Deep Learning for Computer Vision

Course Introduction

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Outline

What is Computer Vision?

Where is the gluestick? Find the book - what's its full title?

Credit: Bharath Kishore, Flickr CC License

What is wrong with this image?

Credit: Erik Johansson

What is Computer Vision?

Where is the gluestick? Find the book - what's its full title?

Credit: Bharath Kishore, Flickr CC License

What is wrong with this image?

Credit: Erik Johansson

Can a machine answer the above questions?

What is Computer Vision?

Computer Vision

- A field that seeks to automate and endow a computing framework with the ability to interpret images the way humans do.
- A sub-topic of Artificial Intelligence.

What is Computer Vision?

Computer Vision

- A field that seeks to automate and endow a computing framework with the ability to interpret images the way humans do.
- A sub-topic of Artificial Intelligence.

Other Definitions

- "the construction of explicit, meaningful descriptions of physical objects from images" (Ballard & Brown, 1982)
- "computing properties of the 3D world from one or more digital images" (Trucco & Verri, 1998)
- "to make useful decisions about real physical objects and scenes based on sensed images" (Sockman & Shapiro, 2001)

Why? Applications of Computer Vision

Autonomous Vehicles

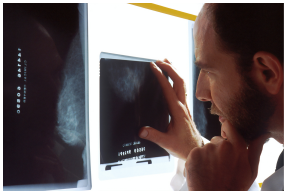
Credit: smoothgrover22, Flickr CC License

Surveillance

Credit: Yeong Nam, Flickr CC License

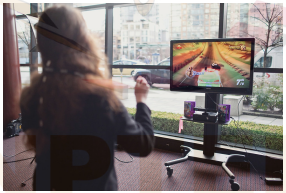
Factory Automation

Credit: KUKA Roboter GmbH, Bachmann



Medical Imaging

Credit: National Cancer Institute



Human-Computer Interaction

Credit: Vancouver Film School

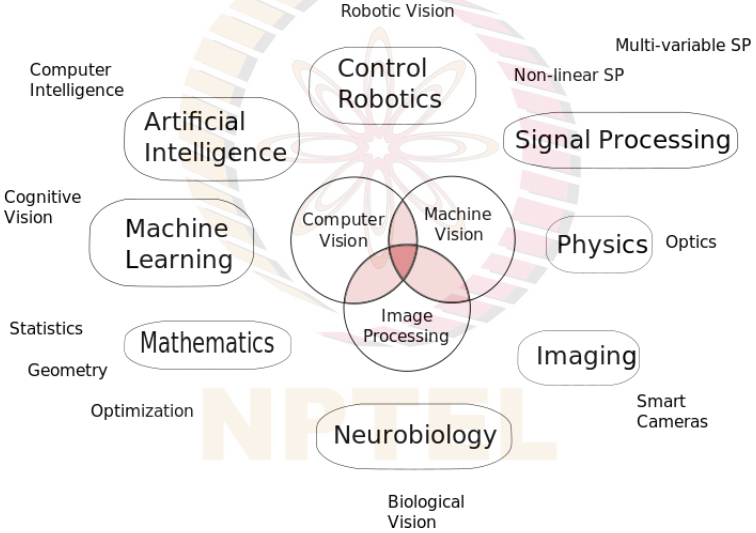
Visual Effects

Credit: AntMan3001, Flickr CC License

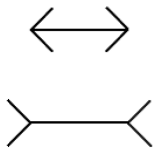
Applications of Computer Vision: More...

- **Retail and Retail Security** (Amazon Go, Virtual Try-on, StopLift)
- **Healthcare** (Blood Loss Detector, DermLens)
- **Agriculture** (SlantRange, Cainthus - Livestock facial recognition)
- **Banking and Finance** (Mobile Deposit, Insurance Risk Profiling)
- **Remote Sensing** (Land Use Understanding, Forestry Modeling)
- **Structural Health Monitoring** (Oilwell Inspection, Drone-based Bridge Inspection and 3D Reconstruction)
- **Document Understanding** (Optical Character Recognition, Robotic Process Automation)
- **Tele- and Social Media** (Image Understanding, Brand Exposure Analytics)
- **Augmented Reality** (TechSee Visual Support, Warehouse and Enterprise Management)

Perspectives of Study



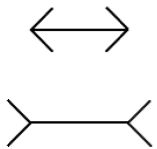
Why is it hard?[1]



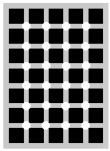
Müller-Lyer illusion: Which line is longer?

[1] Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

Why is it hard?[1]



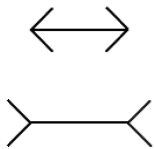
Müller-Lyer illusion: Which line is longer?



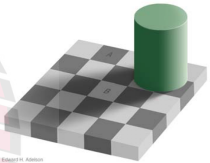
Variation of Hermann grid illusion: What do you see at the intersections?

[1] Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

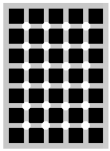
Why is it hard?[1]



Müller-Lyer illusion: Which line is longer?



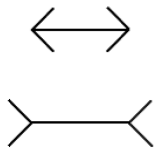
Adelson's brightness constancy illusion: Which is brighter, A or B?



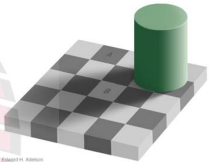
Variation of Hermann grid illusion: What do you see at the intersections?

[1] Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

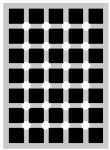
Why is it hard?[1]



Müller-Lyer illusion: Which line is longer?



Adelson's brightness constancy illusion: Which is brighter, A or B?

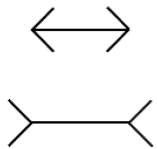


Variation of Hermann grid illusion: What do you see at the intersections?

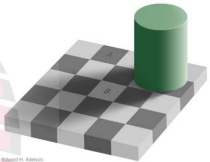
Count the red Xs in both figures, which is harder?

[1] Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

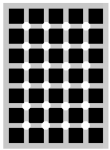
Why is it hard?[1]



Müller-Lyer illusion: Which line is longer?



Adelson's brightness constancy illusion: Which is brighter, A or B?



Variation of Hermann grid illusion: What do you see at the intersections?

Count the red Xs in both figures, which is harder?

[1] Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

Why is it hard?

- Many practical use cases are **inverse model** applications
  - No knowledge of how an image was taken or camera parameters - but need to model the real world in which picture/video was taken (shape, lighting, color, objects, interactions). $\Longrightarrow$ Need to almost always model from incomplete/partial noisy information
  - *Forward models* are used in physics (radiometry, optics, and sensor design) and in computer graphics

Why is it hard?

- Many practical use cases are **inverse model** applications
  - No knowledge of how an image was taken or camera parameters - but need to model the real world in which picture/video was taken (shape, lighting, color, objects, interactions). $\Longrightarrow$ Need to almost always model from incomplete/partial noisy information
  - *Forward models* are used in physics (radiometry, optics, and sensor design) and in computer graphics
- High-dimensional data $\Longrightarrow$ heavy computational requirements

Why is it hard?

- Many practical use cases are **inverse model** applications
  - No knowledge of how an image was taken or camera parameters - but need to model the real world in which picture/video was taken (shape, lighting, color, objects, interactions). $\Longrightarrow$ Need to almost always model from incomplete/partial noisy information
  - *Forward models* are used in physics (radiometry, optics, and sensor design) and in computer graphics
- High-dimensional data $\Longrightarrow$ heavy computational requirements
- Computer vision is AI-complete

Credit: Anish Chopra, Medium.com

Why is it hard?

- No complete models of the human visual system exist
  - Existing models largely related to subsystems, not holistic
  - What is perceived, and what is cognized? When is an object important for a task, and when is the context important?

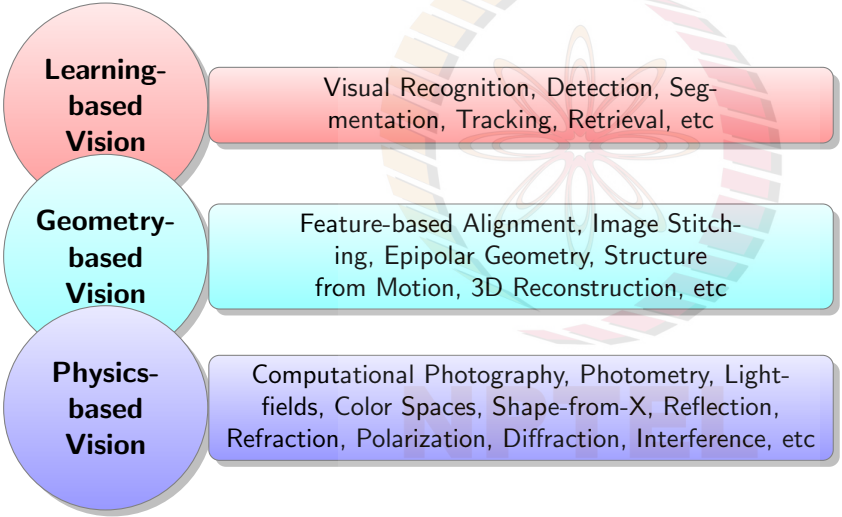
Why is it hard?

- No complete models of the human visual system exist
  - Existing models largely related to subsystems, not holistic
  - What is perceived, and what is cognized? When is an object important for a task, and when is the context important?
- Verifiability of mathematical/physical models non-trivial
  - How should similarity/dissimilarity between representations be defined? Is this a distance metric? Do all images follow such a distance metric?
  - How would a manipulation (counterfactual) in a given (potentially noisy) environment behave, w.r.t. the captured image/video? Can a physical model capture this?

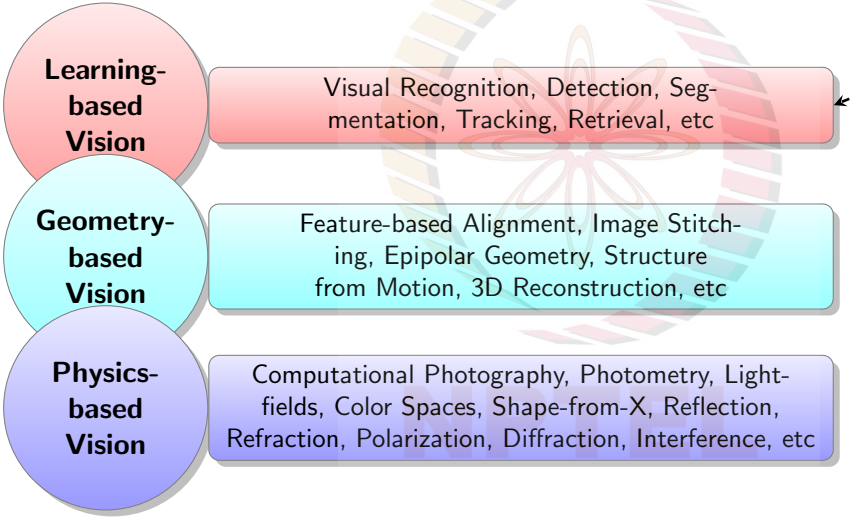
Credit: Anish Chopra, Medium.com

Outline

Computer Vision: Topics



Computer Vision: Topics



Course Topics

Segment 1: The Journey So Far

- Image Formation, Linear Filtering
- Edges, Blobs, Features
- Visual Descriptors, Matching

Course Topics

Segment 1: The Journey So Far

- Image Formation, Linear Filtering
- Edges, Blobs, Features
- Visual Descriptors, Matching

Segment 2: The Building Blocks

- Review of Neural Networks
- Convolutional Neural Networks (CNNs)
- CNN Architectures, Visualizing and Understanding CNNs

Course Topics

Segment 1: The Journey So Far

- Image Formation, Linear Filtering
- Edges, Blobs, Features
- Visual Descriptors, Matching

Segment 2: The Building Blocks

- Review of Neural Networks
- Convolutional Neural Networks (CNNs)
- CNN Architectures, Visualizing and Understanding CNNs

Segment 3: The Many Forms and Uses

- Recognition, Verification, Retrieval, Detection, Segmentation

Course Topics

Segment 1: The Journey So Far

- Image Formation, Linear Filtering
- Edges, Blobs, Features
- Visual Descriptors, Matching

Segment 2: The Building Blocks

- Review of Neural Networks
- Convolutional Neural Networks (CNNs)
- CNN Architectures, Visualizing and Understanding CNNs

Segment 3: The Many Forms and Uses

- Recognition, Verification, Retrieval, Detection, Segmentation

Segment 4: A Dimension Beyond

- Recurrent Neural Networks
- Spatio-Temporal Models
- Attention, Vision-Language Tasks

Course Topics

Segment 1: The Journey So Far

- Image Formation, Linear Filtering
- Edges, Blobs, Features
- Visual Descriptors, Matching

Segment 2: The Building Blocks

- Review of Neural Networks
- Convolutional Neural Networks (CNNs)
- CNN Architectures, Visualizing and Understanding CNNs

Segment 3: The Many Forms and Uses

- Recognition, Verification, Retrieval, Detection, Segmentation

Segment 4: A Dimension Beyond

- Recurrent Neural Networks
- Spatio-Temporal Models
- Attention, Vision-Language Tasks

Segment 5: Staying Contemporary

- Deep Generative Models
- Learning with Limited Supervision
- Recent Trends

Course Eligibility

Theory

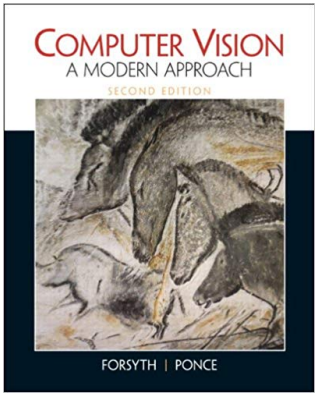
- Completion of a basic course in Machine Learning
- Completion of a basic course in Deep Learning highly recommended
- Knowledge of basics in probability, linear algebra, and calculus

Programming

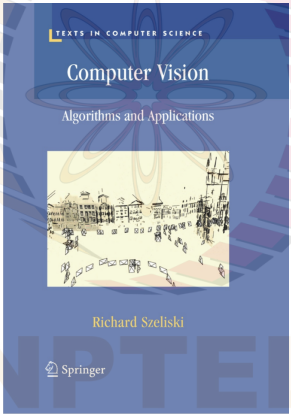
- Comfort with programming in Python
- Knowledge of a deep learning framework (PyTorch or TensorFlow) highly recommended

Outline

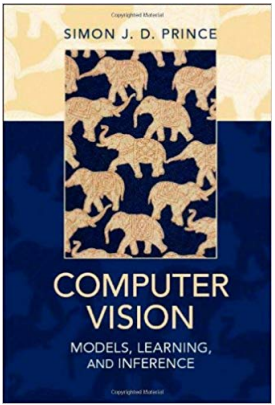
Traditional Computer Vision: References



Book website

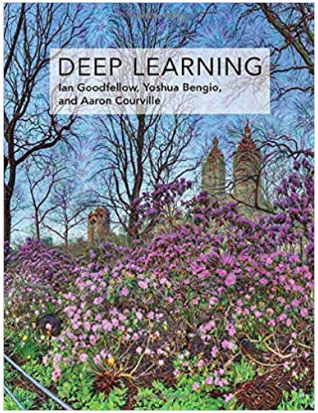


Book website

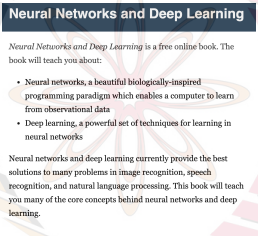


Book website

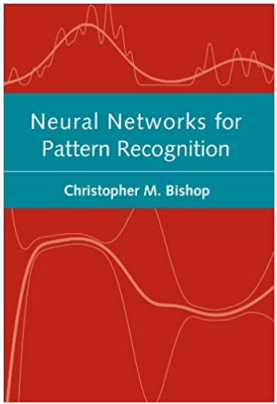
Deep Learning: References



Book website

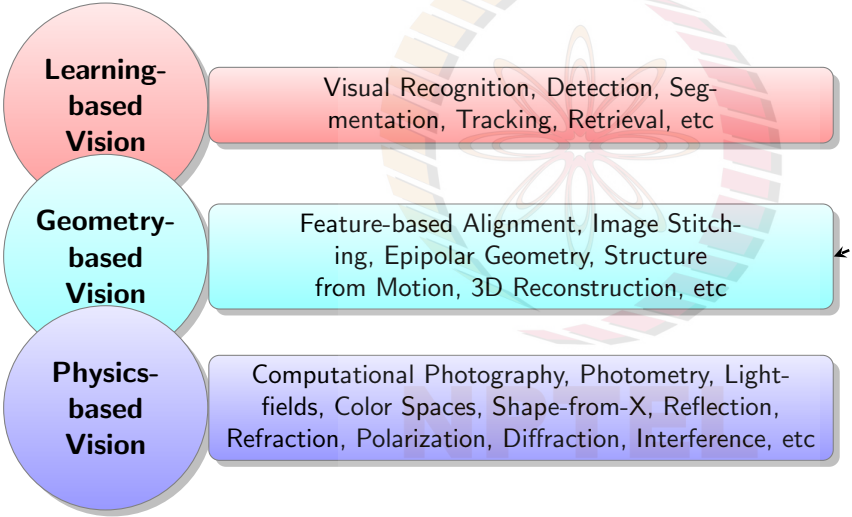


A nice, short online book by Michael Nielsen



Book website

Want to Learn Other Topics?

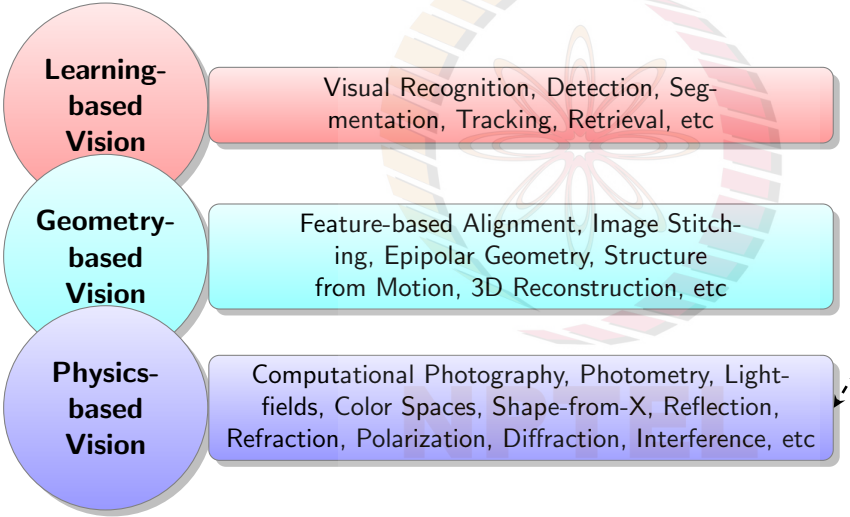


Book Link: Multiple View Geometry in Computer Vision

+

Relevant Course Links: Link 1, NPTEL Course, See Weeks 2-5

Want to Learn Other Topics?



Book Link: Physics-Based Vision: Principles and Practice
+
Relevant Course Links: Link 1

Homework!

- Go through all links on the Applications of Computer Vision slide (Slide 6) - they are interesting views/reads!

Acknowledgements and Disclaimers

- We are grateful to the deep learning/computer vision courses and their contents that are publicly available online. Wherever possible and relevant, these sources have been cited. If you notice an oversight, please let us know, and we will be glad to acknowledge.
- Any errors in the material are our own. Please point out such issues, and we will be glad to rectify.
- To the extent possible, all images used in these materials have been chosen from free stock photos to avoid any copyright violations. If you notice an oversight, please let us know.