

# Multi-View Interpolation

Luiz Velho  
IMPA

(Based on Slides by Alyosha Efros, and Angjoo Kanazawa)

## Problem Statement

### Input:

A set of calibrated Images



### Output:

A 3D scene representation that  
renders novel views



# Outlook

- Plenoptic Function Revisited
- Light Fields and Lumigraphs
- NeRFs

*Plenoptic Function Revisited*

# Plenoptic Function

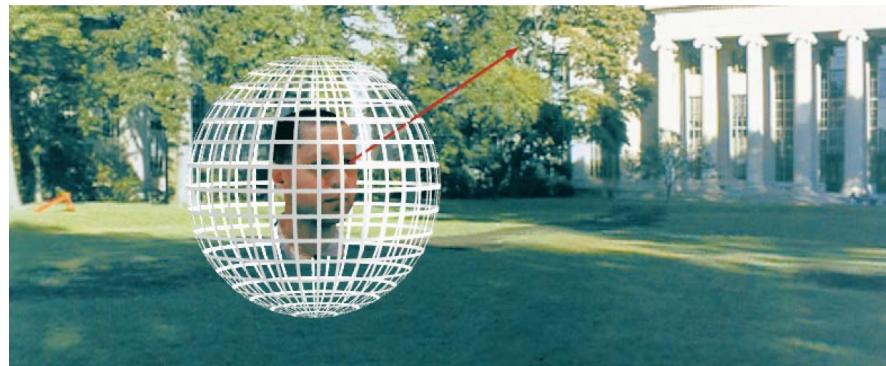


Figure by Leonard McMillan

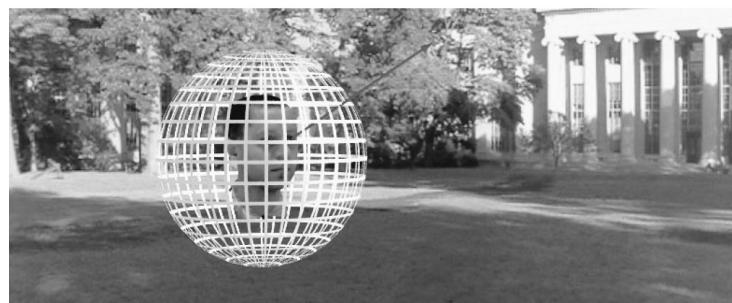
Q: What function describes ALL visual information in a 3D Scene?

A: The Plenoptic Function (Adelson & Bergen '91)

Let's start with a stationary person and try to parameterize everything that they can see...

Slide credit:  
Alyosha Efros

## Grayscale Snapshot



$$P(\theta, \phi)$$

- is intensity of light
  - Seen from a single position (viewpoint)
  - At a single time
  - Averaged over the wavelengths of the visible spectrum

Slides from Alyosha Efros

# Color snapshot



$$P(\theta, \phi, \lambda)$$

- is intensity of light
  - Seen from a single position (viewpoint)
  - At a single time
  - As a function of wavelength

Slides from Alyosha Efros

# A movie

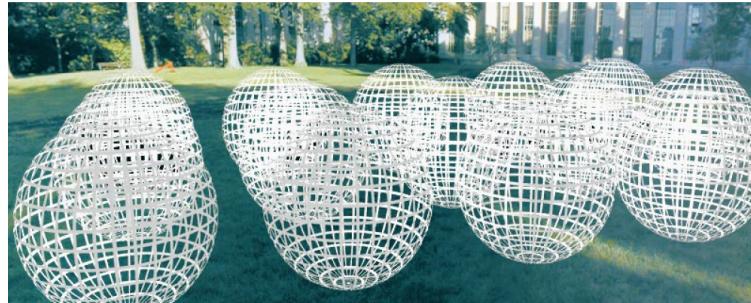


$$P(\theta, \phi, \lambda, t)$$

- is intensity of light
  - Seen from a single position (viewpoint)
  - Over time
  - As a function of wavelength

Slides from Alyosha Efros

# A holographic movie

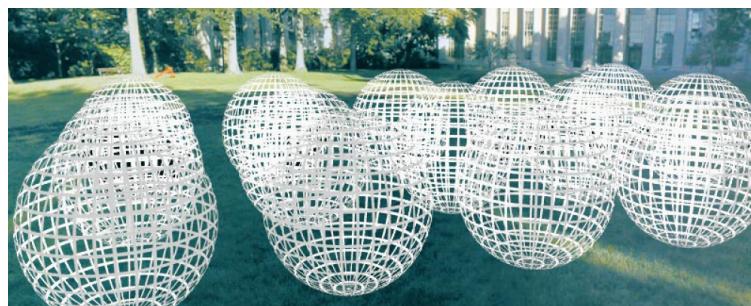


$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$$

- is intensity of light
  - Seen from ANY position and direction
  - Over time
  - As a function of wavelength

Slides from Alyosha Efros

# The plenoptic function

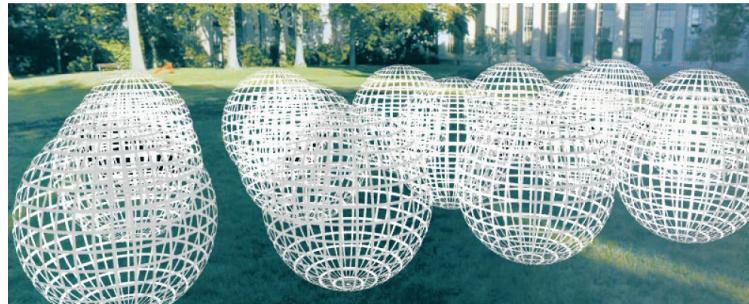


$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$$

7D function, that can reconstruct every position & direction,  
at every moment, at every wavelength  
= it recreates the entirety of our visual reality!

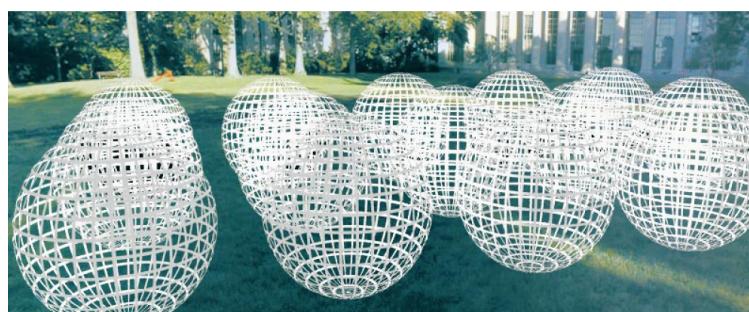
Slides from Alyosha Efros

## Goal: Plenoptic Function from a set of images



- Objective: Recreate the visual reality
- All about recovering photorealistic pixels, not about recording 3D point or surfaces
  - Image Based Rendering
  - aka **Novel View Synthesis**

## Plenoptic Function from a set of images

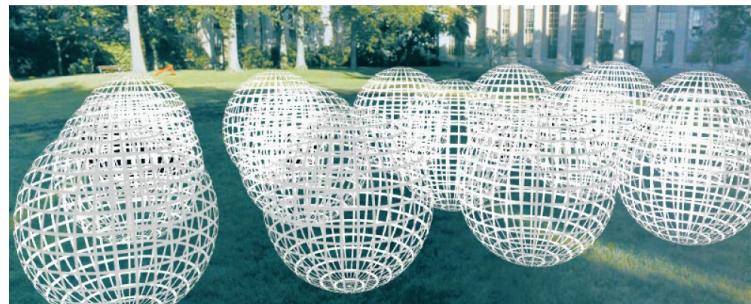


It is a conceptual device

Adelson & Bergen do not discuss how to solve this

# Plenoptic Function

7D function:  
2 – direction  
1 – wavelength  
1 – time  
3 – location



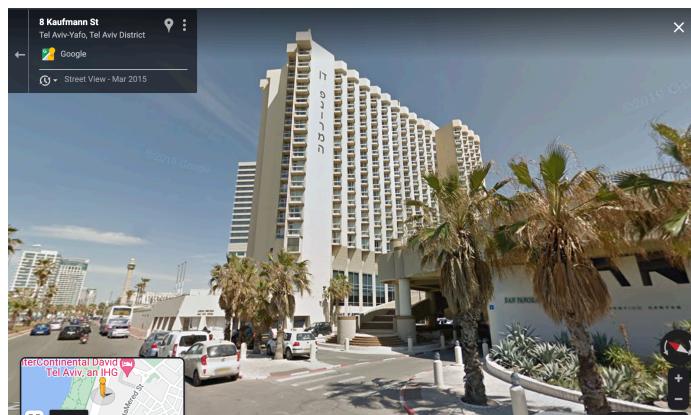
Look familiar  
😊?

$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z) \longrightarrow P(\theta, \phi, V_x, V_y, V_z)$$

Let's simplify:

1. Remove the time
2. Remove the wavelength & let the function output RGB colors

## Example of a 1D sample of the Plenoptic Function



If street view was super dense  
(360 view from any view point)  
then it is the Plenoptic Function

# *Light Field / Lumigraph*

Levoy and Hanrahan, SIGGRAPH 1996  
Gortler et al. SIGGRAPH 1996

# Lightfield / Lumigraph

- An approach for modeling the Plenoptic Function
- Take a lot of pictures from many views

*(sampling)*

**Stanford Gantry**  
128 cameras



Levoy and Hanrahan, SIGGRAPH 1996

Gortler et al. SIGGRAPH 1996

# Lightfield / Lumigraph

- An approach for modeling the Plenoptic Function
- Interpolate the rays to render a novel view  
*(reconstruction)*

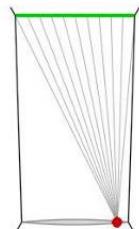


Figure from Marc Levoy

Levoy and Hanrahan, SIGGRAPH 1996  
Gortler et al. SIGGRAPH 1996

# Lightfield / Lumigraph

- An approach for modeling the Plenoptic Function
- Take a lot of pictures (*sampling*)
- Interpolate the rays (*reconstruction*)

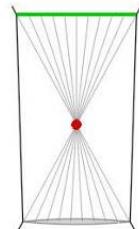


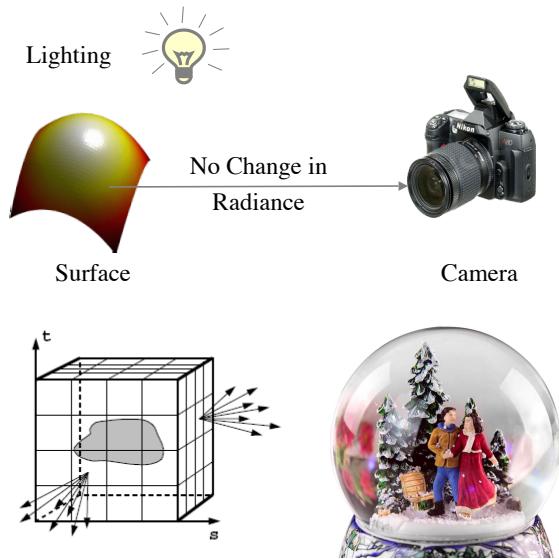
Figure from Marc Levoy

Levoy and Hanrahan, SIGGRAPH 1996

Gortler et al. SIGGRAPH 1996

# Lightfield / Lumigraph

Lightfields assume that the ray shooting out from a pixel is never occluded.



Because of this it only models the  
plenoptic surface:

Figure 1: The surface of a cube holds all the radiance information due to the enclosed object.

*Back to Structure From Motion*

# Structure from Motion

Or Photogrammetry (1850~)  
Long history in Computer Vision

*Proc. R. Soc. Lond. B. 203, 405–426 (1979)*

*Printed in Great Britain*

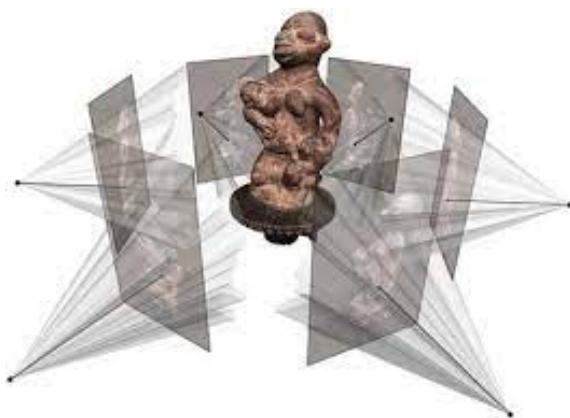
## The interpretation of structure from motion

By S. ULLMAN

*Artificial Intelligence Laboratory, Massachusetts Institute of Technology,  
545 Technology Square (Room 808), Cambridge, Massachusetts 02139 U.S.A.*

# Multi-view Stereo

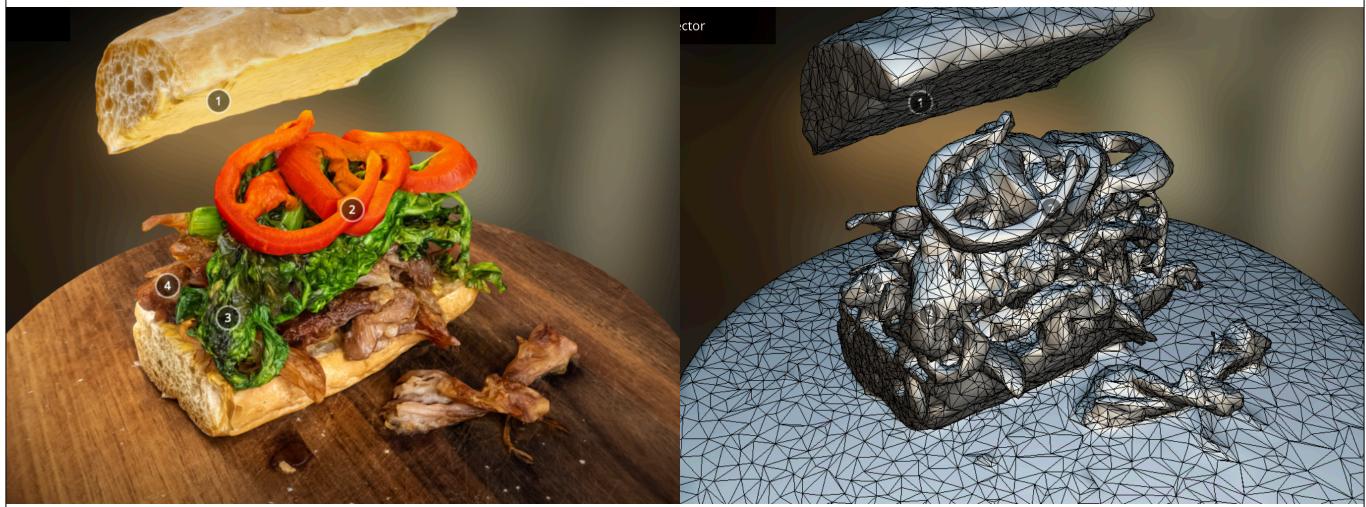
- Problem: Given calibrated cameras, recover highly detailed 3D **surface** model
- Dense photogrammetry, often the output is textured meshes



Figures by Carlos Hernandez, Yasutaka Furukawa

# Multi-View Stereo

Solutions to MVS is what you see for any existing 3D scanning system, ie sketchfab, or what's in your video game



# Multi-View Stereo

Because they often model surfaces, struggles on Thin / Amorphous / Shiny objects



# *Neural Radiance Fields*

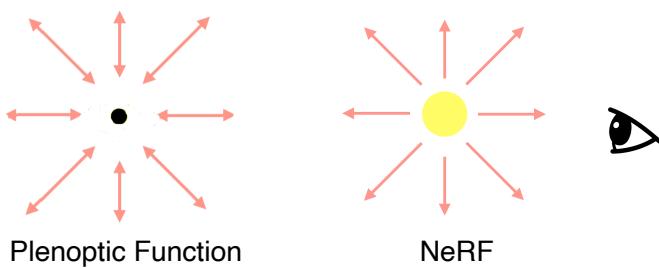
## How NeRF models the Plenoptic Function

$$P(\theta, \phi, V_x, V_y, V_z)$$

Look familiar  
😊?

NeRF takes the same input as the Plenoptic Function!

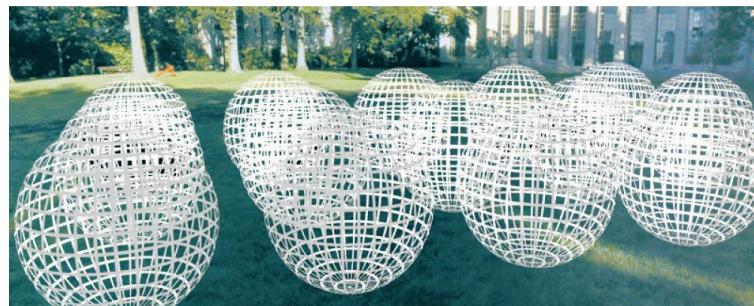
A subtle difference:



NeRF requires the integration along the viewing ray to compute the Plenoptic Function

NeRF assumes a “Baked” Illumination (i.e., All 3D points only emits Radiance)

# 5D function



- For every location (3D), all possible views (2D)
- NeRF models this space with a continuous view-dependent volume with opacity
- The color emitted by every point is composited to render a pixel
- Unlike a light field, the entire 5D plenoptic function can be modeled (you can fly through the world)

## Visualizing the 2D function on the sphere



Outgoing radiance distribution  
for point on side of ship



Outgoing radiance distribution  
for point on water's surface

# Baking in Light

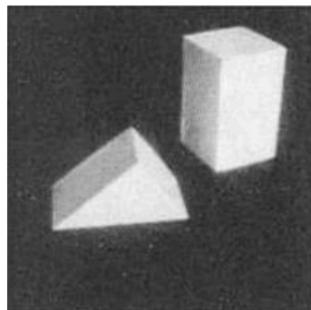


- NeRF can capture non-Lambertian (specular, shiny surfaces) because it models the color in a view-dependent manner
- This is hard to do with meshes unless you model the physical materials & lighting interactions
- But, with Image Based Rendering — All lighting effects are baked in

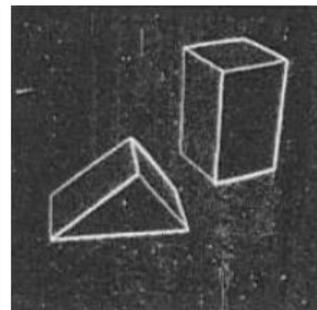
# Analysis-by-Synthesis



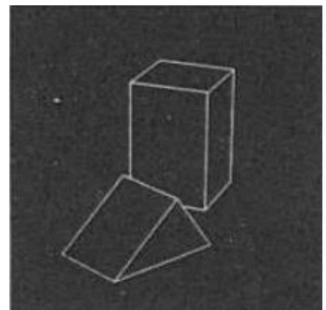
Larry Roberts  
“Father of Computer Vision”



Input image



2x2 gradient operator



computed 3D model  
rendered from new viewpoint

- History goes way back to the **first** Computer Vision paper!  
Roberts: Machine Perception of Three-Dimensional Solids, MIT, 1963

## Power of Analysis-by-Synthesis



- Space Carving: A MVS method that used Colored voxels
- But the optimization method was bottom up then.
- Key is optimization via Analysis-by-Synthesis [Plenoxels, Yu et al. 2022]



Input Image (1 of 45)



Reconstruction



Reconstruction



Reconstruction



Input Image  
(1 of 100)



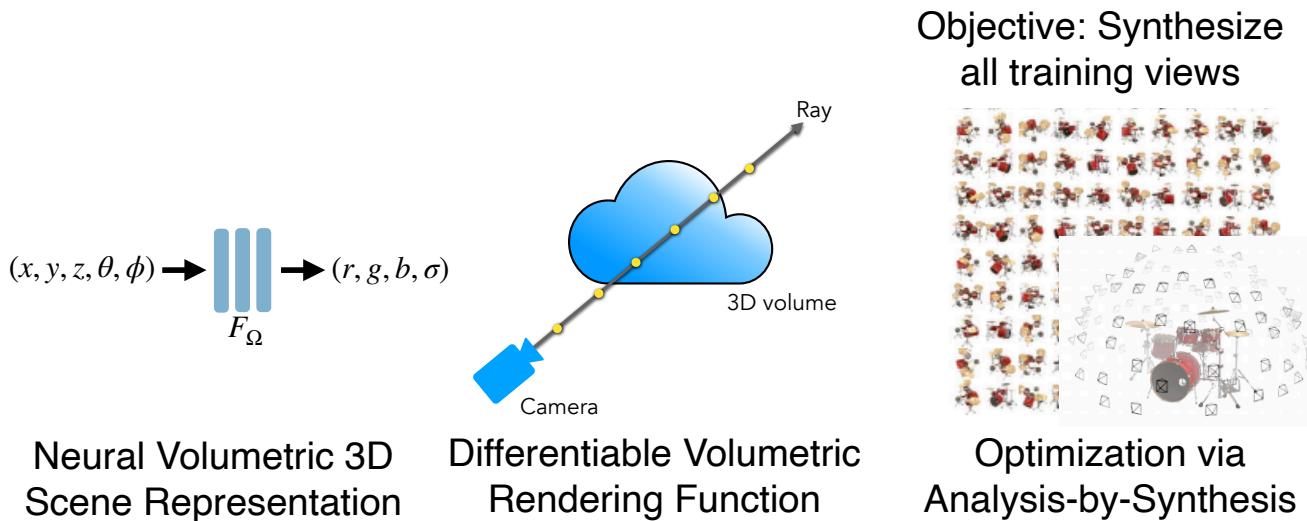
Views of Reconstruction

Kutulakos and Seitz, A Theory of Shape by Space Carving IJCV 2000

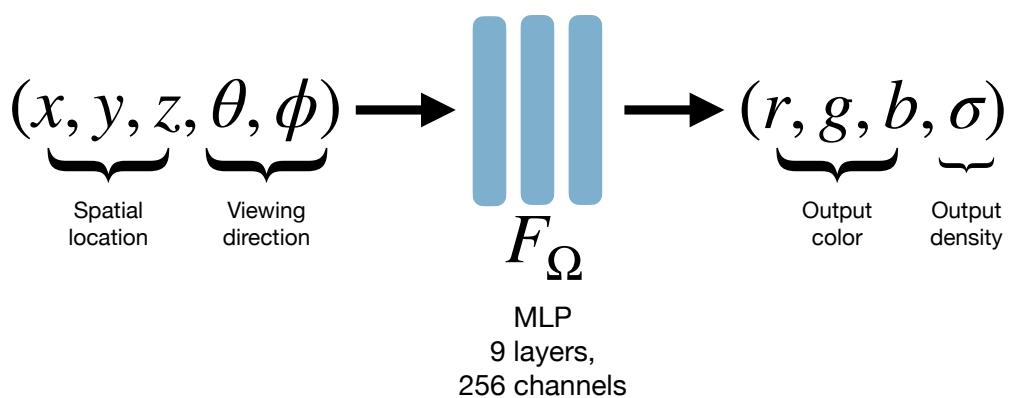
## Analysis by Synthesis Requires Differentiable Renderers

Next: Deep dive into Volumetric Rendering Function

# Three Key Components

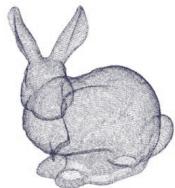
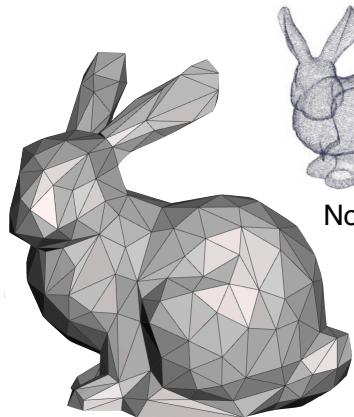


Representing a 3D scene as a continuous 5D function



What kind of a 3D representation is this?

# It is not a Mesh



Not a point cloud  
either



# It is volumetric

It's *continuous voxels* made of shiny transparent cubes

## Unmentioned caveat so far

- Training a NeRF requires a **calibrated** camera!!!!
- Need to know the camera parameters: extrinsic (viewpoint) & intrinsics (focal length, distortion, etc)



**How do we get this from images?**

# NeRF is AFTER Structure from Motion

- In order to train NeRF you need to run SfM/SLAM on the images to estimate the camera parameters
- In this sense, the problem category is same as that of **Multi-view Stereo**



Colmap: Schönberger et al. 2016

## Where NeRF stands

