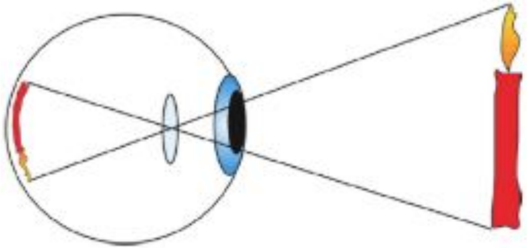
Computational 3D Photography

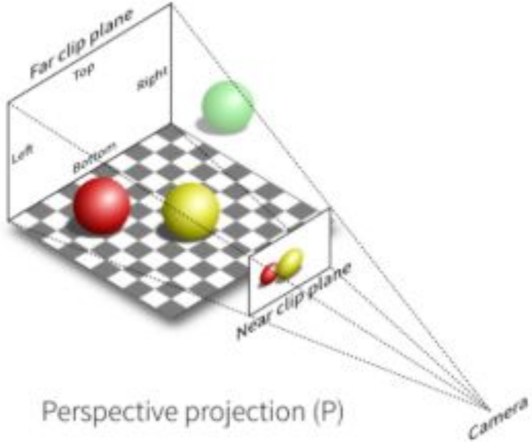
EC522: Computational Optical Imaging

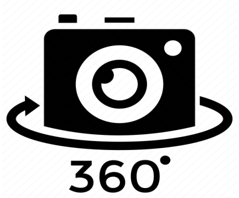
Wangyi Chen, Ian Lee, Susan Zhang

Introduction

- **Depth Perception:** the ability to perceive the distance and spatial relationships between objects in a three-dimensional space







Computational Photography: subset of computational imaging technologies that aim to push the imaging limit in vision systems

Goal

- Existing techniques do not take full advantage of the available monocular depth cues

- Proposal: a nonlinear occlusion-aware optical image formation that models defocus blur at occlusion boundaries more accurately than previous approaches

  Using:

  - an occlusion-aware image formation model
  - a rotationally symmetric aperture
  - an effective preconditioning approach

Monocular refers to a single-camera setup (as opposed to stereo or multi-camera setups to capture images from different viewpoints simultaneously

Related Work

- Monocular Depth Estimation (MDE)
- Computational Imaging for Depth Estimation
- Deep Optics

Related Works: Monocular Depth Estimation (MDE)

- **Deep Learning Approaches:**
  - excel at MDE as they can uncover depth cues not immediately apparent to human observers.
  - Techniques vary in supervision level, loss functions, and constraints used.
- **Combining Surface Normal Estimation:**
  - Some approaches enhance depth estimation by also learning surface normal estimation, leveraging tools like conditional random fields and two-stream CNNs,
    - demonstrating strong performance on datasets such as KITTI and NYU Depth.
- **Incorporating Physical Camera Parameters:**
  - To improve generalization across different datasets,
  - Some methods include camera parameters like defocus blur and focal length in the learning process
    - implicit encoders of depth information.

Related Works: Computational Imaging for Depth Estimation

- **Depth from Defocus (DfD) Variants:**
    - Employ two or more images for depth estimation.
    - Apply computational methods like the sum-modified-Laplacian operator.
    - Experiment with amplitude- and phase-coded apertures.
    - Traditional methods generally lack end-to-end optimization.
- **Utilizing Dual-Pixel Sensors:**
    - Capture stereo image pairs offering enough disparity for depth estimation.
    - Provide innovative ways to capture depth information.
    - Typically depend on hand-crafted designs.
    - Use conventional lenses rather than optimized systems.

Related Works: Deep Optics

- **Joint Design of Optics and Image Processing:**
  - Emphasizes simultaneous design of camera optics and computational methods.
  - Has led to advancements in color filtering, spectral imaging, and high-dynamic-range (HDR) imaging.
- **Recent Advances in Joint Optimization:**
  - Utilization of concentric rings in phase masks to address chromatic aberrations.
  - Joint optimization of phase mask and CNN-based reconstruction targeted at depth estimation.
  - Expansion of deep optics principles to new imaging tasks like simultaneous depth mapping and multispectral scene information extraction.

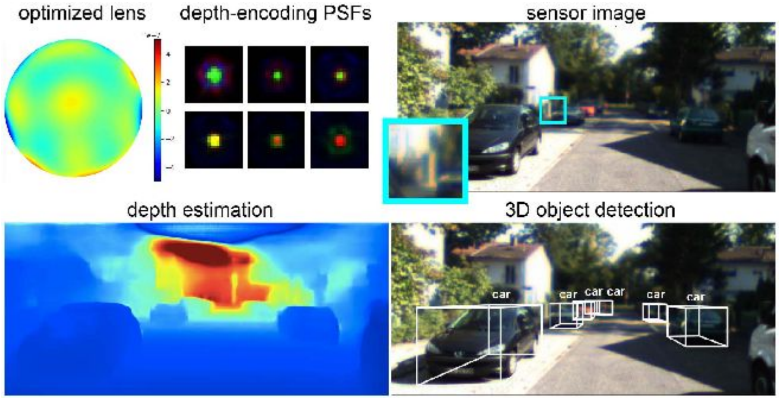
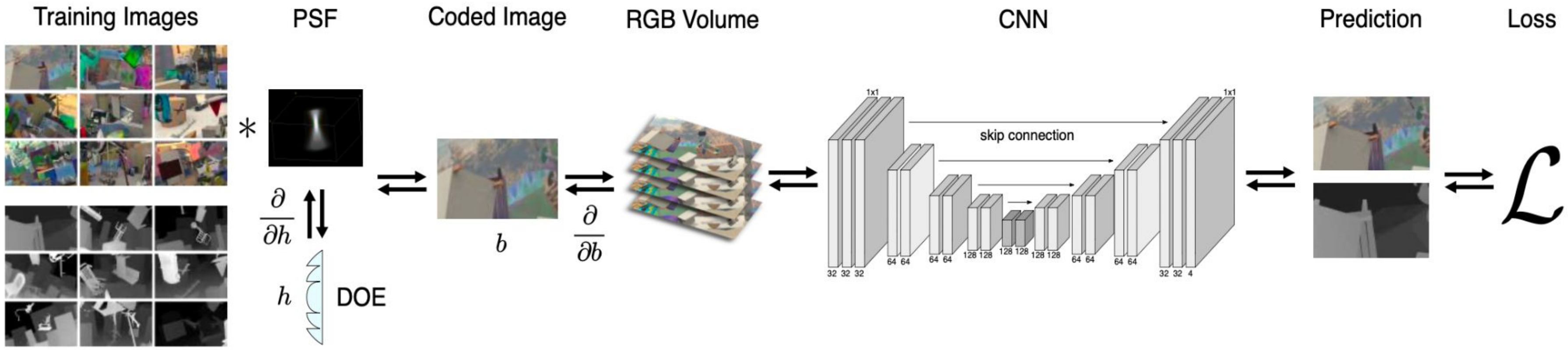


Figure 1. We apply deep optics, i.e. end-to-end design of optics

PHASE-CODED 3D IMAGING SYSTEM



IMAGING SYSTEM: Camera

- **Modified Optical System:**
  - Utilizes a conventional photographic compound lens to focus the scene onto a sensor.
  - Includes modifications to incorporate a diffractive optical element (DOE) within the aperture plane.
- **Learnable Phase-Coded Aperture:**
  - Enables direct control over the depth-dependent point spread function (PSF).
  - Alterations in the DOE's surface profile facilitate this control.
- **Purpose of Modifications:**
  - The camera modifications support end-to-end (E2E) optimization.
  - Aim to improve depth estimation from a single image through the imaging system.

IMAGING SYSTEM: PSF

- **Depth-Dependent Modeling:**
  - The point spread function (PSF) is modeled to change with depth.
  - Incorporates the effects of wavelength and radial distances from the aperture and sensor planes.
  - Essential for embedding depth information within images.
- **Radially Symmetric Design:**
  - Simplifies computational demands for the imaging system.
  - Significantly reduces the number of parameters required for optimization.
  - Lowers memory requirements during the optimization process.
- **Impact on Depth Estimation:**
  - Depth-variant PSF is critical for capturing defocused images with depth information.
  - Enables the system to perform accurate depth estimation from these images.

IMAGING SYSTEM: PSF (Cont.)

- **Mathematical Formulation:** The PSF is defined by the equation:

$$\mathrm{PSF}(\rho, z, \lambda) = \left| \frac{2\pi}{\lambda s} \int_0^\infty r D(r, \lambda, z) P(r, \lambda) J_0(2\pi \rho r)\, dr \right|^2 .$$

Where:

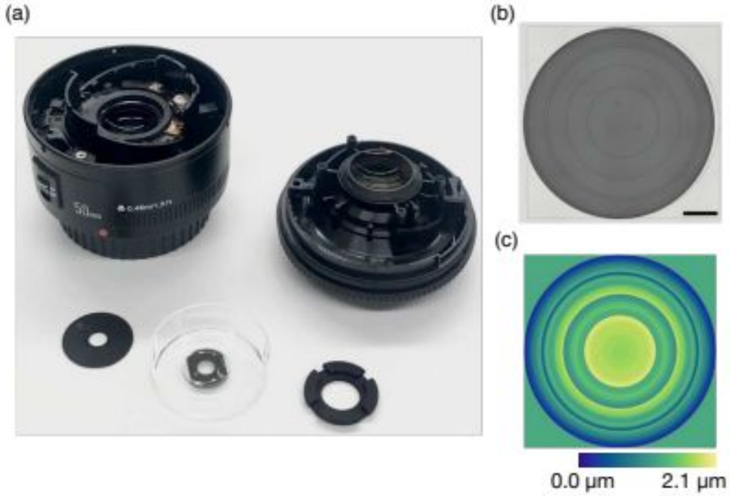
- ρ and r are the radial distances on the sensor and aperture planes, respectively.
- λ is the wavelength of light.
- s is the distance between the lens and the sensor.
- D(r,λ,z) is the defocus factor, modeling how the PSF varies with depth (z) for a point at some distance from the lens.
- P(r,λ) represents the phase delay introduced by the DOE, which is a function of the radial position and wavelength.
- J0 is the zeroth-order Bessel function of the first kind, accounting for the radial symmetry of the PSF.

IMAGING SYSTEM: Phase Code

- **Definition and Purpose:**
  - A phase code is a specific pattern on the DOE that modulates light to create a precise PSF.
  - It is essential for embedding depth information into an image.
- **Implementation through DOE:**
  - Implemented via the DOE's surface profile, which is fine-tuned during end-to-end (E2E) training.
  - Optimization aims to find a surface profile that enhances depth estimation accuracy.
- **Rotational Symmetry for Efficiency:**
  - The phase code's design is rotationally symmetric, streamlining the computational process.
  - This symmetry helps reduce both the complexity and the computational load during optimization.

IMAGING SYSTEM: DOE Design

- **Optimized for Depth Estimation:**
  - specifically designed and optimized to work in conjunction with the neural network for the task of depth estimation
  - Aims to maximize the system's performance.
- **Utilization of Rotational Symmetry:**
  - A rotationally symmetric design is adopted for the DOE, which significantly reduces computational demands.
  - This design choice facilitates more efficient training and optimization.
- **Effectiveness in E2E Optimization:**
  - The DOE's design is a critical component of the end-to-end (E2E) optimization process.
  - Its structure is iteratively refined, improving the system's ability to accurately estimate depth from defocused images.



IMAGING SYSTEM: Non-linear Image Formation

$$b(\lambda) = \sum_{k=0}^{K-1} \tilde{l}_k \prod_{k'=k+1}^{K-1} (1 - \tilde{\alpha}_{k'}) + \eta$$

- $l_k := \frac{(PSF_k(\lambda) * l_k)}{E_k(\lambda)}$: Normalized contribution of each depth layer to the image formation

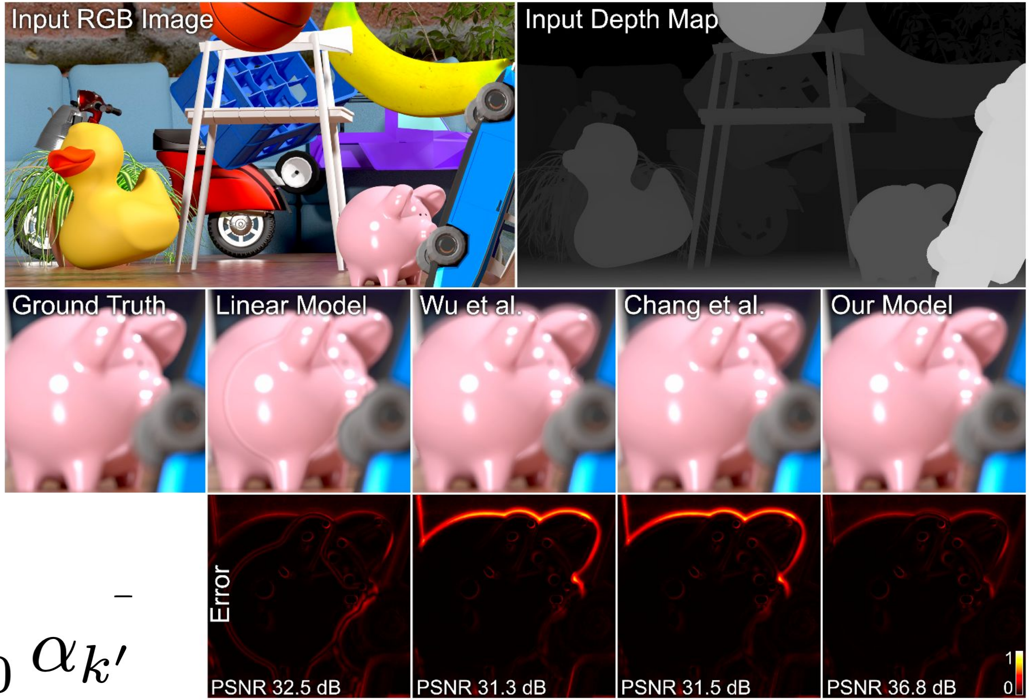
- $\alpha_k := \frac{(PSF_k(\lambda) * \alpha_k(\lambda))}{E_k(\lambda)}$: Normalized binary mask for each depth layer, indicating occlusion effects.

- $E_k(\lambda) := PSF_k * \sum_{k'=0}^{k} \alpha_{k'}$: Normalization factor for realistic energy levels at depth transitions

- η: Additive noise

IMAGING SYSTEM: Depth Image Formation

- Generate defocused images with RGBD input
- Performance excels at depth incontinuities
- More realistic than linear models



Energy at transitions recovered by normalization:

$$E_k(\lambda) \;:=\; \mathrm{PSF}_k * \textstyle\sum_{k'=0}^{k} \alpha_{k'}$$

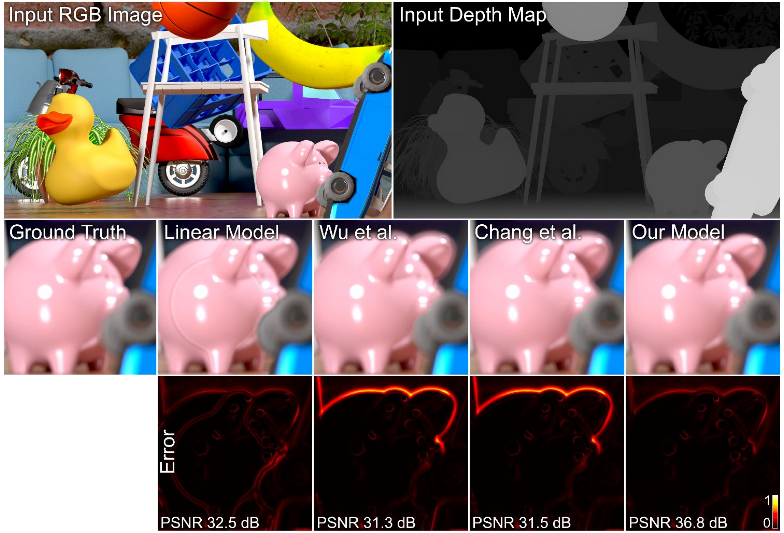
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9466261

IMAGING SYSTEM: CNN

Goal: Accurately predict depth map

Ray tracing: computation and storage intensive

CNN: Less accurate but efficient



https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9466261

Image Segmentation

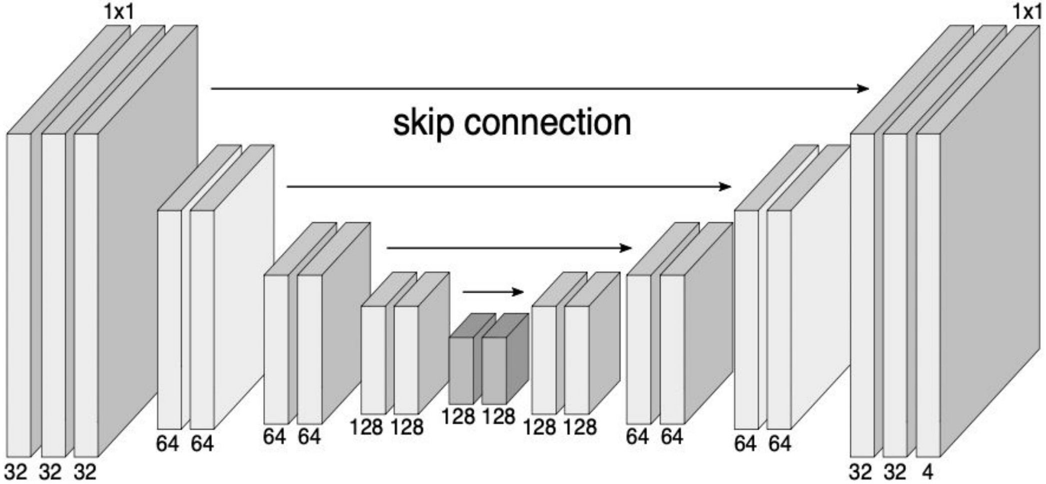
Pixel Level Object Recognition
Pixel to {categories}

Pixel Level Depth Segmentation
Pixel to {depths}

https://medium.com/anolytics/image-segmentation-the-deep-learning-approach-1e48035dfade
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9466261

IMAGING SYSTEM: U-net CNN

- Encoder-Decoder Structure
- Efficient Use of Data
- Low level features(conv layer) + high level features(skip connection)
- Precise Localization
- Training efficiency
- 1 million parameters



https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9466261

IMAGING SYSTEM: CNN Loss Function

$$\mathcal{L} = \psi_{\mathrm{RGB}}\mathcal{L}_{\mathrm{RGB}} + \psi_{\mathrm{Depth}}\mathcal{L}_{\mathrm{Depth}} + \psi_{\mathrm{PSF}}\mathcal{L}_{\mathrm{PSF}}$$

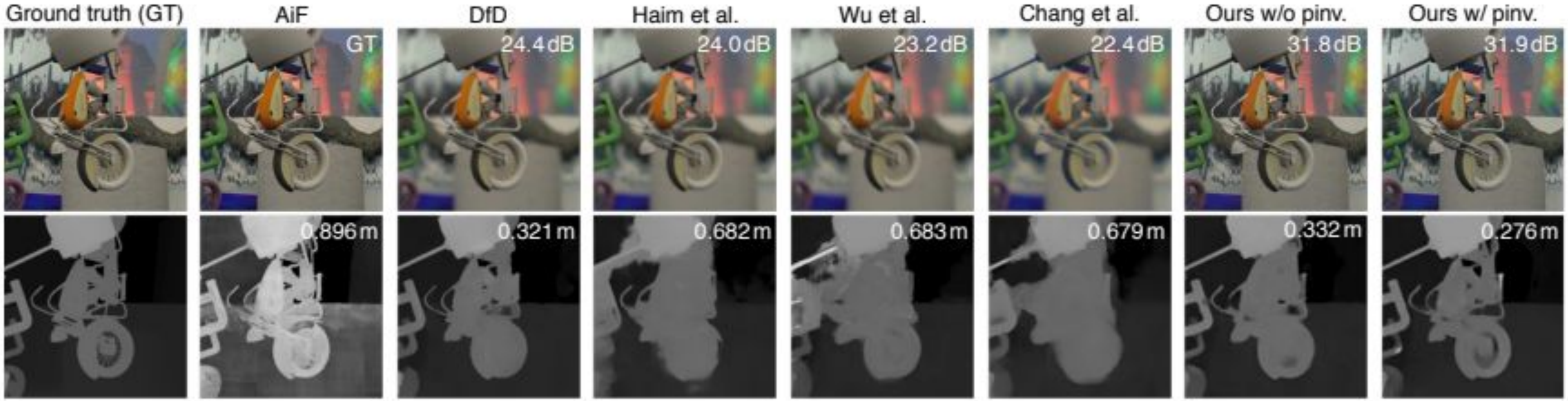
L_{RGB}: Loss for RGB image estimation.
L_{Depth}: Loss for depth map estimation.
L_{PSF}: Regularization loss for the PSF.
ψ_{RGB}: Weighting factor for the RGB image estimation loss.
ψ_{Depth}: Weighting factor for the depth estimation loss.
ψ_{PSF}: Weighting factor for the PSF regularization loss.

IMAGING SYSTEM: CNN Training Details

- 100 epochs
- Adam Optimizer(β_1 = 0.9, β_2 = 0.99)
- Batch size 3
- Best model taken from lowest validation set loss

Experiment

- *FlyingThings3D* - pairs of an RGB image and its corresponding depth maps
  - Training: 22K
    - Training: 18K
    - Validation: 4K
  - Testing: 8K



https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9466261

Limitations and Future Work

- Limitations:
  - Unable to fully represent the continuous nature of physical systems
    - Discretization of depth layers and PSF simulation at discrete wavelengths
  - Costly
    - Increase memory consumption when treating image and depth reconstruction tasks separately could enhance network capacity.
- Future Work:
  - Optimizing fabrication processes and diffraction efficiency of optical elements, alignments, and calibration of integrated systems from differences between the image formation model and the physical system
  - Explore different network architectures optimized to capture both physical information provided by coded defocus blur and contextual cues encoded by pictorial scene information

Thank you