldo, and D Zorzetto. Photogrammetric 3d building reconstruction from thermal images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:25, 2017. 3

[51] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4460–4470, 2019. 2

[52] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2

[53] Gyeongsik Moon and Kyoung Mu Lee. I2l-meshnet: Image-to-lixel prediction network for accurate 3d human pose and

mesh estimation from a single rgb image. In *European Conference on Computer Vision*, pages 752–768. Springer, 2020. 2

[54] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv preprint arXiv:2201.05989*, 2022. 2

[55] Xuecheng Nie, Jiashi Feng, Jianfeng Zhang, and Shuicheng Yan. Single-stage multi-person pose machines. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6951–6960, 2019. 3

[56] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11453–11464, 2021. 2

[57] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: A retargetable forward and inverse renderer. *ACM Transactions on Graphics (TOG)*, 38(6):1–17, 2019. 2

[58] Michael Oren and Shree K Nayar. Generalization of lambert's reflectance model. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 239–246, 1994. 2

[59] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 2, 3, 4

[60] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. 2

[61] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10975–10985, 2019. 2, 3, 4

[62] Sameera Ramasinghe, Salman Khan, Nick Barnes, and Stephen Gould. Spectral-gans for high-resolution 3d point-cloud generation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8169–8176. IEEE, 2020. 2

[63] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv preprint arXiv:2007.08501*, 2020. 2

[64] Davis Rempe, Tolga Birdal, Aaron Hertzmann, Jimei Yang, Srinath Sridhar, and Leonidas J Guibas. Humor: 3d human motion model for robust pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11488–11499, 2021. 2

[65] Rafael E Rivadeneira, Angel D Sappa, Boris X Vintimilla, Jin Kim, Dogun Kim, Zhihao Li, Yingchun Jian, Bo Yan, Leilei Cao, Fengliang Qi, et al. Thermal image super-resolution challenge results-pbvs 2022. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 418–426, 2022. 3

[66] Agata Sage, Daniel Ledwoń, Jan Juszczyk, and Paweł Badura. 3d thermal volume reconstruction from 2d infrared images—a preliminary study. In *International Scientific Conference Advances in Applied Biomechanics*, pages 371–379. Springer, 2020. 3

[67] Sebastian Schramm, Phil Osterhold, Robert Schmoll, and Andreas Kroll. Combining modern 3d reconstruction and thermal imaging: Generation of large-scale 3d thermograms in real-time. *Quantitative InfraRed Thermography Journal*, 19(5):295–311, 2022. 3

[68] Leonid Sigal, Alexandru O Balan, and Michael J Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision*, 87(1):4–27, 2010. 8

[69] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben @articleli2018differentiable, title=Differentiable monte carlo ray tracing through edge sampling, author=Li, Tzu-Mao and Aittala, Miika and Durand, Frédo and Lehtinen, Jaakko, journal=ACM Transactions on Graphics (TOG), volume=37, number=6, pages=1–11, year=2018, publisher=ACM New York, NY, USA Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7495–7504, 2021. 2

[70] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5693–5703, 2019. 2

[71] Garvita Tiwari, Dimitrije Antić, Jan Eric Lenssen, Nikolaos Sarafianos, Tony Tung, and Gerard Pons-Moll. Pose-ndf: Modeling human pose manifolds with neural distance fields. In *European Conference on Computer Vision*, pages 572–589. Springer, 2022. 3

[72] Andre Treptow, Grzegorz Cielniak, and Tom Duckett. Real-time people tracking for mobile robots using thermal vision. *Robotics and Autonomous Systems*, 54(9):729–739, 2006. 3

[73] Delio Vicini, Sébastien Speierer, and Wenzel Jakob. Differentiable signed distance function rendering. *ACM Transactions on Graphics (TOG)*, 41(4):1–18, 2022. 2

[74] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. *Advances in neural information processing systems*, 29, 2016. 2

[75] Shangzhe Wu, Christian Rupprecht, and Andrea Vedaldi. Unsupervised learning of probably symmetric deformable 3d objects from images in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1–10, 2020. 2

[76] Zhijie Wu, Xiang Wang, Di Lin, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Sagnet: Structure-aware gen-

erative network for 3d-shape modeling. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019. 2

[77] Yufei Ye, Shubham Tulsiani, and Abhinav Gupta. Shelf-supervised mesh prediction in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8843–8852, 2021. 2

[78] Xiang Yu, Feng Zhou, and Manmohan Chandraker. Deep deformation network for object landmark localization. In *European conference on computer vision*, pages 52–70. Springer, 2016. 3

[79] Sergey Zakharov, Wadim Kehl, Arjun Bhargava, and Adrien Gaidon. Autolabeling 3d objects with differentiable rendering of sdf shape priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12224–12233, 2020. 4

[80] Björn Zeise and Bernardo Wagner. Temperature correction and reflection removal in thermal images using 3d temperature mapping. In *ICINCO (2)*, pages 158–165, 2016. 2

[81] Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. A differential theory of radiative transfer. *ACM Transactions on Graphics (TOG)*, 38(6):1–16, 2019. 2