Project 4 Results

Image Depth Estimates

• Why would depth be useful?
  – segmentation, navigation, interaction, and even recognition.

• How can we estimate it?
  – stereo / structured lighting / structure-from-motion, vanishing point, parallel line reasoning, explicit scene and object recognition, time of flight measurement, haze measurement.
Another depth cue

- Are these at the same depth?
Image and Depth from a Conventional Camera with a Coded Aperture

Anat Levin, Rob Fergus, Frédéric Durand, William Freeman

MIT CSAIL
Single input image:

Output #1: Depth map
Single input image:

Output #1: Depth map

Output #2: All-focused image
Build your own coded aperture
Voila!
Lens and defocus

Lens’ aperture

Image of a point light source

Focal plane

Lens

Camera sensor

Point spread function
Lens and defocus

Lens’ aperture

Image of a defocused point light source

Object

Lens

Camera sensor

Focal plane

Point spread function
Lens and defocus

Image of a defocused point light source

Object

Lens

Camera sensor

Point spread function

Focal plane
Lens and defocus

Lens' aperture → Image of a defocused point light source

Object → Lens → Camera sensor

Focal plane

Lens' aperture

Point spread function
Lens and defocus

Lens’ aperture

Image of a defocused point light source

Object

Lens

Camera sensor

Focal plane

Point spread function
Depth and defocus

Depth from defocus:
Infer depth by analyzing local scale of defocus blur
Challenges

• Hard to discriminate a smooth scene from defocus blur

• Hard to undo defocus blur

Out of focus

Ringing with conventional deblurring algorithm
Key contributions

• Exploit prior on natural images
  - Improve deconvolution
  - Improve depth discrimination

• Coded aperture (mask inside lens)
  - make defocus patterns different from natural images and easier to discriminate
Related Work

• Depth from (de)focus  
  e.g. Pentland, Chaudhuri, Favaro et al.

• Plenoptic/ light field cameras  
  e.g. Adelson and Wang, Ng et al.

• Wave front coding  
  e.g. Cathey & Dowski

• Coded apertures for light gathering:  
  e.g. Fenimore and Cannon

• Blind Deconvolution  
  e.g. Kundur and Hatzinakos, Fergus et al, Levin

Never recover both depth AND full resolution image from a single image

Except: Veeraraghavan, Raskar, Agrawal, Mohan, Tumblin  SIGGRAPH07  
optimize debluring while we optimize depth discrimination
Defocus as local convolution

Input defocused image

Calibrated blur kernels at different depths
Defocus as local convolution

$$y' = f_k \otimes \chi$$

**Input defocused image**

- **Local sub-window**
- **Calibrated blur kernels at depth $k$**
- **Sharp sub-window**

**Depth $k=1$:**

**Depth $k=2$:**

**Depth $k=3$:**
Overview

Try deconvolving local input windows with different scaled filters:

- Larger scale
- Correct scale
- Smaller scale

Somehow: select best scale.
Challenges

• Hard to deconvolve even when kernel is known

![Image of input and deconvolved images showing ringing with traditional Richardson-Lucy deconvolution algorithm.]

• Hard to identify correct scale:

  - Larger scale
  - Correct scale
  - Smaller scale
Deconvolution is ill posed

\[ f \otimes x = y \]
Deconvolution is ill posed

\[ f \otimes x = y \]

Solution 1:

\[ \begin{array}{c}
\text{hexagon} \\
\otimes \\
\text{54} \\
\otimes \\
\text{?:} \\
\end{array} = \begin{array}{c}
\text{54} \\
\end{array} \]

Solution 2:

\[ \begin{array}{c}
\text{hexagon} \\
\otimes \\
\text{54} \\
\otimes \\
\text{?:} \\
\end{array} = \begin{array}{c}
\text{54} \\
\end{array} \]
Idea 1: Natural images prior

What makes images special?

Natural images have sparse gradients

- put a penalty on gradients
Deconvolution with prior

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

- Convolution error
- Derivatives prior

Equal convolution error

Low

High
Comparing deconvolution algorithms

(Non blind) deconvolution code available online:
http://groups.csail.mit.edu/graphics/CodedAperture/

$$\rho(\nabla x) = \|\nabla x\|^2$$
“spread” gradients

$$\rho(\nabla x) = \|\nabla x\|^{0.8}$$
“localizes” gradients
Comparing deconvolution algorithms

(Non blind) deconvolution code available online: http://groups.csail.mit.edu/graphics/CodedAperture/

\[ \rho(\nabla x) = \|\nabla x\|^2 \]

“spread” gradients

\[ \rho(\nabla x) = \|\nabla x\|^{0.8} \]

“localizes” gradients

Input

Richardson-Lucy

Gaussian prior

Sparse prior
Try deconvolving local input windows with different scaled filters:

- Larger scale
- Correct scale
- Smaller scale

Somehow: select best scale.

Challenge: smaller scale not so different than correct
Idea 2: Coded Aperture

- Mask (code) in aperture plane
  - make defocus patterns different from natural images and easier to discriminate

Conventional aperture

Our coded aperture
Solution: lens with occluder
Solution: lens with occluder

Aperture pattern

Image of a defocused point light source

Object

Focal plane

Lens with coded aperture

Camera sensor

Point spread function
Solution: lens with occluder

Aperture pattern

Image of a defocused point light source

Object

Focal plane

Lens with coded aperture

Camera sensor

Point spread function
Solution: lens with occluder

Aperture pattern → Image of a defocused point light source

Object → Lens with coded aperture → Camera sensor

Focal plane
Solution: lens with occluder

Aperture pattern

Image of a defocused point light source

Object

Lens with coded aperture

Camera sensor

Point spread function

Focal plane
Solution: lens with occluder

- Aperture pattern
- Image of a defocused point light source
- Lens with coded aperture
- Camera sensor
- Point spread function
Why coded?

Coded aperture - reduce uncertainty in scale identification

Larger scale
Correct scale
Smaller scale
Filter Design

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

Account for image prior and physical constraints

More discrimination between scales

Less discrimination between scales

Sampled aperture patterns

Conventional aperture

See paper for details
Zero frequencies - pros and cons

Previous talk:

No zero frequencies:
- Filter can be easily inverted
- Weaker depth discrimination

Our solution:

Include zero frequencies:
- Zeros improve depth discrimination
- Inversion difficult
- Inversion made possible with image priors
Depth results
Regularizing depth estimation

Try deblurring with 10 different aperture scales

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

\[ \left| f \otimes x - y \right|^2 \]

Convolution error

Derivatives prior

Keep minimal error scale in each local window + regularization

Input

Local depth estimation

Regularized depth
Regularizing depth estimation

Input

Local depth estimation

Regularized depth
Sometimes, manual intervention

Input

Local depth estimation

Regularized depth

After user corrections
All focused results
All-focused (deconvolved)
Close-up

Original image

All-focus image
All-focused (deconvolved)
Close-up

Original image

All-focus image

Naïve sharpening
Comparison - conventional aperture result

Ringing due to wrong scale estimation
Comparison- coded aperture result
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Application: Digital refocusing from a single image
Coded aperture: pros and cons

+ Image AND depth at a single shot
+ No loss of image resolution
+ Simple modification to lens
- Depth is coarse
  unable to get depth at untextured areas, might need manual corrections.
+ But depth is a pure bonus
- Lose some light
+ But deconvolution increases depth of field
Deconvolution code available

http://groups.csail.mit.edu/graphics/CodedAperture/
50mm f/1.8: $79.95
Cardboard: $1
Tape: $1
Depth acquisition: priceless