Seeing the World through Your Eyes
Supplementary Materials

A. Details on Real Experiments

A.1. Sample input images for the method

The input to our method is a sequence of images captured by a *static* camera of a moving person that moves within the camera field of view. The views of the moving head essentially provide us with multi-view observations of what the person is looking at that we can leverage to reconstruct the observed scene. In Figures 3 and 4, we show the entire sequence of images used to reconstruct the object of interest and the recovered texture and reconstructed view. In Figure 6, we also visualize how the estimated eye poses change after the eye-pose optimization process. Note that while we also optimize the rotation/orientation of each eye, we only visualize the optimized translations.

A.2. Qualitative analysis of radial loss

Separating the iris texture and the reflected scene by jointly optimizing a 2D texture field and 3D radiance field of the scene can be ill-posed for textureless regions that have no motion across all captured images. As a result, some regions of the scene can get *absorbed* into the 2D texture field. However, one observation to make is that the eye colors are rotationally consistent, as shown in Figure 1b, by visualizing the extracted texture in polar coordinates. To alleviate the issue, we use this observation to penalize regions that violate the rotational consistency (as stated in Eq. 5 of the main paper). In Figure 1c, we show what regions in the extracted texture are penalized and show that it consistently prevents the scene regions from being absorbed into the texture.

A.3. Capture setup of controlled results

We tested our method with different illuminations and capture setups. For the structured captures, we use two external area lights facing the objects of interest to enhance the reflection off the user's eyes, as shown in Figure 2. We used this setup to validate and test our method. We also show results on unstructured captures of indoor and outdoor environments in Figure 9 of the main paper.

A.4. Effects of eye disease on reconstruction

Our method relies on healthy humans having the same cornea shape. However, eye diseases can affect the cornea shape. For example, Keratoconus is a condition in which the cornea bulges outward into a cone shape. To analyze the effects of such eye disease, we captured a sequence of images of a person with Keratoconus, as well as control capture where the user is wearing corrective contacts that mimic the shape of a healthy cornea. In Figure 7 we show that Keratoconus affects the shape of the reflected object, as we can see that the reflected objects are vertically compressed. Note that despite the cornea shape not matching our model of the cornea accurately, we can still recover the reflected object. We hypothesize that this is due to the eye-pose optimizer, as it has degrees of freedom to ensure the consistency of the reflected views and the 3D reconstruction despite the error in the reflected ray directions and origins, which shows that our method can be robust to minor errors in the eye geometry.

A.5. Depth results

Even though we only have a monocular camera in our capture setup, the views we get from the eye reflections make it effectively multi-camera setup. As a result, we can also get an estimated depth map from the radiance field reconstruction. In Figure 5, we show the disparity we get from an unstructured capture of office space as well as the disparity of an object placed in front of the user. Note that since we have an estimate of the eye location in 3D space, this could be a promising approach to obtaining metric depth with a monocular view by observing moving heads.

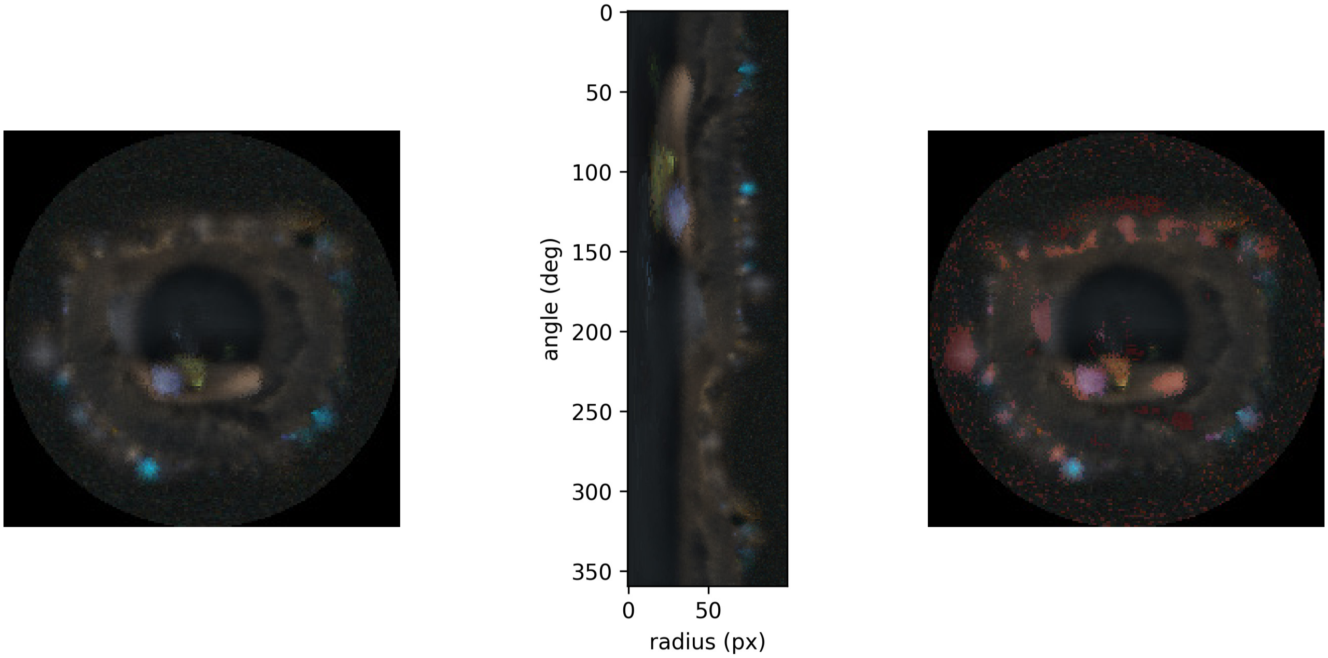
B. Additional Synthetic Results

B.1. Additional qualitative results

In Figure 8, we show additional results on novel view synthesis in simulated environments as well as sample corneas used in the input.

B.2. Pose optimization ablation

In Figure 9, we show how different noise levels affect the reconstruction quality. Note that while noisy poses can signif-



(a) Extracted texture (b) Polar coords. (c) Penalized regions

Figure 1. **Texture analysis.** When attempting to sepearte the iris texture from the scene by simply jointly training a radiance field and a texture field, regions of the scene can be absorbed into the iris texture (as shown in (a)). When rewarping the extracted texture into polar coordinates (b), we observe that the iris regions are generally consistent in color for the different radii, unlike the scene regions absorved into the eye. With the radial regularization loss (see Eq. 5 in the main paper), we penalize regions that deviate by a standard deviation in color from the 10th and 90th percentile colors in the radius they are located in. In (c) we show that the penalized regions indeed penalize regions that are absorbed from the scene into the iris texture.

Figure 2. **Capture setup.** We illuminate the objects of interest with two area lights to ensure that sufficient amount of light is reflected off the eyes.

icantly affect the reconstruction when not using our cornea pose optimizer, the reconstruction result is not visibly affected when using the pose optimization. Additionally, we directly visualize the effect of our pose optimizer on depth errors in the poses in 3D in Figure 10. We do not visualize X/Y axis or rotation errors in the same way because a similar percentage of eye center estimation error (which leads to XY error) as iris radius estimation error (which leads to depth error) results in less change in X/Y value than in Z value and small rotations are difficult to see in this kind of 3D visualization.

B.3. Texture decomposition ablation

Because of the iris texture in the eye observation, without an explicit texture decomposition, the radiance field must compensate by explicitly modeling the iris in the scene. As a result, the novel views rendered are a mix of the scene and the iris texture, as shown in Figure 11. By jointly training the radiance field and texture field, we can recover clean views of the scene.

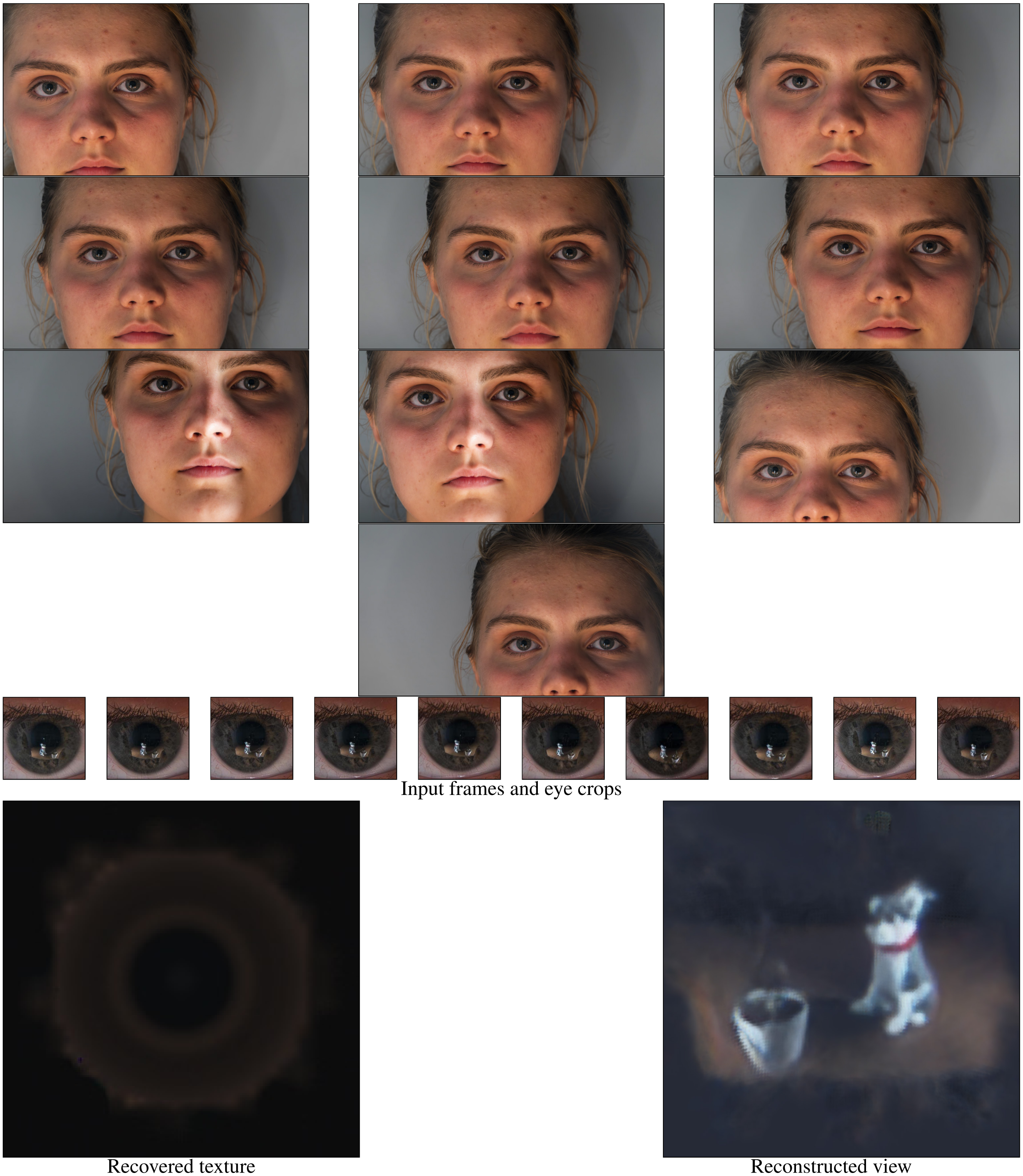


Figure 3. **Full input.** Example of a full input for our method, and the reconstructed view rendered from the radiance field alongside the recovered texture.

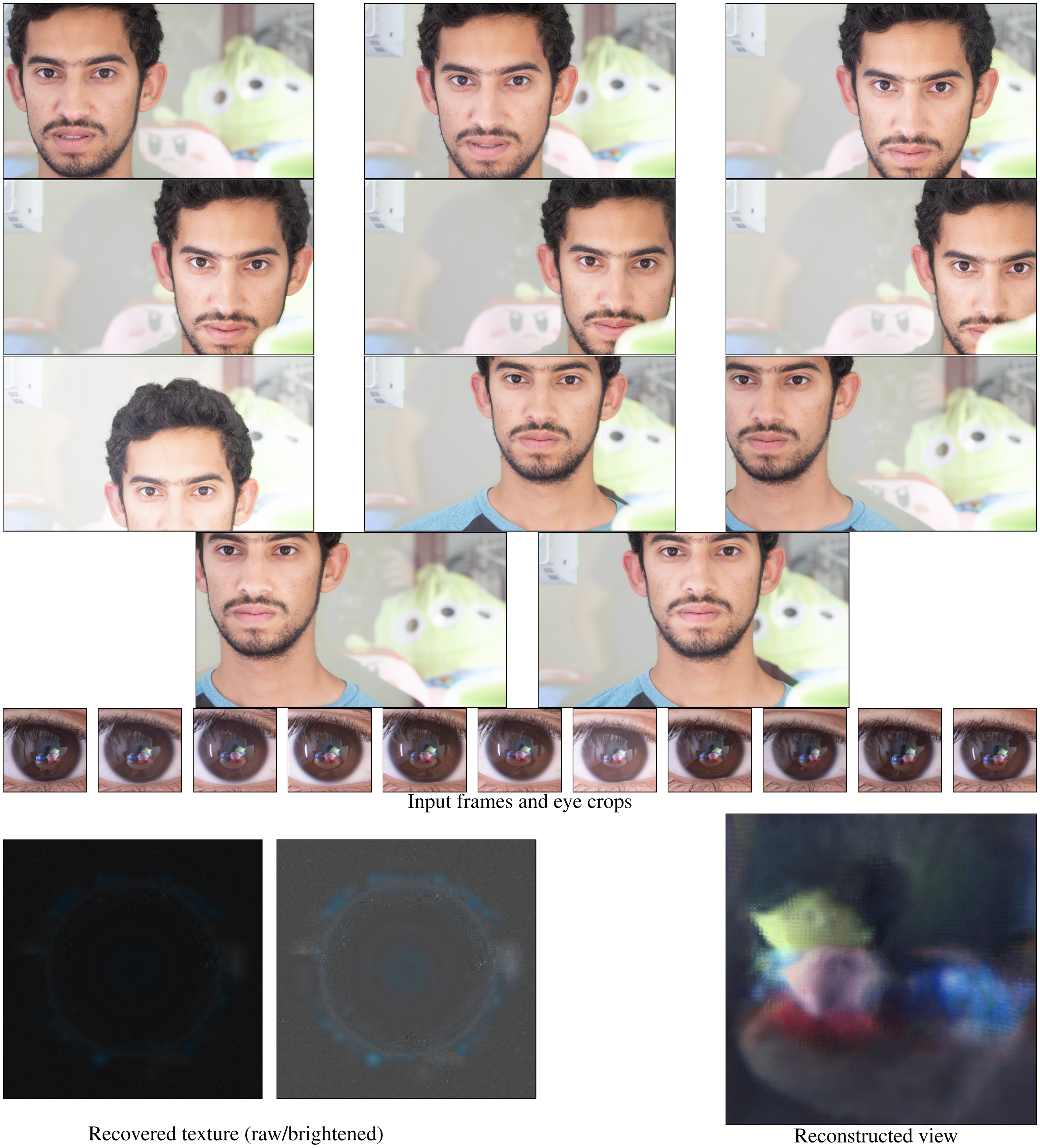
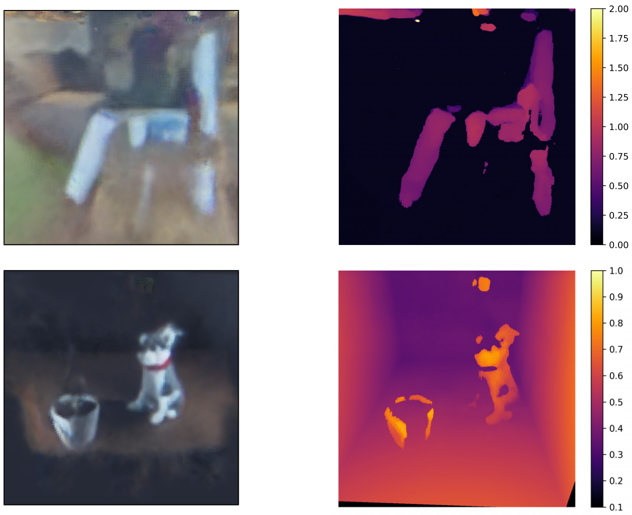
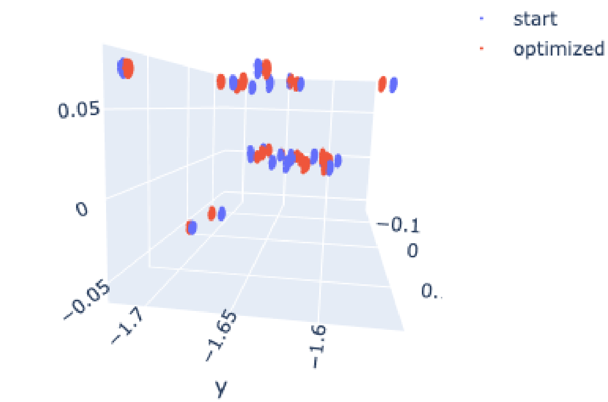


Figure 4. **Full input.** Example of a full input for our method, and the reconstructed view rendered from the radiance field alongside the recovered texture. We brighten the iris texture on the right to assist with visualizing the recovered content only.

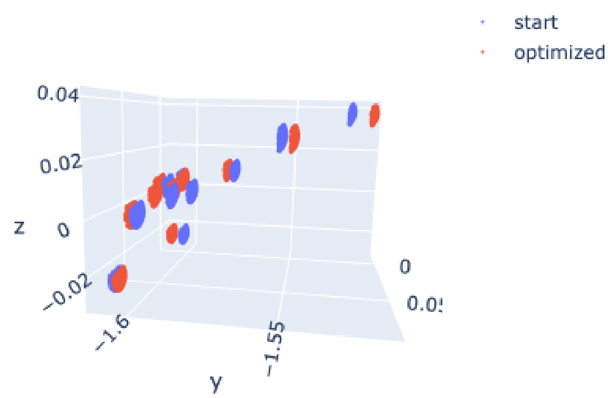


RGB Disparity

Figure 5. **Depth recovery.** Since we have multiple views of the scene through the person's moving eyes, we can get the depth of the scene using a static monocular camera.



Poses of Fig. 4 data



Poses of Fig. 3 data

Figure 6. **Poses visualization.** we visualize the initial estimated location of the user's eyes against the optimized poses.

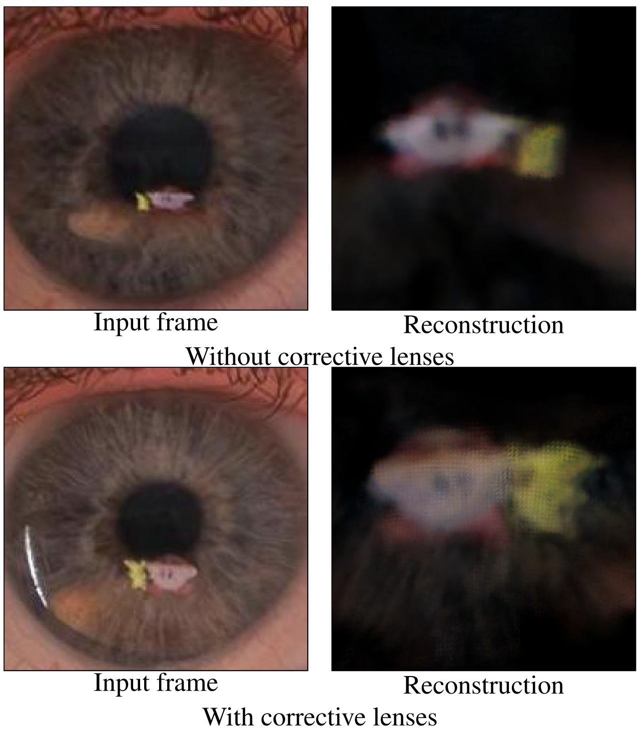


Figure 7. **Effects of eye disease.** We ran our method with a user who suffers from Keratoconus, an eye disease that distorts the shape of the cornea. Notice that the reflection in the input frame is visibly distorted—vertically compressed by a severe astigmatism. When using our method, we recover a distorted version of the reflected object, indicating that the pose optimization compensates for the inconsistency in eye topology across users.

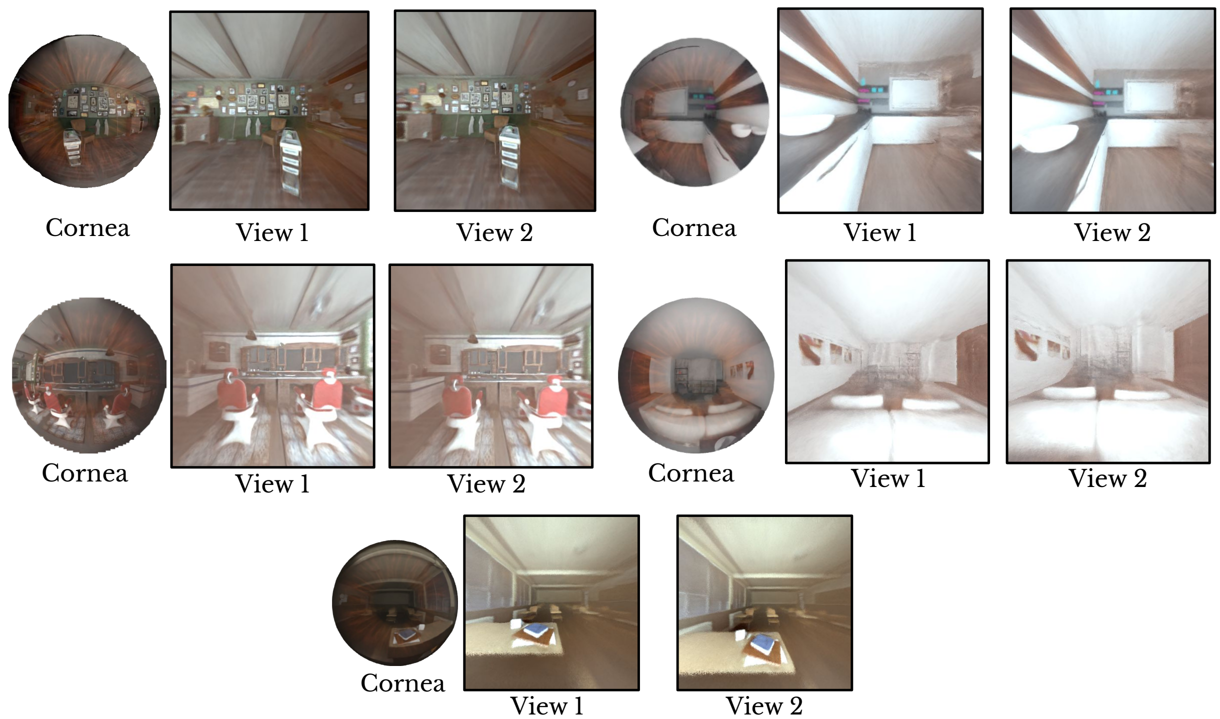


Figure 8. **Qualitative synthetic results.** We show that our method can achieve reasonable radiance field reconstructions suitable for novel view synthesis from challenging measurements in simulation by visualizing the rendered views at two different eye locations facing the scene.

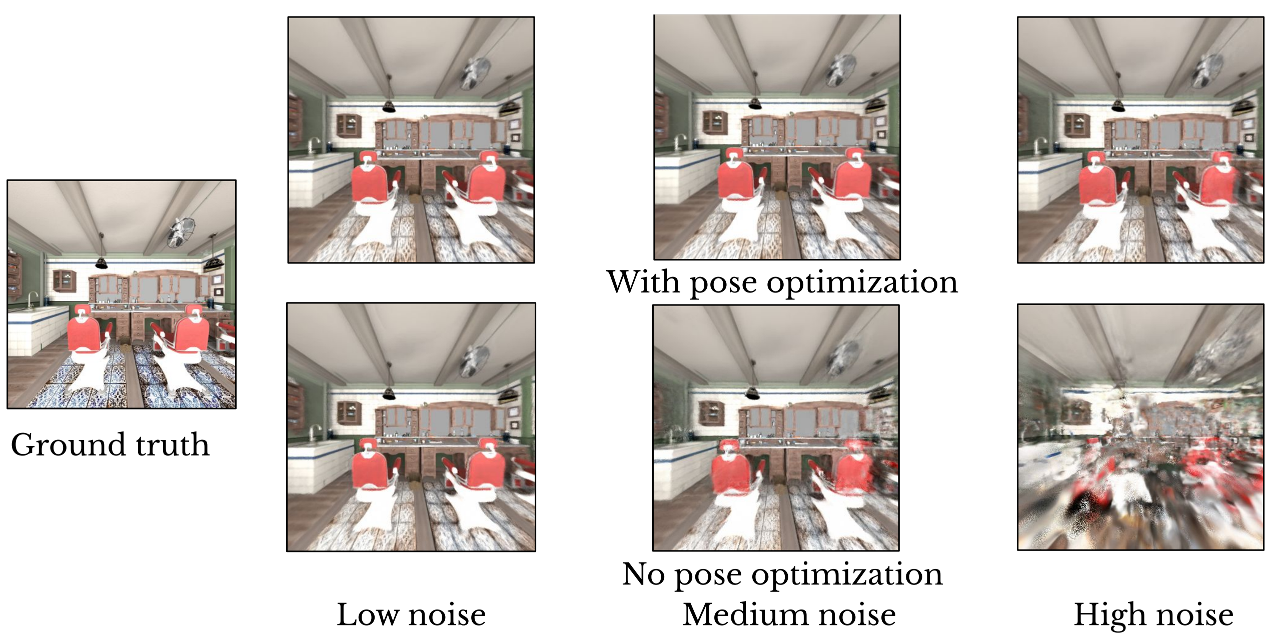


Figure 9. **Synthetic pose optimization ablation.** In simulation, the cornea pose optimization refines the noisy initial poses and results in clearer reconstruction.

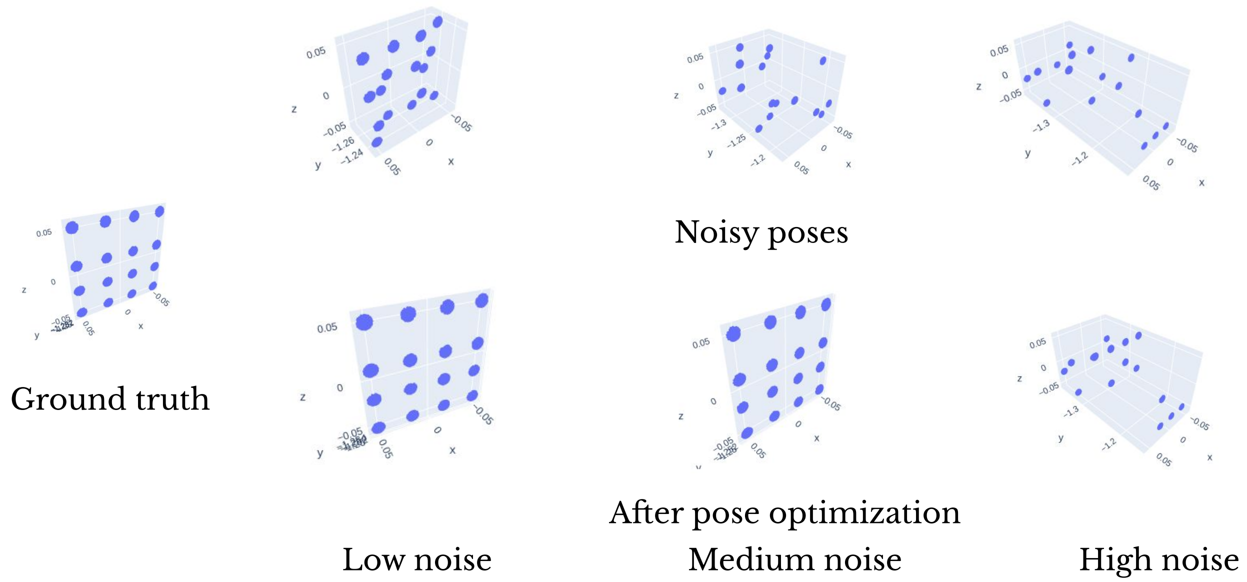


Figure 10. **Synthetic pose optimization visualization.** We directly visualize the effect that the pose optimization has on errors on in the z (depth) direction. Our implementation of pose optimization succeeds in fixing all errors up to moderate pose noise

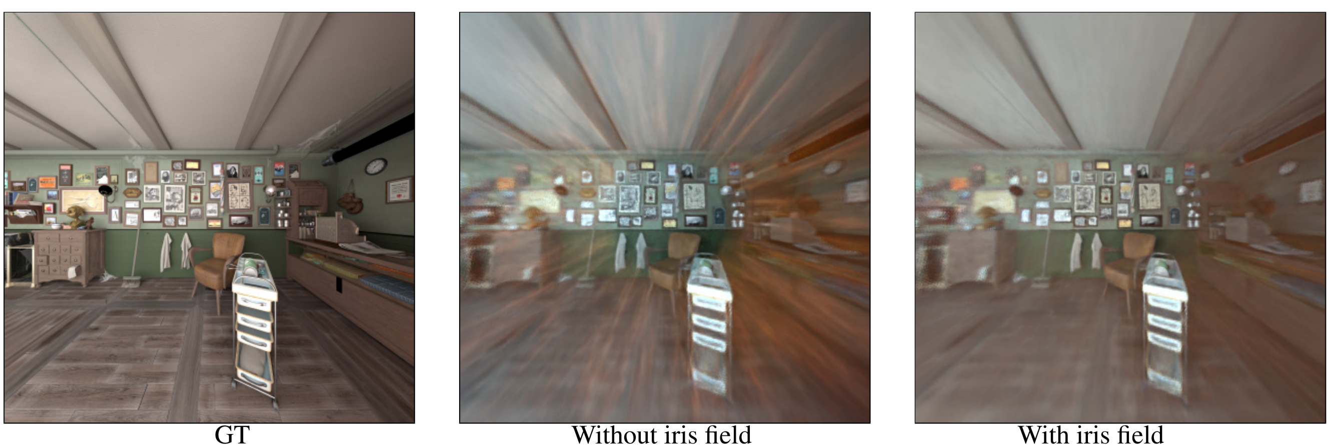


Figure 11. **Texture decomposition and novel view synthesis.** Without the texture decomposition, the novel view rendering is an overlay of the iris texture and the scene, but the texture decomposition helps us avoid the iris floaters. We rescale the output color channels for clarity.