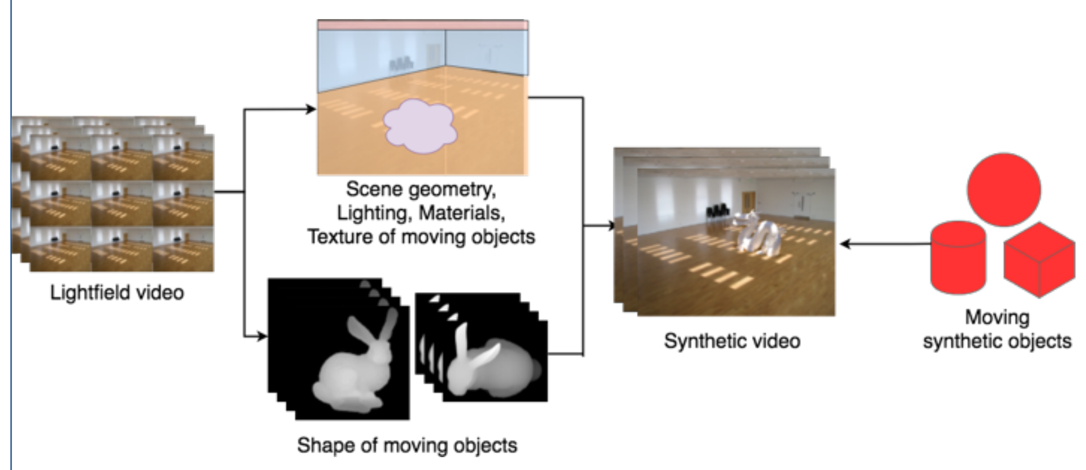
Scene2Synth: Inserting synthetic objects into a dynamic scene using lightfield videos

Anubhav Ashok (**anubhava**); Sree Harsha Kalli (**skalli**)
ProjectID: 1
Robotics Institute, Carnegie Mellon University

Introduction

- We explore an approach to photorealistically insert synthetic objects into dynamic scenes.
- Inserting synthetic objects into a video is challenging as it requires accounting for lighting and physical interactions between moving objects and the scene.



- In our pipeline, we first reconstruct the static 3D scene using the approach described in [2].
- Determining the geometry of deformable moving objects from a video is challenging, we deal with this by using a lightfield video to compute the geometry from the disparity map for each frame.
- Lastly, we add the synthetic objects into the reconstructed dynamic scene and produce a photorealistic rendering.
- We show that this approach is able to generate realistic videos from pre-existing datasets such as [1].

Lightfield Data

- For the dataset we used [1], 16 views are captured using a light field camera with a baseline of 70mm between each camera.
- In our project we show results on the Painter sequence which contains 371 (2048x1088) lightfield images. This scene contains a moving person and dynamic painting in scene.
- We rectify images with respect to middle view

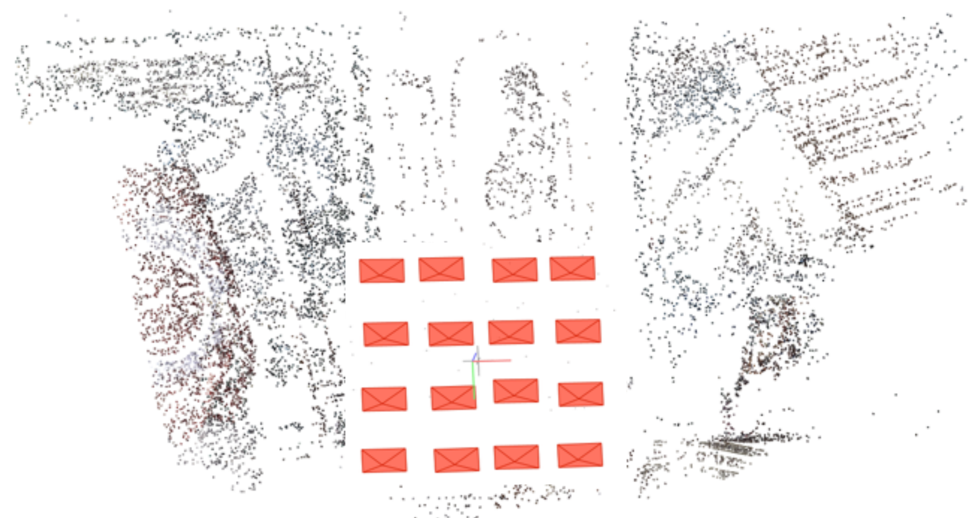


Figure 1. Configuration of cameras

Scene Reconstruction

- We model the geometry, lighting and materials in the scene from a single frame using the method described in [2].
- This required manual annotation of light sources, 3D bounds of the room, 2 pairs of parallel lines and surfaces in the image.
- From these annotations, we could determine the 3 vanishing points, camera pose and focal length of the camera.
- Since we know the 3D bounds of the room, surfaces that are annotated in the image could be extruded to form a 3D object in the scene as shown in **Figure 2**.

- The lighting parameters are expressed as a 6x1 non-negative vector **L** and is determined by iteratively optimizing the following equation.

$$\underset{\mathbf{L}}{\operatorname{argmin}} \sum_{i \in \text{pixels}} \alpha_i (R_i(\mathbf{L}) - R_i^*)^2 + \sum_{j \in \text{params}} w_j (\mathbf{L}_j - \mathbf{L}_{0_j})^2$$
$$\text{subject to: } 0 \leq \mathbf{L}_j \leq 1 \; \forall j$$

- The material properties such as albedo (**ρ**) and direct light (**D**) were determined jointly by the optimization described below.

$$\underset{\rho, D}{\operatorname{argmin}} \sum_{i \in \text{pixels}} |\Delta \rho|_i + \gamma_1 m_i (\nabla \rho)_i^2 + \gamma_2 (D_i - D_{0_i})^2 + \gamma_3 (\nabla D)_i^2$$
$$\text{subject to} \quad B = D + \rho \Gamma, \quad 0 \leq \rho \leq 1, \quad 0 \leq D,$$

Figure 2. Reconstructed scene

Moving Objects From Lightfield

- A disparity map (**d**) is computed from the lightfield images using a Semi-Global Block Matching approach.
- This disparity map is reprojected to 3D (**X,Y,Z**) using the camera intrinsic matrix (**K**) to get a dense point cloud. We assume 0 skew for **K,** non-zero principal point and unequal aspect ratio.

$$K = \begin{bmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \qquad Z = \frac{f * b}{d}, \quad X = \frac{(x_l - p_x) Z}{f_x}, \quad Y = \frac{(y_l - p_y) Z}{f_y}$$

- A surface mesh is formed from this point cloud using Poisson Surface Reconstruction.
- We segmented the moving object from the point cloud reconstruction by using an optical flow based approach to determine motion in the scene.
- Since lighting conditions also varied in the scene, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to obtain the mask for each frame.
- **Figure 3-6** shows the resulting intermediate outputs of each stage of this process.

Figure 3. Original image

Figure 4. Disparity map

Figure 5. Segment map

Figure 6. Reconstruction

Results

- **Figure 8/9** show the final composite image for time t and t+1 where a **metal sphere is moved from one surface to another**.
- Notice the reflection of the person and the table on the sphere.

Figure 7. Original image

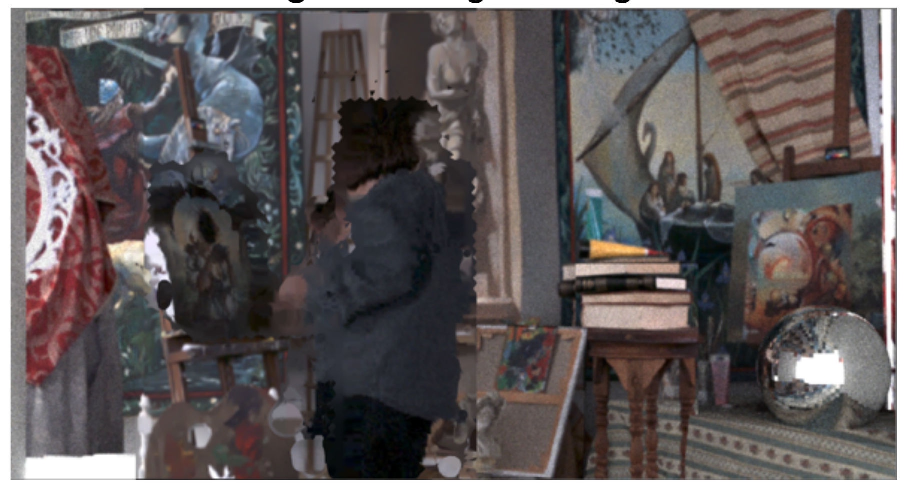


Figure 8. Synthetic image at time t

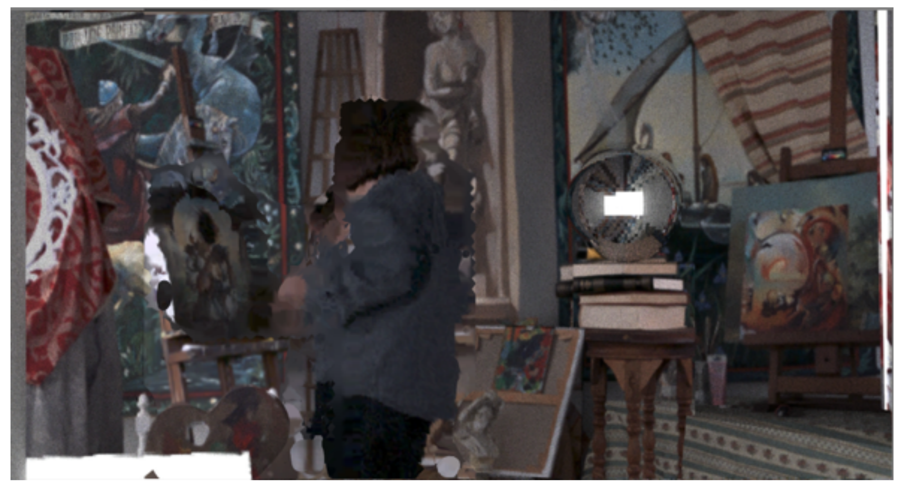


Figure 9. Synthetic image at time t+1

Conclusion

- In our project we have shown that it is possible to render synthetic objects into moving scenes.
- While the synthetic object looks realistic in the scene, the reconstruction of the moving object could be improved by better depth estimation techniques (as described in the paper).
- Overall, this seems like a promising initial result.
- With light field videos being in its early development, we hope to see greater improvement in this method as more comprehensive datasets are released in the future.

References

1. Dataset and Pipeline for Multi-View Light-Field Video. N. Sabater, G. Boisson, B. Vandame, P. Kerbiriou, F. Babon, M. Hog, T. Langlois, R. Gendrot, O. Bureller, A. Schubert, V. Allie. CVPR Workshops, 2017.
2. Karsch, K.; Hedau, V.; Forsyth, D.; Hoiem, D. ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia), 30(6), 2011.