Lecture 1: Introduction and Overview
Hello! Meet Each Other!

Prinaya Choubey  Dominique Dorvil  Adam Inskip  Jiangqi Su  Wayne Xu  Kevin Martinez  Asser E.a.a. Elfeky

Yi Yang  Ryan Maciel  Elias Neuman-Donihue  Arnav Sharma  Sudhanshu Gupta  Jack Valinsky  Leah Goodwin
Scope of the Course

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Computer Vision (Semantics understanding)

Image Sensing (Optical to electrical signal transformation) → Computer Graphics (Visibility + Shading) → Display (Generating lights) → Video/Image De/Compression → Cloud/Storage

Modeling (Scene, optics) → Human Vision System (Eye, visual cortex)

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Computer Vision (Semantics understanding)
Conventional Digital Camera

Today’s Digital Camera

Today's Digital Camera: Portrait Mode
Digital Camera System
Digital Camera System

Physical Scene
(Objects, lights)
Digital Camera System

Physical Scene
(Objects, lights)

Human Vision System
(Eye, visual cortex)
Digital Camera System

Physical Scene
(Objects, lights)

Human Vision System
(Eye, visual cortex)
Digital Camera System

Physical Scene
(Objects, lights)

Display
(Generating lights)

↓

Human Vision System
(Eye, visual cortex)
Digital Camera System

Physical Scene
(Objects, lights)

Display
(Generating lights)

Human Vision System
(Eye, visual cortex)

Electrical to optical signal transformation
Digital Camera System

Physical Scene
(Objects, lights) → Converting optical signals (light) to electrical signals (pixels)

Display
(Generating lights) → Human Vision System
(Eye, visual cortex)

Electrical to optical signal transformation
Digital Camera System

Physical Scene
(Objects, lights)

Optics
(Gather light)

Display
(Generating light)

Human vision system
(Eye, visual cortex)
Role of Optics in a Camera

- Focus light and capture energy
- At a cost of introducing “imperfections”
- Distorting incident light signals to another form of light signals

Desired Distortion

- Some optical “imperfections” give you an artistic feeling.

https://funny.pho.to/vignetting/
Desired Distortion

- We even intentionally create lenses that do not mimic our eyes.

https://www.bhphotovideo.com/c/product/1500028-REG/venus_optics_ve428rfnt_laeva_4mm_f_2_8_fisheye.html
Digital Camera System

Physical Scene
(Objects, lights)

Introduce “imperfections”

Optics
(Gather light)

Display
(Generating lights)

Human Vision System
(Eye, visual cortex)
Digital Camera System

Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Display (Generating lights) → Human Vision System (Eye, visual cortex)

Introduce “imperfections”

Physical Scene (Objects, lights)
Image Sensing

- Converts optical signals (lights) to electrical signals (pixels)

Photoelectric effect
Convert Optical Signals to Electrical Signals

- Introduces “imperfections”, which, unlike optical distortions, are almost always unwanted.

Digital Camera System

Physical Scene
(Objects, lights)

Optics
(Gather light)

Image Sensing
(Optical to electrical signal transformation)

Display
(Generating lights)

Human Vision System
(Eye, visual cortex)
Digital Camera System

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Display (Generating lights) → Human Vision System (Eye, visual cortex)
Low-Level Image Signal Processing

- Reconstruct/recover the original information carried in the incident light from the “distorted” pixels.

Digital Camera System

Physical Scene (Objects, lights) → Introduce “imperfections”
Optics (Gather light) → Introduce “imperfections”
Image Sensing (Optical to electrical signal transformation) → Correct “imperfections”
Image Signal Processing (Signal reconstruction)

https://4mobiles.net/iphone-6s-iphone-6s-plus/iphone-6s-space-grey/
Digital Camera System

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction)

Introduce “imperfections”
Introduce “imperfections”
Correct “imperfections”

https://4mobiles.net/iphone-6s-iphone-6s-plus/iphone-6s-space-grey/
**Digital Camera System**

<table>
<thead>
<tr>
<th>Physical Scene (Objects, lights)</th>
<th>Optics (Gather light)</th>
<th>Image Sensing (Optical to electrical signal transformation)</th>
<th>Image Signal Processing (Signal reconstruction)</th>
</tr>
</thead>
</table>

Conventional cameras use complex and expensive lenses and sensors to minimize imperfections.

[https://4mobiles.net/iphone-6s-iphone-6s-plus/iphone-6s-space-grey/](https://4mobiles.net/iphone-6s-iphone-6s-plus/iphone-6s-space-grey/)
Digital Camera System

Conventional cameras use complex and expensive lenses and sensors to minimize imperfections.
### Digital Camera System

<table>
<thead>
<tr>
<th>Physical Scene</th>
<th>Introduce “imperfections”</th>
<th>Introduce “imperfections”</th>
<th>Correct “imperfections”</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Objects, lights)</td>
<td>Optics (Gather light)</td>
<td>Image Sensing (Optical to electrical signal transformation)</td>
<td>Image Signal Processing (Signal reconstruction)</td>
</tr>
</tbody>
</table>

Conventional cameras use complex and expensive lenses and sensors to minimize imperfections.

Today’s cameras use better signal processing algorithms to make up for the weak optics and sensors.

[iPhone 6s](https://4mobiles.net/iphone-6s-iphone-6s-plus/iphone-6s-space-grey/)
Computational Photography

Use image signal processing algorithms to:

1. simplify optics and sensors.

- Portrait mode: simulate a large aperture
- HDR mode: simulate a high dynamic range sensor
Computational Photography

- Use image signal processing algorithms to:
  - (2) do things that conventional optics/sensors alone can’t do. Usually requires optics, sensor, algorithm co-design.

Digital Camera System

Physical Scene
(Objects, lights)

Optics
(Gather light)

Image Sensing
(Optical to electrical signal transformation)

Image Signal Processing
(Signal reconstruction)

Display
(Generating lights)

Human Vision System
(Eye, visual cortex)
Photorealistic Rendering

https://notrianglestudio.com/all-categories-blog/why-photorealistic-renderings-should-be-part-of-your-real-estate-marketing-strategy
Photorealistic Rendering

https://www.youtube.com/watch?v=aiY4cE_ngaY
Real-Time Photorealistic Rendering (e.g., Gaming)

https://www.digitaltrends.com/gaming/battlefield-v-dx12-ray-tracing-tested/
Graphics Rendering

- Why do those images look “real” to you?
Why do those images look “real” to you?
Photorealism: a rendered image that looks like taken by a camera
Why do those images look “real” to you?
Photorealism: a rendered image that looks like taken by a camera
Why do those images look “real” to you?
Photorealism: a rendered image that looks like taken by a camera
Simulating the camera pipeline (assuming perfect optical to electrical signal transformation)
Could simulate some desired optical distortions
Graphics Rendering

Physical Scene
(Objects, lights)

Optics
(Gather light)

Image Sensing
(Optical to electrical signal transformation)

Image Signal Processing
(Signal reconstruction)

Display
(Generating lights)

Human Vision System
(Eye, visual cortex)

Simulate object behaviors, light-object interactions, light-lens interactions
Graphics Rendering

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Computer Vision (Semantics understanding)

Modeling (Scene, optics) → Rendering (Visibility + Shading) → Display (Generating light) → Human vision system (Eye, visual cortex)
3D Modeling

Parametric surface

\[ x^2 + y^2 + z^2 = R^2 \]

Polygon mesh

Point cloud

https://twitter.com/dmitriatsafe/status/1066037693174341632
https://commons.wikimedia.org/wiki/File:Sphere_-_monochrome_simple.svg
Modeling: Mesh

http://www.cgarena.com/freestuff/tutorials/max/thomas_highway/sergeant.html
Modeling: Parametric Surface

\[ p(u, v) = \sum_{i=0}^{n} \sum_{j=0}^{m} B_i^n(u) B_j^m(v) \mathbf{k}_{i,j} \]

evaluated over the unit square, where

\[ B_i^n(u) = \binom{n}{i} u^i (1 - u)^{n-i} \]

is a Bernstein polynomial, and

\[ \binom{n}{i} = \frac{n!}{i!(n-i)!} \]

is the binomial coefficient.

Utah teapot

1975 Newell's drawing

Modeling: Point Cloud

Modeling: Point Cloud

From Mike Jarvis, Prof. @ History Dept. Scanning Elmina Castle, a UNESCO World Heritage Site at Ghana.
Modeling: Point Cloud

Velodyne LiDAR (Light Detection and Ranging)

Uber’s self-driving car equipped with LiDARs

https://iftt.de/Main/VelodyneHDL64E?setlang=en
https://news.voyage.auto/an-introduction-to-lidar-the-key-self-driving-car-sensor-a7e405590c?ig=78c92169e8e
https://news.voyage.auto/an-introduction-to-lidar-the-key-self-driving-car-sensor-a7e405590c?ig=78c92169e8e
Modeling Optics

Geometrical optics?
Wave optics?
Other abstractions?

Graphics Rendering

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Display (Generating lights) → Human Vision System (Eye, visual cortex) → Modeling (Scene, optics)
Graphics Rendering

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Display (Generating lights) → Human Vision System (Eye, visual cortex)

Modeling (Scene, optics) → Rendering (Visibility + Shading)
Simulate Light Transport

Direct Lighting Only

Direct + Indirect Lighting

Physics Simulation

https://www.flickr.com/photos/65945817@N07/6333871893
Graphics Rendering: Not All About Photorealism

Physical Scene
(Objects, lights)

Optics
(Gather light)

Image Sensing
(Optical to electrical signal transformation)

Image Signal Processing
(Signal reconstruction)

Display
(Generating lights)

Human Vision System
(Eye, visual cortex)

Modeling
(Scene, optics)

Rendering
(Visibility + Shading)
Computer Vision: Semantics Understanding

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Computer Vision (Semantics understanding)

Modeling (Scene, optics) → Rendering (Visibility + Shading) → Display (Generating lights) → Human Vision System (Eye, visual cortex)
Computer Vision: Semantics Understanding

Physical Scene  (Objects, lights)  ➔  Optics  (Gather light)  ➔  Image Sensing (Optical to electrical signal transformation)  ➔  Image Signal Processing (Signal reconstruction)  ➔  Computer Vision (Semantics understanding)

Modeling  (Scene, optics)  ➔  Rendering  (Visibility + Shading)

Display  (Generating lights)  ➔  Human Vision System  (Eye, visual cortex)

Consumer: human
Computer Vision: Semantics Understanding

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction)

Modeling (Scene, optics) → Rendering (Visibility + Shading)

Display (Generating lights)

Human Vision System (Eye, visual cortex)

Consumer: human

Computer Vision (Semantics understanding)

Consumer: computer!
Object Detection and Tracking
Segmentation
Video/Image Compression

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Computer Vision (Semantics understanding)

- Physical Scene: Objects, lights
- Optics: Gather light
- Image Sensing: Optical to electrical signal transformation
- Image Signal Processing: Signal reconstruction
- Computer Vision: Semantics understanding

- Modeling: Scene, optics
- Computer Graphics: Visibility + Shading
- Display: Generating lights
- Human Vision System: Eye, visual cortex
- Cloud/Storage: Video/Image De/Compression
Video Compression

Numbers credit: Kayvon Fatahalian
30-second video @ 1080p resolution (1920 x 1080 pixels per frame) @ 30 frames per second (FPS)
3 colors per pixel + 1 byte per color → 6.2 MB/frame → 6.2 MB x 30 s x 30 FPS = 5.2 GB total size
Actual H.264 video file size: 65.4 MB (80-to-1 compression ratio).
Compression/encoding done in real-time without you even realizing it!
Video Compression

Getting Computer Systems Ready for Visual Computing in Ten Years

Yuhan Zhu
https://cs.rit.edu
https://petlab.cs.rit.edu/research

41m 27s  3.81 GB est.

Add to Theater

fastpath_2020

Description: This video is about fastpath_2020
Tags: iMovie
Format: Video and Audio
Resolution: 720p
Quality: High
Compress: Faster

Cancel Next...
Video Compression

Getting Computer Systems Ready for Visual Computing in Ten Years

Yuhao Zhu
http://www.cs.utoronto.ca/
https://github.com/yuhaozhu

41m 27s  3.81 GB est.

Add to Theater

---

fastpath_2020

Description: This video is about fastpath_2020
Tags: iMovie
Format: Video and Audio
Resolution: 720p
Quality: High
Compress: Faster

---

fastpath_2020.mp4
Modified: August 19, 2020 at 9:12 AM

264.1 MB

Add Tags...

General:
Kind: MPEG-4 movie
Size: 264,068,677 bytes (268.4 MB on disk)
Video Compression

WiFi bandwidth test at my home
Video Compression

WiFi bandwidth test at my home

My home

Viewer

Getting Computer Systems Ready for Visual Computing in Ten Years

Yuhao Zhu

Video and Audio

Resolution: 720p
Quality: High
Compress: Faster

Fastpath_2020.mp4

Modified: August 19, 2020 at 9:12 AM

Add Tags...

General:
Kind: MPEG-4 movie
Size: 264,068,677 bytes (268.4 MB on disk)

Fastpath_2020

Description: This video is about fastpath_2020
Tags: iMovie

3.81 GB
11.68 Mb/s
≈ 45 min

264.1 MB
11.68 Mb/s
≈ 3 min

WiFi bandwidth test at my home

Add to Theater

41m 27s
3.81 GB est.
Scope of the Course

Physical Scene (Objects, lights) → Optics (Gather light) → Image Sensing (Optical to electrical signal transformation) → Image Signal Processing (Signal reconstruction) → Computer Vision (Semantics understanding) → Cloud/Storage

Modeling (Scene, optics) → Computer Graphics (Visibility + Shading) → Display (Generating lights) → Human Vision System (Eye, visual cortex) → Video/Image De/Compression

Electrical to optical signal transformation
Display: Transform Electrical Signals to Optical

Holographic Display

Sequence filmed with 4DViews’ volumetric capture technology

Holographic Display

Holographic Display

Holographic Display

Light-field Display: One Pathway Toward Holography

[Image of light-field display prototype]

Human Vision System

- How do we perceive light?
- How do we perceive color?
- Why do we have a sense of depth?
- Why do we have a sense of motion?

https://askabiologist.asu.edu/rods-and-cones
Why Study Visual Computing?

The most exciting applications on the horizon are visual computing applications, and they challenge the way we build computers.
Augmented Reality

Optics and Sensing:
sense the environment
**Computer vision:** recognizing table and estimate its orientation

**Optics and Sensing:** sense the environment
**Computer vision:** recognizing table and estimate its orientation

**Computer graphics:** drawing virtual content to overlay

**Optics and Sensing:** sense the environment


Computer vision: recognizing table and estimate its orientation

Computer graphics: drawing virtual content to overlay

Optics and Sensing: sense the environment

Optics and Display: mirrors and/or waveguides
360° (VR/Panoramic) Videos and Photos

VR Video Capturing Device

The stitching algorithm must handle a large amount of data. Input is sixteen $2704 \times 2028$ 30 FPS video streams and the output is a single $8192 \times 8192$ video stream, with a $8192 \times 4096$ panorama for each eye (with the left eye on top of the right). The algorithm takes about 60 seconds per frame (where each frame consists of 16 images) on a single machine, meaning an hour of video would take 75 days to process. To allow for timely processing we use a

Facebook Surround 360

https://www.wired.com/2016/04/heres-360-degree-video-camera-facebook-giving-away/
https://twitter.com/vtechnology/status/918116670559336448

Google Jump VR
Autonomous Machines

Autonomous Machines

https://www.wired.com/2017/05/the-physics-of-drones/
Why Study Visual Computing?

Humanity needs visual computing.
Cultural Heritage Preservation

https://accademia.stanford.edu/mkh/
Cultural Heritage Preservation

https://accademia.stanford.edu/mih/
Cultural Heritage Preservation

https://www.imb.org/image/elmina-castle-10/
Cultural Heritage Preservation
Cultural Heritage Preservation

https://slickdeals.net/article/buying-guide/oculus-quest-review/
http://digitalelmina.org/
https://www.sciencemag.org/fulltext/full/369/1676/1231a
What Challenge?

What Challenge?

<table>
<thead>
<tr>
<th>Energy Budget</th>
<th>Smartphone</th>
<th>~ 4000 mAh</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR/(VR)</td>
<td>~ 500 mAh</td>
<td></td>
</tr>
</tbody>
</table>
What Challenge?

Energy Budget

Smartphone

AR/(VR)

~ 4000 mAh

~ 500 mAh

“Retina” resolution: ~60 pixels per degree (PPD)
What Challenge?

<table>
<thead>
<tr>
<th>Energy Budget</th>
<th>Smartphone</th>
<th>~ 4000 mAh</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR/(VR)</td>
<td>~ 500 mAh</td>
<td></td>
</tr>
</tbody>
</table>

“Retina” resolution: ~60 pixels per degree (PPD)
What Challenge?

**Energy Budget**
- Smartphone: ~ 4000 mAh
- AR/(VR): ~ 500 mAh

**Processing Requirement**
- Smartphone: ~ 1 MPixel @ 30 FPS
- VR/AR: ~ 220 MPixel @ 120 FPS
What is This Course About?

- Understand the fundamental principles and structures of the problems in visual computing.

- Design more efficient algorithms, software, and hardware systems for visual computing applications on the horizon.
Necessarily Inter-Disciplinary
Necessarily Inter-Disciplinary

- Your goal is **NOT** to understand everything in one semester (no one does).
Necessarily Inter-Disciplinary

- Your goal is **NOT** to understand everything in one semester (no one does).
- Your goal is to understand the basic **principles** so that you can learn things on your own when need to.
Rochester is a Great Place for Visual Computing Research

- Augmented/Virtual Reality Initiative
  - [https://www.rochester.edu/augmented-virtual-reality/](https://www.rochester.edu/augmented-virtual-reality/)

- AR/VR National Research Traineeship (NRT) Program:
  - [https://www.rochester.edu/augmented-virtual-reality/education/index.html](https://www.rochester.edu/augmented-virtual-reality/education/index.html)

- Studio X:
  - [https://www.rochester.edu/advancement/studiox/](https://www.rochester.edu/advancement/studiox/)

- Center for Visual Science
  - [https://www.cvs.rochester.edu/](https://www.cvs.rochester.edu/)

- Collaboration between CS/ECE/BCS/Neuroscience/Optics/URMC
Logistics

- Course website: https://www.cs.rochester.edu/courses/572/fall2020/
- We “meet” on Wednesdays and Fridays (11:50 — 13:05).
- Synchronous delivery with videos for asynchronous consumption.
- All announcements will be made through blackboard.
- Feel free to send me emails.
- Office hours: Friday 2-3pm. Zoom link different from the lecture zoom. Sent in Blackboard. Come and discuss anything you want. There will be waiting room so that we can talk privately.
“Pre-requisites”

- Basic geometry (similar triangles, etc.)
- Basic physics (Power = Energy/Time, etc.)
- Basic calculus (integration, etc.)
- Knowing basic linear algebra (e.g., matrix multiplication, solving systems of linear equations) will be very helpful.
- Knowing basic computer systems (CSC 252 equivalent) could be helpful depending on what you want to get out of this course.
Reference Materials

- There is no definitive textbook. The course material is self-contained.
- The syllabus page has links to reference books: https://www.cs.rochester.edu/courses/572/fall2020/syllabus.html. Most of them are either PDFs online made available by the authors or are accessible through UR digital library (login required).
- There will be readings assigned (required and recommended).
How You Will Be Evaluated

‣ **3%**: finish the student information sheet and come to my office hours within the first 2 weeks to introduce yourself.

‣ **7%**: news flash. Take turns to present a 3-min news flash related to visual computing. Sign-up sheet to be posted soon.

‣ **30%**: written assignments; mostly every 2—3 weeks.

‣ **20%**: mini programming assignments; probably 2.

‣ **40%**: a final project of your choice that fits your interests and background; could be programming, systems, art, design, etc.
  ▶ Need approval from me. Must finalize within the first 2 months. Could be team of up to 2.
  ▶ UGs: think of it as an independent study.