

Deep Learning for Computer Vision

# Course Introduction

Vineeth N Balasubramanian

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



NPTEL

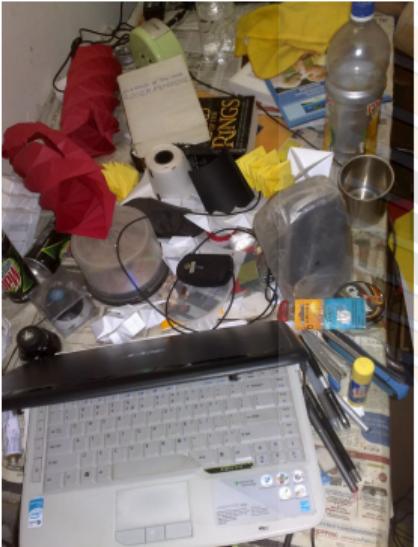
# Outline



- 1 Computer Vision: What and Why
- 2 This Course: Topics, Structure, Objectives
- 3 Resources and References

NPTEL

# What is Computer Vision?



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

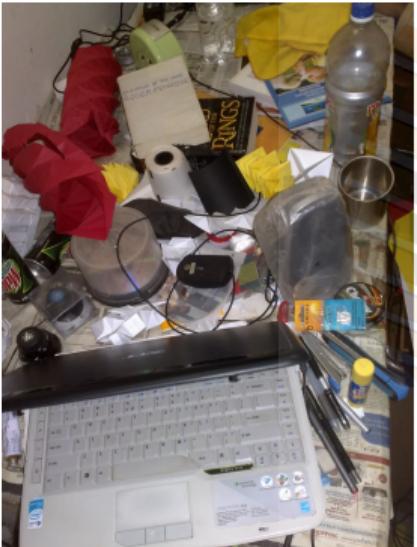
Credit: Bharath Kishore, Flickr CC License



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?

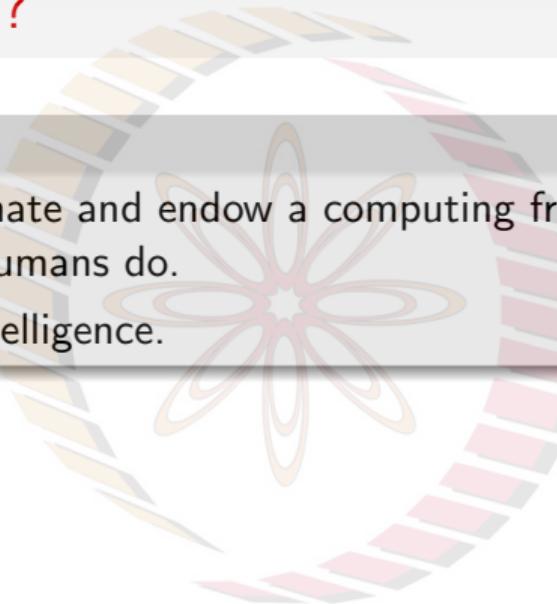
Credit: Erik Johansson

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.

A large, semi-transparent watermark of the NPTEL logo is centered on the slide. The logo consists of a stylized orange and red flower-like emblem with radiating lines, set against a background of concentric curved bands in shades of orange, yellow, and red.

NPTEL

# 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

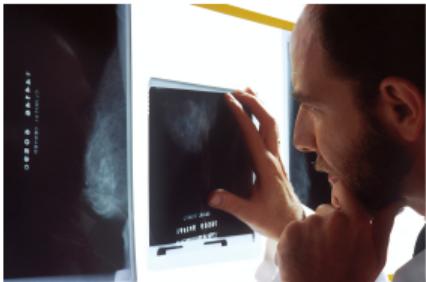
- “the construction of explicit, meaningful descriptions of physical objects from images” (Ballard & Brown, 1982)
- “computing properties of the 3D world from one or more digital images” (Trucco & Verri, 1998)
- “to make useful decisions about real physical objects and scenes based on sensed images”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

# Why? Applications of Computer Vision



## Autonomous Vehicles

Credit: smoothgrovers22, Flickr CC License



## Medical Imaging

Credit: National Cancer Institute

Vineeth N B (IIT-H)



## Surveillance

Credit: Yeong Nam, Flickr CC License



## Human-Computer Interaction

Credit: Vancouver Film School

§1.1 Introduction



## Factory Automation

Credit: KUKA Roboter GmbH, Bachmann



## Visual Effects

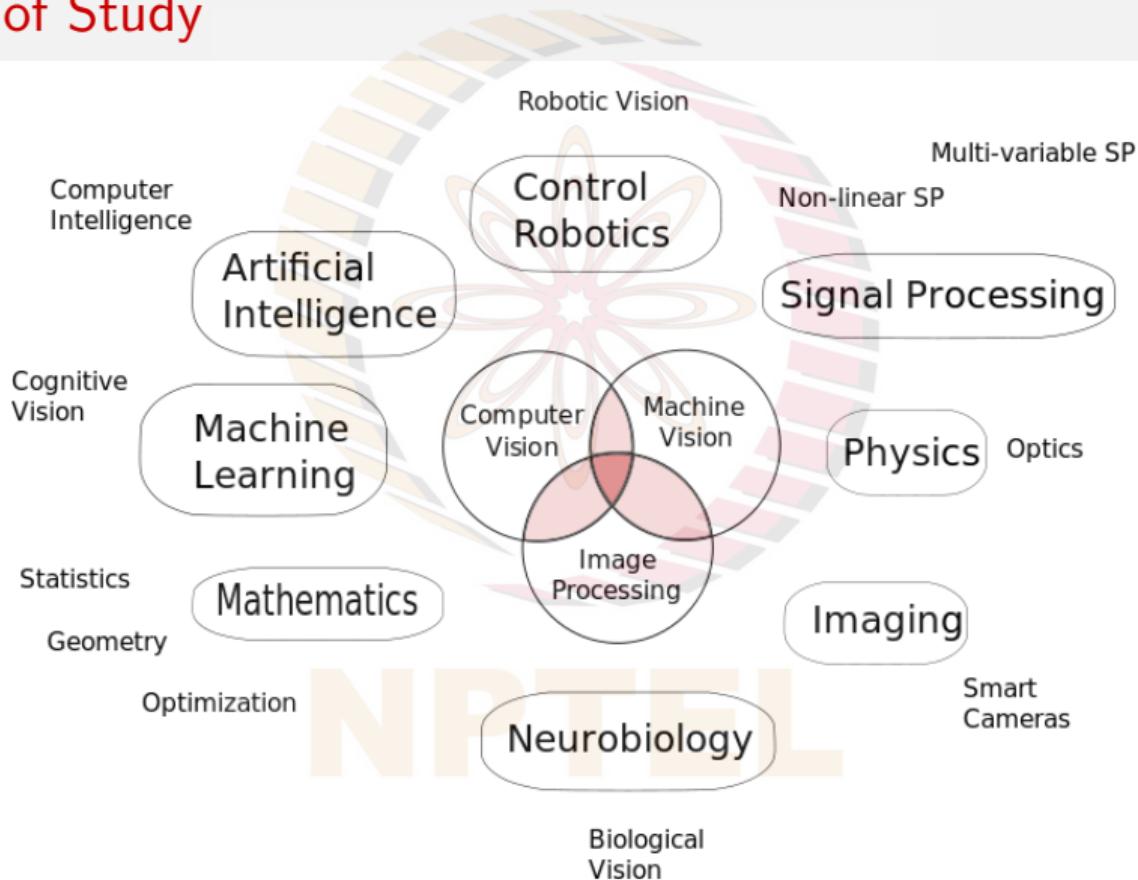
Credit: AntMan3001, Flickr CC License

# Applications of Computer Vision: More...

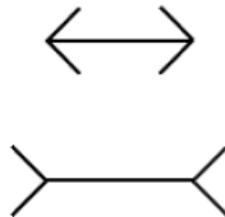
- **Retail and Retail Security** ([Amazon Go](#), [Virtual Try-on](#), [StopLift](#))
- **Healthcare** ([Blood Loss Detector](#), [DermLens](#))
- **Agriculture** ([SlantRange](#), [Cainthus - Livestock facial recognition](#))
- **Banking and Finance** ([Mobile Deposit](#), [Insurance Risk Profiling](#))
- **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](#))

NPTI

# Perspectives of Study



# Why is it hard?<sup>1</sup>

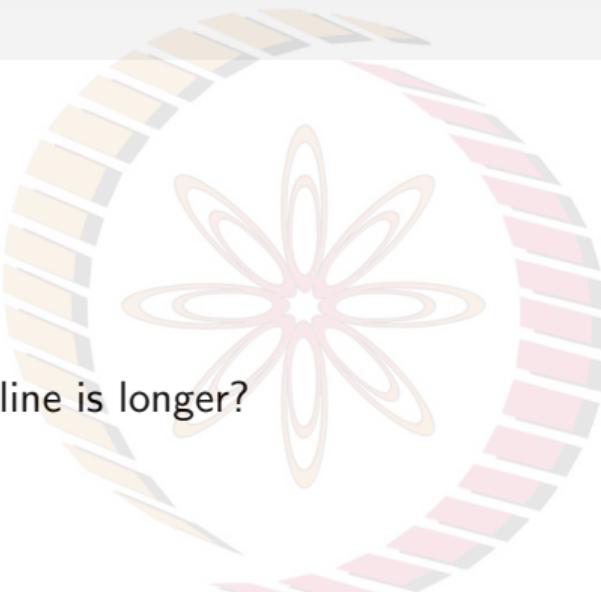
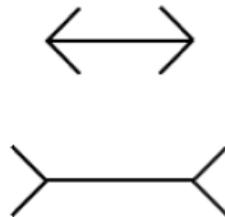


Müller-Lyer illusion: Which line is longer?

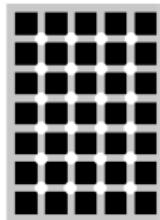
NPTEL

<sup>1</sup>Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

# Why is it hard?<sup>1</sup>



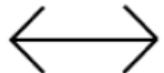
Müller-Lyer illusion: Which line is longer?



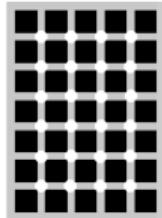
Variation of Hermann grid illusion: What do you see at the intersections?

<sup>1</sup>Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

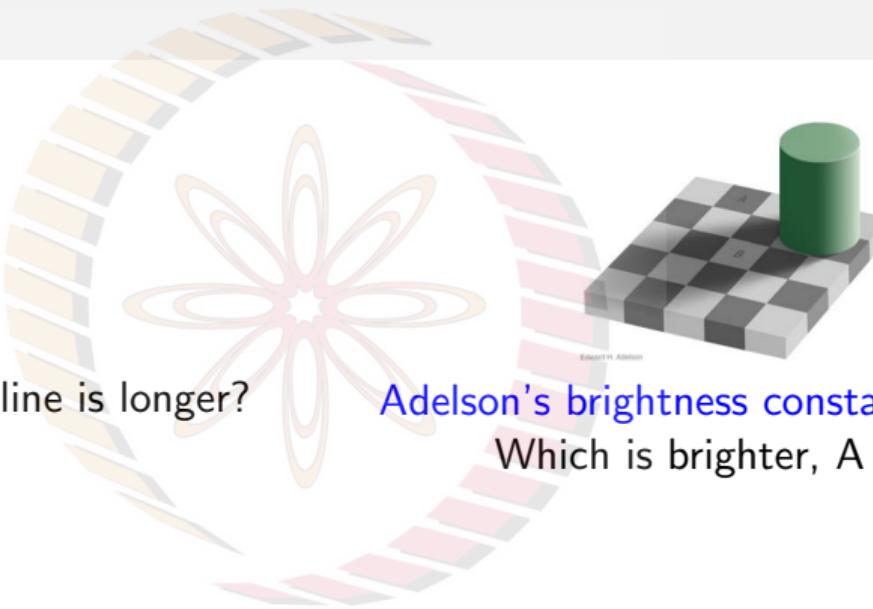
# Why is it hard?<sup>1</sup>



Müller-Lyer illusion: Which line is longer?



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



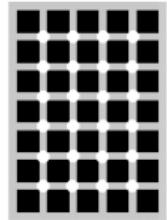
Adelson's brightness constancy illusion:  
Which is brighter, A or B?

<sup>1</sup>Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

# Why is it hard?<sup>1</sup>



Müller-Lyer illusion: Which line is longer?



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



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

x x x x x x x x	o x o x o x x
x x x x x x x x	x o x x x o x
x x x x x x x x	o x x o x x o
x x x x x x x x	x x o x o o x
x x x x x x x x	o x x o x x x
x x x x x x x x	x o x x x x o
x x x x x x x x	o x x o o x x
x x x x x x x x	x o x x x x o
x x x x x x x x	x x x o o x x
x x x x x x x x	x o x x x x o

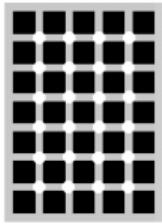
Count the red Xs in both figures, which is harder?

<sup>1</sup>Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

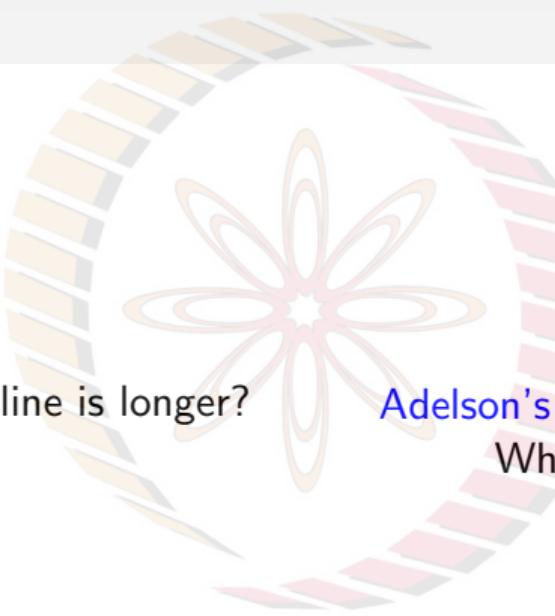
# Why is it hard?<sup>1</sup>



Müller-Lyer illusion: Which line is longer?



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



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

x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

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

<sup>1</sup>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).  $\Rightarrow$  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

The NPTEL logo is a large, stylized, blocky text "NPTEL" in a light beige color. Behind the letters, there is a circular graphic composed of several concentric rings and radial bars in shades of brown, tan, and light orange, resembling a stylized sun or a complex lens flare effect.

# 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).  $\Rightarrow$  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  $\Rightarrow$  heavy computational requirements

The NPTEL logo is a watermark-style graphic located at the bottom center of the slide. It consists of the letters "NPTEL" in a large, bold, sans-serif font. The letters are partially transparent, allowing the background content to be seen through them.

# 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).  $\Rightarrow$  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  $\Rightarrow$  heavy computational requirements
- Computer vision is **AI-complete**

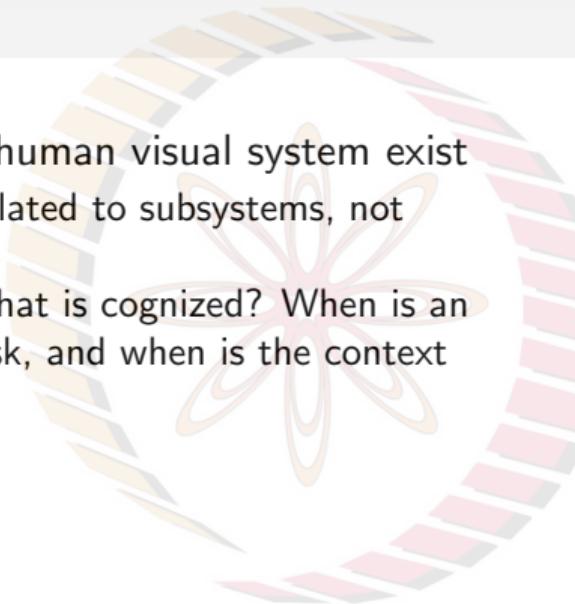


IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

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?

A circular watermark logo for NPTEL. It features a stylized flower or sunburst design in the center, composed of many thin, curved, overlapping lines in shades of orange, yellow, and red. This central design is surrounded by two concentric rings of rectangular blocks. The inner ring is light orange and the outer ring is light pink. Both rings have a slight gradient and some internal shading.

NPTEL

# 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?



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

Credit: Anish Chopra, Medium.com

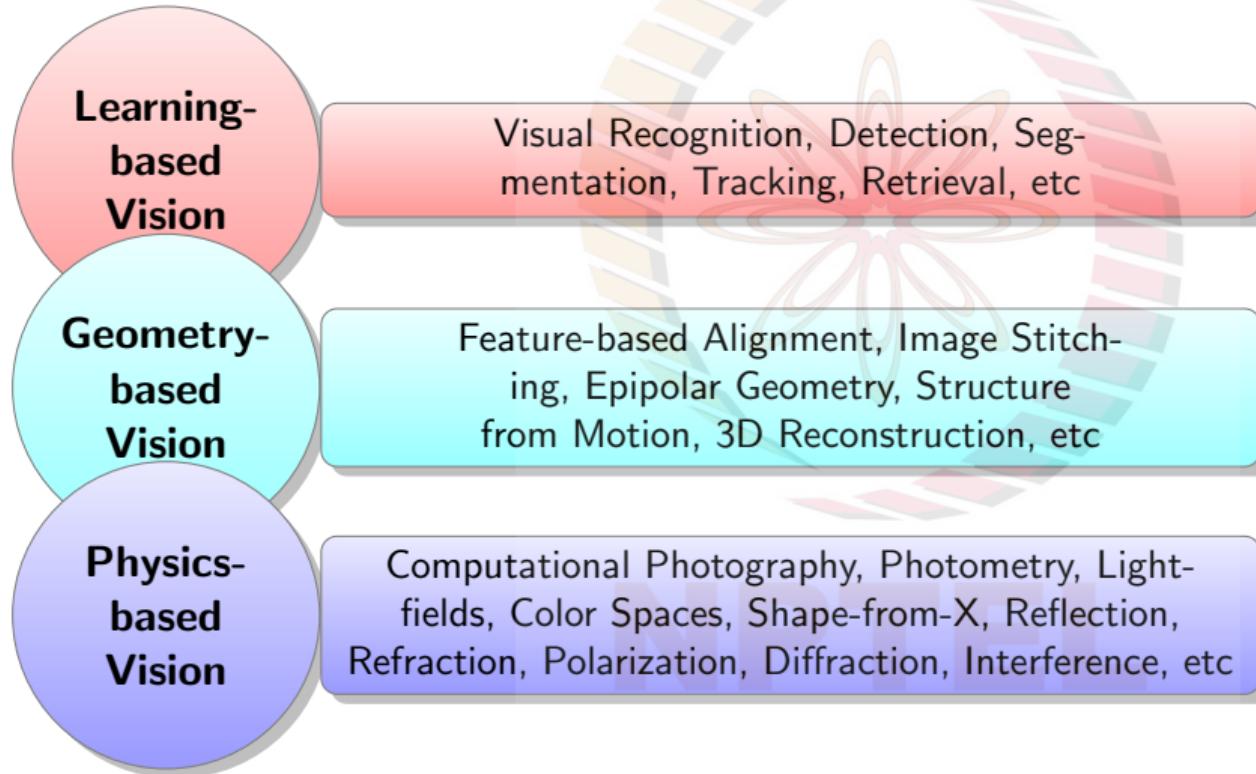
# Outline



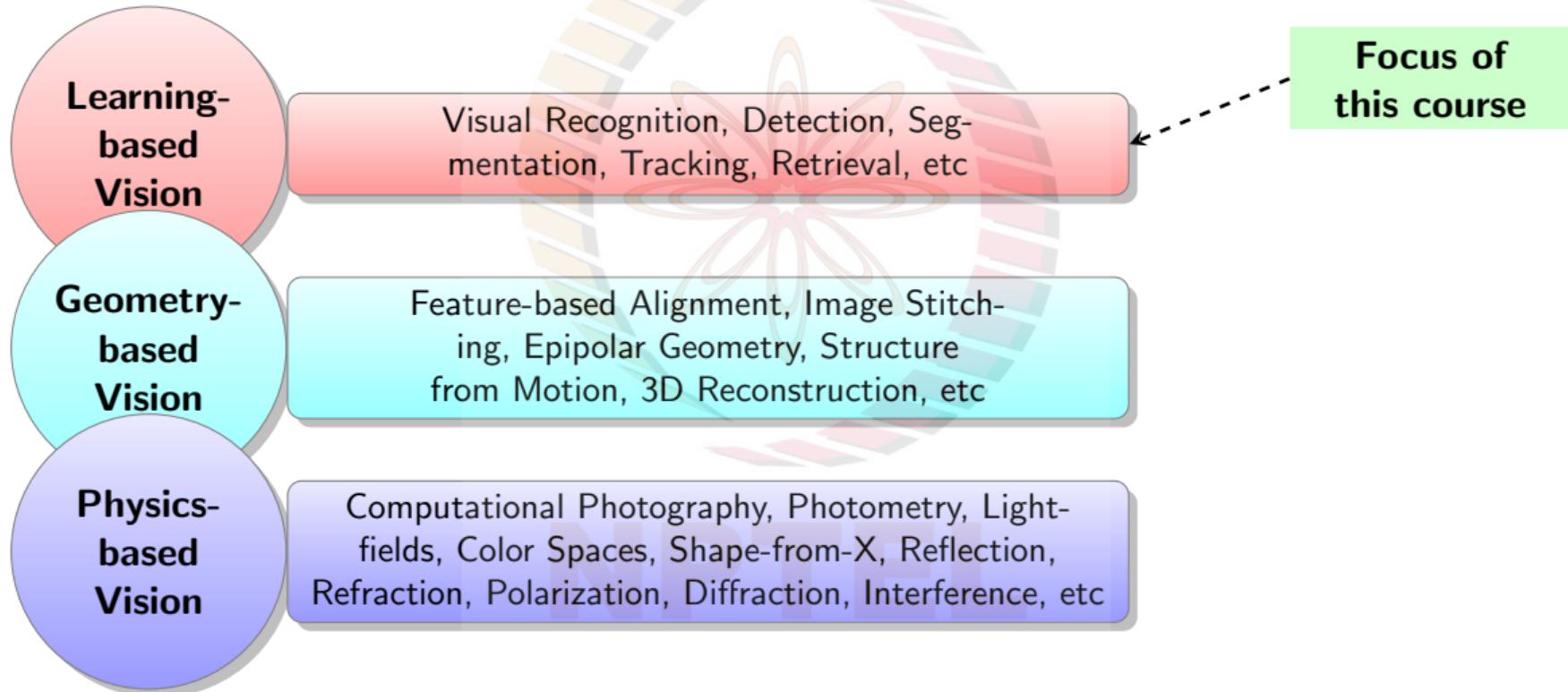
- 1 Computer Vision: What and Why
- 2 This Course: Topics, Structure, Objectives
- 3 Resources and References

**NPTEL**

# Computer Vision: Topics



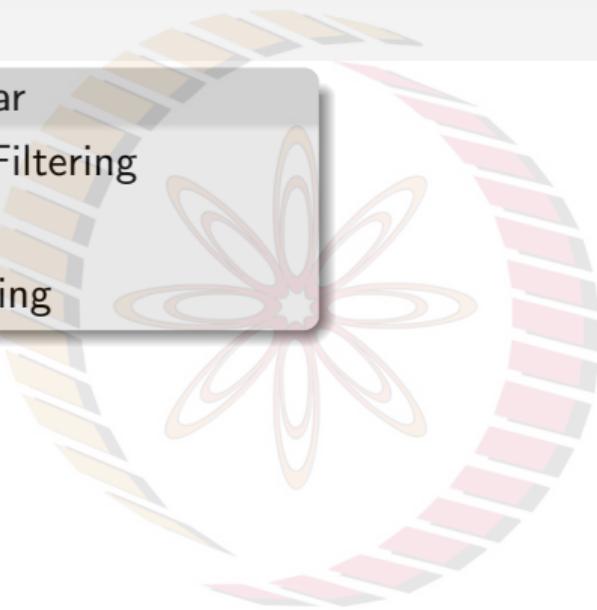
# Computer Vision: Topics



# Course Topics

## Segment 1: The Journey So Far

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



NPTEL

# 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

The NPTEL logo is displayed prominently in the center of the slide. It consists of the word "NPTEL" in a large, bold, sans-serif font. The letters are colored in a gradient that transitions from light blue at the top to light orange at the bottom. The letters are slightly overlapping, giving a sense of depth.

# 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



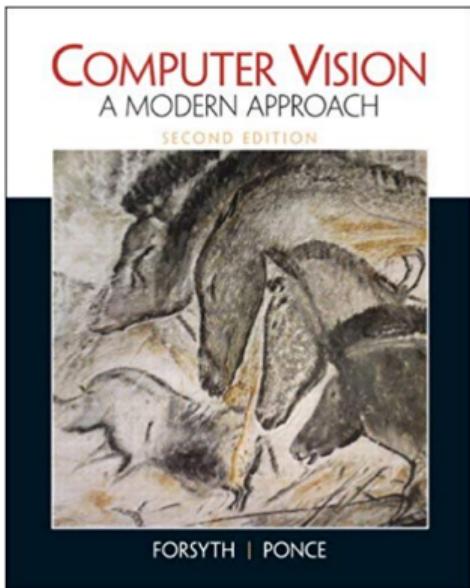
1 Computer Vision: What and Why

2 This Course: Topics, Structure, Objectives

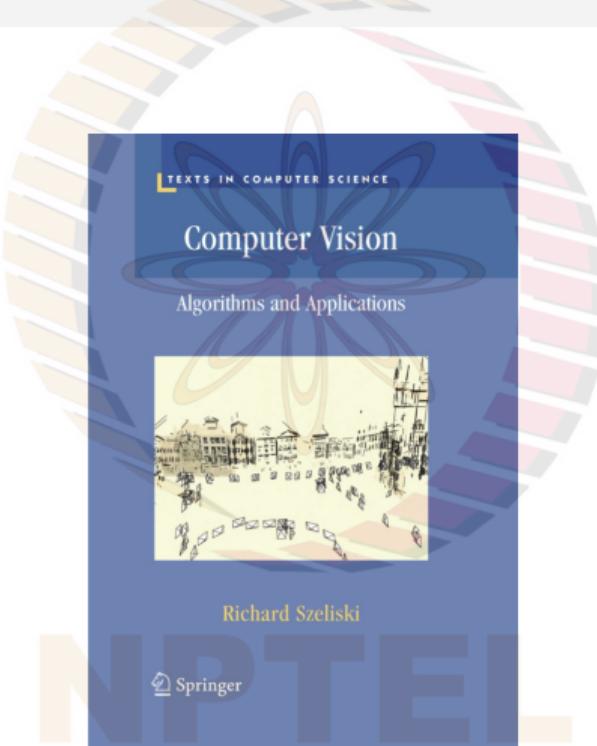
3 Resources and References

**NPTEL**

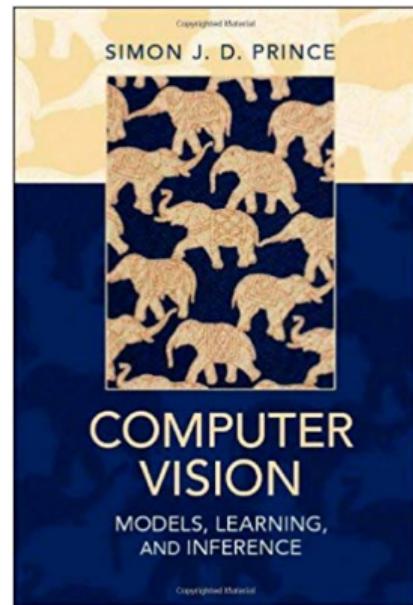
# Traditional Computer Vision: References



[Book website](#)

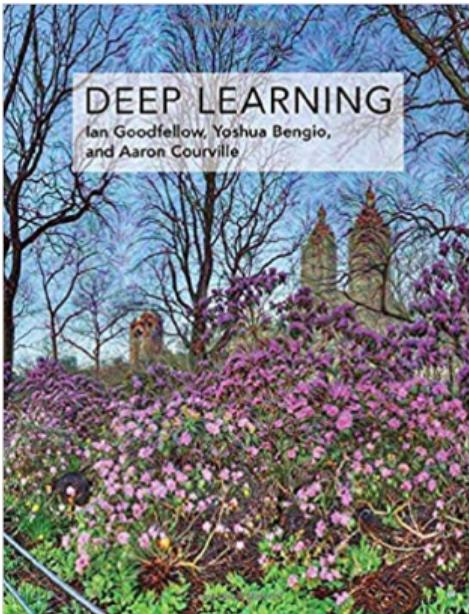


[Book website](#)



[Book website](#)

# Deep Learning: References



[Book website](#)

**Neural Networks and Deep Learning**

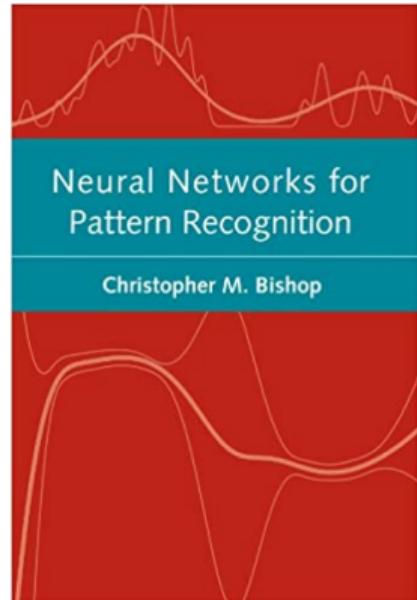
*Neural Networks and Deep Learning* is a free online book. The book will teach you about:

- Neural networks, a beautiful biologically-inspired programming paradigm which enables a computer to learn from observational data
- Deep learning, a powerful set of techniques for learning in neural networks

Neural networks and deep learning currently provide the best solutions to many problems in image recognition, speech recognition, and natural language processing. This book will teach you many of the core concepts behind neural networks and deep learning.

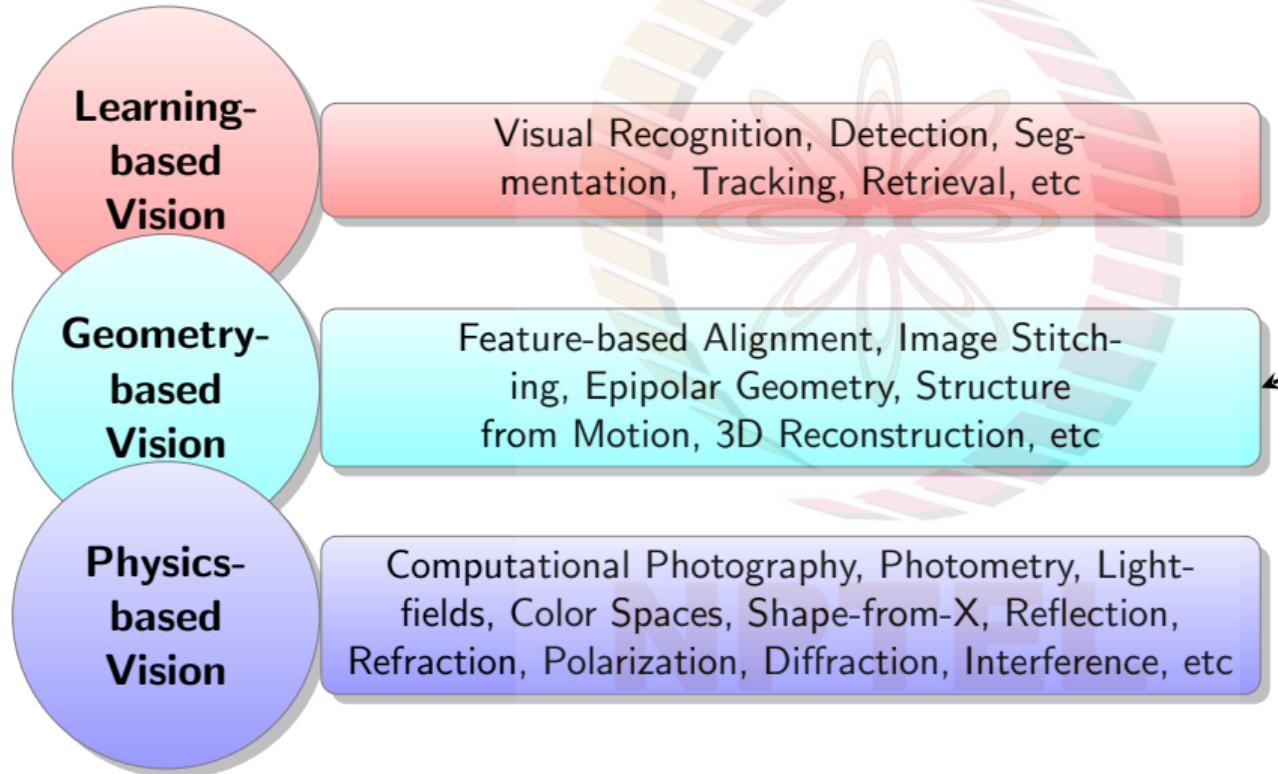
A nice, short online book  
by Michael Nielsen

**NPTEL**



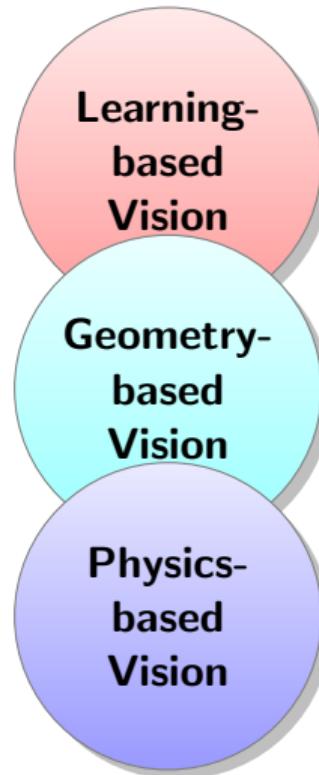
[Book website](#)

# 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!



# NPTEL

# 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.