Lecture 6: Optics in Digital Camera

Making your DSLR into a 3D camera

Vivek Boominathan
Rice University
vivekboominathan.com

CSC 292/572, Fall 2020
Mobile Visual Computing
Camera optics

DSLR lenses

Smartphone cameras

Rear Camera
Quad Camera
Ultra Wide: 12MP, F2.2 (120°)
Wide-angle: 108MP, PDAF, OIS, F1.8 (79°)
Telephoto: 48MP, PDAF, OIS, F3.5 (24°)
DepthVision

Front Camera
Selfie: 40MP, PDAF, F2.2 (80°)

Samsung Galaxy S20 Ultra
Camera optics

• Lens specifications:
  – Focal length
  – Aperture size

• Lens effects:
  – Field-of-view (optical zoom)
  – Depth-of-field
  – Perceptual distortion
Smartphone lens choices

Shortest focal length, Small aperture
Long focal length, Large aperture
Longest focal length, Largest aperture

Samsung Galaxy S20 Ultra

Rear Camera
Quad Camera
Ultra Wide: 12MP, F2.2 (120°)
Wide-angle: 108MP, PDAF, OIS, F1.8 (79°)
Telephoto: 48MP, PDAF, OIS, F3.5 (24°)
DepthVision

Front Camera
Selfie: 40MP, PDAF, F2.2 (80°)
Lens effects example: perceptual distortion

When you take a selfie

When you look in the mirror

Changing focal length
Lens effects (Today’s focus) – In focus and Out of focus
Lens effects – In focus and Out of focus
Lens effects – In focus and Out of focus

In focus

Circular lens aperture

Point spread function (PSF)

Image of a point light source

Lens

Camera sensor

Point spread function

Focal plane
Lens effects – In focus and Out of focus

Out of focus

Object

Focal plane

Lens

Circular lens aperture

Point spread function (PSF)

Image of a point light source

Camera sensor

Point spread function
Jazz up your photography – Bokeh effects
Lens effects – In focus and Out of focus

5-blade Circular Aperture (with No Optical Vignetting)
1/5.6 (4 Stops Closed)

Out of focus

Object

Focal plane

5-blade Circular Aperture

Circular lens aperture

Point spread function (PSF)

Image of a point light source

Camera sensor

Point spread function

Lens
Lens and defocus

Image of a point light source

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Lens and defocus

Lens’ aperture

Image of a point light source

Object

Lens

Camera sensor

Focal plane

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Lens and defocus

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Lens and defocus

Object → Lens → Camera sensor

Lens’ aperture

Image of a point light source

Focal plane

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Depth and defocus

Infer depth by analyzing local scale of defocus blur

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Defocus as local convolution

Input defocused image

Calibrated blur kernels at different depths

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Defocus as local convolution

$$y = f_k \otimes x$$

Input defocused image

Local sub-window

Calibrated blur kernels at depth $k$

Sharp sub-window

Depth $k=1$:

Depth $k=2$:

Depth $k=3$:

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Convolution intuition

In focus:
Replace every point light with in-focus psf.
“Convolution with in-focus psf.”
“Sharp”

Out of focus:
Replace every point light with out-of-focus psf.
“Convolution with out of focus psf.”
“Blurry”

Circular lens aperture

In focus

Out of focus

Point spread functions (PSFs)

Image of a point light source
Defocus as local convolution

\[ y = f_k \otimes x \]

Local sub-window \hspace{2cm} Calibrated blur kernels at depth \( k \) \hspace{2cm} Sharp sub-window

Depth \( k=1 \):

Depth \( k=2 \):

Depth \( k=3 \):

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Challenges

• Hard to deconvolve even when kernel is known

Input \[ \Rightarrow \text{Deconvolution Algorithm} \rightarrow \text{Ringing with the traditional Richardson-Lucy deconvolution algorithm} \]

• Hard to identify correct scale:

\[ AB \]  

\[ \Rightarrow \text{Incorrect scale} \]

\[ AB \]  

\[ \Rightarrow \text{Correct scale} \]

\[ AB \]  

\[ \Rightarrow \text{Incorrect scale} \]

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Deconvolution is ill posed

\[ f \otimes x = y \]

**Solution 1:**

**Solution 2:**

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Conventional lens blurring

Conventional aperture

Calibrated blur kernels at different depths

Bad for depth estimation

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
• Mask (code) in aperture plane
  
  - make defocus patterns different from natural images and easier to discriminate

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Solution: lens with occluder

Aperture pattern

Image of a defocused point light source

Object

Focal plane

Lens

Camera sensor

Point spread function

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Solution: lens with occluder

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Solution: lens with occluder

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Solution: lens with occluder

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Solution: lens with occluder

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Why coded?

Coded aperture - reduce uncertainty in scale identification

Larger scale

Correct scale

Smaller scale

Exaggerates the differences

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Filter Design

Analytically search for a pattern maximizing discrimination between images at different defocus scales (KL-divergence)

Account for image prior and physical constraints

See paper for details

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Build your own coded aperture

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Voila!

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Optimization problem

Try deblurring with 10 different aperture scales

\[
x = \arg \min_x \left[ \| f \otimes x - y \|^2 + \lambda \sum_i \rho(\nabla x_i) \right] + \text{regularization}
\]

Keep minimal error scale in each local window
Depth estimation

Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
All-focused (deconvolved)
Anat Levin et. al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Comparison – Conventional aperture result

Ringing due to wrong scale estimation
Comparison – Conventional aperture result

Anat Levin et al. “Image and Depth from a Conventional Camera with a Coded Aperture”
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Recap until now

- Add coded aperture to enable depth reconstruction.

- Design:
  - Find the patterns such that the properties of scaled pattern differ analytically.
  - Greedy search in a combinatorial solution space

- Depth reconstruction:
  - Reconstruct with different scales of the pattern.
  - Assign depth when the reconstruction is sharpest.
Recap until now

• Design and reconstruction are **disjoint**.

• Depth reconstruction is based on **heuristics**.
  – Lack deeper understanding of images.

• Depth reconstructions are **ok-ish**.

To improve depth reconstruction:

• Can we jointly design the optics and reconstruction?

• Use deep learning based methods that have deeper understanding of images?
JOINT DESIGN OF OPTICS AND RECONSTRUCTION
Joint design with Machine Learning

Scene → Coded aperture → Lens → coded mask → Sensor → Optical System → Depth reconstruction algorithm → Digital network → End-to-end neural network → Depth map
Defocus of general lens

Far out of focus
In focus
Near out of focus

PSFs at different depths

× Identical PSF at both sides of the focal plane.
× Impossible to tell the depth based on the blur size.
Independent-crafted designs

Veeraraghavan et al., 2007

Levin et al., 2007 (previously discussed)

The difference in PSFs across depth is only in scale.
PhaseCam3D sensor – Three upgrades

- Phase-coded aperture
- Digital network
- End-to-end learning


**Model for the optical system**

**PSF formulation**

\[
PSF = |F\{P\}|^2
\]

\[
P = \exp[j\phi^M_{\lambda}(h) + j\phi^{DF}_{\lambda}(z)]
\]

Given the phase mask glass height profile, determines the PSFs for different defocus.

---

Model for the optical system

Knowing the depth and the defocus PSFs, compute how the sensor captures the image.

Image formulation

\[ I_{sensor}(h) = \sum_{z} I_{obj}(z) \otimes PSF_{\lambda}(h, z) + N(0, \sigma^2) \]

Noise

Model for the optical system

PSF formulation

\[
PSF = |F\{P\}|^2 \\
P = \exp[j\phi^M_\lambda(h) + j\phi^DF_\lambda(z)]
\]

- Phase mask
- Defocus

Image formulation

\[
I_{sensor}(h) = \sum_z I_{obj}(z) \otimes PSF_\lambda(h, z) + N(0, \sigma^2)
\]

- Noise

Differentiable model

Phase-coded aperture
Model for the digital network

- Pixel-wise prediction
- Skip connection

Coded image

U-Net

- conv 3x3, ReLU, BN
- copy and concatenate
- max pool 2x2
- upsampling 2x2
- conv 1x1, sigmoid

Estimated depth
Train the network

- RGBD dataset
  - Experimental: boundary mismatch, missing depth
  - Synthetic: precise texture and depth

- Loss functions
  - Root mean square (RMS) loss
  - Gradient loss

- Mask parameters
  - Coefficients of Zernike polynomial basis

PhaseCam3D: an end-to-end learning approach

Scene

Optical System

PhaseCam3D sensor

Sensor

Lens

Phase mask

Digital network

Depth map

- Phase-coded optical model
- Digital network
- End-to-end learning

Designed mask and defocus PSFs

Phase mask
glass height map

PSFs at different depths
Far ↔ Focus ↔ Near

Simulation results

Sharp image

Coded image capture

Estimated disparity

Digital network

Comparison with independent-crafted designs

<table>
<thead>
<tr>
<th></th>
<th>Levin et al.</th>
<th>Veeraraghavan et al.</th>
<th>Ours</th>
<th>PhaseCam3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground truth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coded images</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disparity map</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. RMS loss</td>
<td>0.052</td>
<td>0.054</td>
<td>0.028</td>
<td></td>
</tr>
</tbody>
</table>

(Previously discussed)

(Joint-design)
Fabricate the learned phase mask

Phase mask
glass height map

Fabricated phase mask

Experimental results

Indoor scenes

Outdoor scenes

Coded images

Disparity map

Experimental results

Indoor scenes

Outdoor scenes

Coded images

Depth map

meters

meters
Accuracy evaluation: compare with Kinect

Coded Images

Estimated depth by PhaseCam3D

Estimated depth by Kinect

Error:

$$\sigma_{PhaseCam3D} = 1.25\text{cm}$$
Summary

• Conventional lens aperture has depth-dependent effect but is in-effective for depth estimation.

• Coding the aperture of lens + computation allows us to extract depth information.

• Modifying the camera system + computational algorithm to extract additional information is the essence of Computational Photography/Imaging.

• Jointly designing both the camera system and the information extraction algorithm allows for the best performance to be realized.