

Deep Learning for Computer Vision
Professor. Vineeth N Balasubramanian
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad
Course Introduction

(Refer Slide Time: 0:15)

The screenshot shows a presentation slide with the following content:

- Top bar: Deep Learning for Computer Vision
- Section header: **Course Introduction**
- Speaker: Vineeth N Balasubramanian
- Department: Department of Computer Science and Engineering, Indian Institute of Technology, Hyderabad
- Logos: NPTEL and IIT Hyderabad
- Footer: Vineeth N B (IIT-H) | §1.1 Introduction | 1 / 21
- Video inset: A small video window showing Professor Vineeth N Balasubramanian.

Hello and welcome to this first lecture for this course Deep Learning for Computer Vision. My name is Vineeth Balasubramanian. I am a faculty in the Department of Computer Science and Engineering at IIT, Hyderabad. Hope all of you had a chance to look at the welcome video for this course which may have given you an introduction of what to expect. We will however revisit some of those things in this particular lecture and talk about what to expect and why we need to study this particular course.

(Refer Slide Time: 0:51)

Computer Vision: What and Why

Outline

1

Computer Vision: What and Why

2

This Course: Topics, Structure, Objectives

3

Resources and References



Vineeth N B (IIT-H)

1.1 Introduction

2 / 21

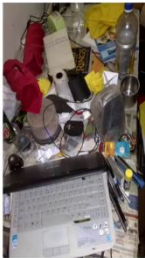


To start with, we will talk about what and why Computer Vision. Then we will talk about the topics that we will cover in this course, the structure that we will follow, what are objectives, what you will not learn and finally we will talk about the resources and references for this course in this first part of the lecture.


(Refer Slide Time: 1:11)

Computer Vision: What and Why

What is Computer Vision?





Where is the glue stick? 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?



Vineeth N B (IIT-H)

1.1 Introduction

3 / 21



Firstly, to start with what is computer vision? So, if you looked at this image of the left and ask the question where is the glue stick? Or can you find the book? What is a full title? You probably take a few seconds but if you observe closely, you will then find this glue stick here standing vertically and there is this book hidden behind other books but that is the book that



slightly to draw your attention and looking at some partial information I would say the books name is Lord of the rings.

The brain is really good at filling in information to get what is hidden in the image. And the image on the right, by looking at it if I ask you the question, what is wrong with this image? It seems very naturally made but from our knowledge of the world and our knowledge how the world behaves, we would probably say the somebody has stampered and created this image and this does not seem to meet some natural laws that we know of.

Now, what we want to find out in this course is, can a machine answer the same questions? Is it trivial for a machine to understand the clutter in an image and isolate one particular glue stick whose pose is very untypical where the name glue stick is not revealed on that particular object? Or can a computer be able to find out that from the word rings and from the font of that particular word, can you fill in and see the book is Lord of the rings.

And same here for the image on the right, can a computer have a knowledge of the world that can help it understand there is something wrong with this image? And be able to say what is wrong. These are very difficult problems for a computer and the human brain is very very good at filling in partial information and bringing in context and all of the knowledge that we have gained to be able to help us understand the world around us visually.

(Refer Slide Time: 3:21)



Computer Vision: What and Why

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" (Sackman & Shapiro, 2001)

Vineeth N.B. (IIT-H) 11.1 Introduction 4 / 21





So, what is Computer Vision? More formally speaking, you could say, Computer Vision is a field that seeks to automate and endow a computing framework with the ability to interpret


There could be other definitions that many researchers over the years have defined, you could call it “the construction of explicit meaningful descriptions of physical objects from images”. This was defined by Ballard and Brown or Trucco and Verri defined it as “computing properties of the 3D world around us from one or more, one or more digital images” or as Sockman and Shapiro say “to make useful decisions about real physical objects and scenes based on sensed images”.

(Refer Slide Time: 4:45)


Computer Vision: What and Why

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
[3.1.1 Introduction](#)



Visual Effects
Credit: Art3d3001, Flickr CC License

Vineeth N B (IIT-H)

5 / 21



Why do we need to study this field apart from the curiosity to understand the world, apart from the need to a computer do what humans do. What are the applications of Computer Vision in today's world? I am quite sure many of you may already know this, but Computer Vision is all around us today. Be it the face detection on your Facebook or Picasa or any other visual platform that you have, to autonomous vehicles where the perception in front of a car is through various cameras including cameras like Ladder cameras.

Surveillance such as CCTV in your Airports or railway stations or wherever it maybe. Factory Automation, this is one of the oldest applications of Computer Vision where cameras are fixed in various settings in manufacturing pipelines or warehouses to be able to automate various aspects of factory, factory cases, medical imaging, very important application today.

Human-Computer Interaction, some of you may have played the Kinect or the X-Box where there is a stereo camera looking at a user and tracking the user to say whether a user is playing, if the user is playing a bowling game or a tennis game, whether the movement is a particular forehand or backhand or any other movement for a particular game. And finally, visual effects and movies where computer vision is used extensively.

These are just some sample applications from different domains but you can imagine that there are many many kinds of applications where one can use cameras and computer vision on the feed that you get from the cameras to be able to get knowledge using in an automated manner.

(Refer Slide Time: 6:33)

Computer Vision: What and Why

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](#))

Vineth N B (IIT-H) 1.1 Introduction 6 / 21

For accessing this content for free (no charge), visit : nptel.ac.in

NPTEL
National Program on Technology Enhanced Learning
© 2014 NPTEL
For more information visit : nptel.ac.in

There are many many more applications. Here are a few more including some recent ones that I have tried to put together on this particular slide. In fact, one of your homework so this particular lecture is going to be to click on some of these links and see how computer vision is used there. And I can guarantee you that some of these are very very interesting applications which you may or may not have come across so far.

One of them is Retail and Retail Security. Some of you may have heard of Amazon Go an initiative of Amazon where the entire shopping experience is without a billing counters. So, you go in into a store, pick up whatever objects that you have in in on the different shelves and you simply walk out of the store with no billing. So, the retail store automatically tries to understand what objects you have picked and put into your basket and just deducts it from your Amazon Wallet or any other thing, any other Wallet that you configure for your shopping experience.

Similarly, this Virtual Try-on where you try on a piece of cloth before buying it online and try to see how it looks on you. StopLift to prevent shop lifting. Healthcare where you can use it for blood loss detection, for dermatology, in agriculture for trying to find out various aspects of agriculture, for trying to find out livestock face recognition, in banking and finance to be able to deposit your cheques automatically through a through a feeder. So, there is no human involved and trying to find out to whom that cheques should be credited and what is the amount all of that is automated through vision system.

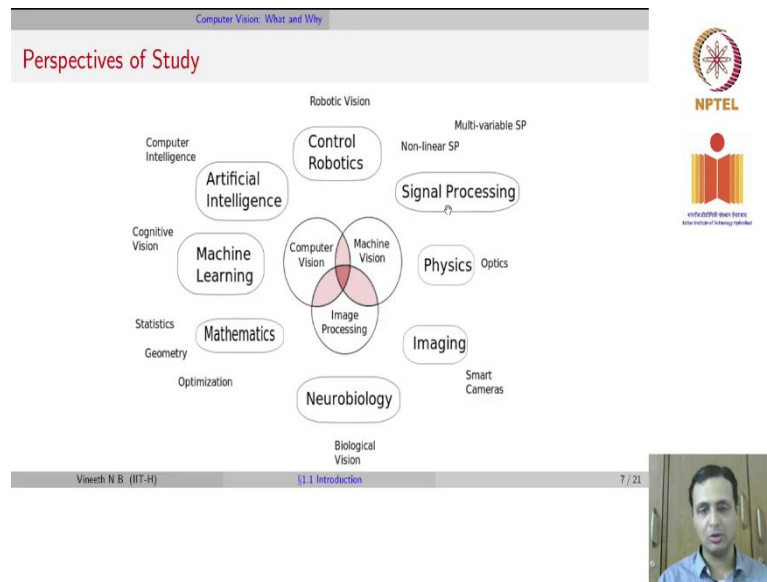
Insurance Risk, Insurance Risk Profiling, Remote Sensing where you do land use understanding using satellite imagery or forestry modelling using drone or satellite imagery. Structural health monitoring, a very important application today where you can do oil well inspection or drone based bridge inspection or 3D reconstructing or try to understand the health of railway tracks using drone imagery and so on and so forth.

This is extremely important in various settings where you may want to monitor the health of an infrastructure without having to physically inspect that infrastructure. For example, it may not be trivial to under a bridge and try to find out if there are cracks under the bridge, but you could use a drone to fly through that area and use the images that you get from that drone to be able to process and understand the health of bridges, railway tracks, oil refineries and any of those any of those large infrastructures.

Document understanding, optical character recognition is one of the oldest successes of computer vision, in fact one of the oldest products that computer vision has delivered to the world. Robotic process automation for document understanding, you can click on these links to understand what each of these terms mean. Tele and Social media, image understanding, brand exposure analytics.

Augmented reality to be able to do visual support or warehouse and enterprise management so on and so forth. In case any of these terms are not clear, please do click on these links and each of those links will take you to a video or a description of what that application is and how computer vision is used in that setting. Please do it to understand the various applications that computer vision has.

(Refer Slide Time: 9:56)



So, in terms of studying computer vision, very broadly speaking there are a few different terms that often overlap with Computer vision, Computer vision, Machine vision and Image processing. So, one of these questions you could ask here is, how are these different? Today it is popularly called computer vision. Image processing often refers to low level image processing. You try to process an image at a low level at a low level such as extracting edges or trying to invert an image or try to do various processing operations straight on that image.

But computer vision is more about understanding higher level abstractions or knowledge of the world from images. So, there is a difference in abstraction of understanding between image processing and computer vision. Although, these days I think the lines between them are very blurry and where the one ends and where the other begins is very difficult to say. We will obviously flow from one to the other in this particular course.

Regarding computer vision and machine vision, this is a little bit more trickier, this is more in terms of the legacy of how this was developed. Machine vision was one of the popular terms that was used in the 80s and 90s when computer vision system was used in industry

automation. So, today when you say machine vision, you often talk about vision systems that are deployed as part of a larger machine or an industry system. That is what you typically mean by machine vision.

Whereas, computer vision refers to the particular set of methods or algorithms that you use to extract knowledge from a camera. So, in a sense, computer vision methods are used as part of machines in machine vision. Once again, I think it is a perspective of usage and context and scope that you have in a particular setting that you use these terms in.

So, in terms of the various topics that are allied to computer vision obviously, Artificial Intelligence, Artificial Intelligence is the larger realm and computer vision forms one aspect of Artificial Intelligence at this time, although computer vision has other facets too. Machine Learning of course, a lot of the way computer vision problems, computer vision problems are solved today is through Machine Learning methods, mathematics, no surprise.

Neurobiology, because that is what helps us understand how the human visual system works. Imaging, how do cameras image is seen around us is important to be able to solve computer vision. The physics or the optics that is involved in lenses and how light falls on a particular object, how an object's appearance changes when the light source changes, when the direction of the light source changes, when the colour of the light source changes, or there are when there are multiple light sources so on and so forth.

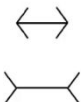
Signal processing because at the end of the day, an image is a 2-dimensional signal. So, a lot of the basic computer vision concepts follow through from signal processing, control and robotics play an important role especially in robotics vision and of course there are various other peripheral areas such as cognitive vision, statistics, geometry, optimization, biological visions, smart cameras, optics, multi-variable and non-linear signal processing, robotic vision so on and so forth.

And these are all connected to those larger topics that you see them organised next too. So, that gives you a prospect of the various topics that are connected to computer vision obviously, we would not have the scope to cover all of them and soon talk about what we are going to cover in this particular course.


(Refer Slide Time: 13:34)

Computer Vision: What and Why

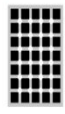
Why is it hard?¹




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?

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

Vineeth N B (IIT-H) 1.1 Introduction 8 / 21



So one question here is why why is this hard? Our brain seems to be so good at perceiving objects around us. So, in fact we will talk about the history of computer vision in subsequent lecture but when people started studying computer vision in the 1960s, it was assumed that computer vision is something that you could solve through a summer project.

And obviously, researchers have long since realised that it is not that simple but let us try to understand why is it really hard? So, one simple way of trying to understand it is by looking at optical illusions because they help you understand why even we do not understand how we perceive certain things in the visual world around us. So, very popular example is this Muller-Lyer illusion where one asks this question, which line is longer?

So, if you did not think too hard, you would probably say that the line that is longer is the one below but the true answer is that both these lines are exactly the same length. Just the contextualization of those lines with those angle parentheses makes you get an illusion that one of the lines is longer than the other.

And there is a very popular illusion called the Hermann grid illusion, named after a German Physiologist called Hermann where if you simply looked at that grid, front of in front of your eyes, you would start hallucinating black dots and white dots at the intersection of those grid points. So, in fact some argue whether this is an illusion or a hallucination, but it still proves the point that human vision is hard to understand and how we perceive things is still hard to understand.

Another popular example is called the Adelson's brightness constancy illusion. So, this again link, please do take a look at that link later on. But the question that we want to ask here is, if you see closely here, there are two squares A and B that are labelled on that checker board. And we ask the question, which is brighter, A or B? What is your answer? So, the truth is obviously, it looks like B is brighter and A is darker.

The truth of it is A and B have exactly the same intensity at the level of a pixel. The pixels that are there in the A block and the B block exactly have the same intensities. This can be a little surprising. But that once again based on how human brain uses context make probably understand, so there is a shadow falling on B and it compensates for that to make B look brighter than A. But A and B exactly have the same pixel intensity.

Once again, the brain does an amazing job here, but to make a computer do this job, is hard. And finally this is an example where you have two sets of crosses. So, on the left part, you have a bunch of red crosses. On the right part, you again have a bunch of red crosses. And the task for you is to count the red crosses in both figures. I let you take a moment so that, clearly if I ask you the question which is harder?


You would say the one on the right because there are some red circles that are putting of your counting process and making it harder to count although otherwise, it is the same red crosses on both sides. So, this is something again that makes understanding how visual perception works hard.

(Refer Slide Time: 17:25)

Computer Vision: What and Why

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). ⇒ 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 ⇒ heavy computational requirements
- Computer vision is AI-complete





Credit: Anish Chopra, Medium.com

Vineeth N B (IIT-H)

§1.1 Introduction

9 / 21

More deeply, if you were you have to ask why computer vision is hard. Most of the practical use cases of computer vision or what is known as inverse model applications. Let us know what does this mean. So, this means that we have no knowledge of how an image was taken or what were the camera parameters when the image was taken. So, you do not know at what angle the image was taken?

What were the settings on your camera when an image was taken? But we need to model the real world in which the picture and the video was taken. So, that is why we call it an inverse problem, we do not know what was the world in which the image was taken? But we actually have to find out from the image as to what are the 3D characteristics of that world in which the image was taken?

So, an often in practice, we almost have to model this from incomplete, partial, noisy information. By that we mean, there could be noise in the image. The noise could come from some motion blur, could come from noisy speck on the lens, could come from noise on the CCD or the CMOS silicon that captures the image or could come from just the processing elements that were involved at various stages of the pipeline. But the noise could come from anywhere.

On the other hand, to just make this clear, forward model. So, we said this is an inverse model so an obvious contour question then is, what is a forward model? A forward model is what could be used in say, physics or computer graphics where you can define the various parameters that you have and you create the image in that setting like animation or graphics for that matter where you say, I am going to place a light source at this angle, the light source is blue in colour and I have an object setting at a particular location, I have another light source that is red colour in some other part of the room falling on the object and now we can actually create the room.

Those kinds of applications should be called forward models of vision, while the computer vision that we are going to talk about in this course is the inverse model and a lot of applications such as object recognition or deduction or segmentation are all inverse models. Secondly, one of the problems in computer vision is that, image data is very high dimensional. So, even if you took a single 1 mega pixel image which today is not very high resolution by the kind of cameras that we all have in our smart phones.

Remember 1 mega pixel image is 10^6 pixels. And once again if you took a colour image, you typically would have as red channel, a green channel and a blue channel in RGB image which means you actually would have 3×10^6 pixels in that particular 1 mega pixel image. So, that is 3 million measurements for that 1 mega pixel image and that is a very huge dimension to work with when you use machine learning algorithms for this kind of a data.

So, which means images have image processing such images and getting understanding from these images has very heavy computational requirements and another thing here is computer vision is said to be what is known as AI-complete. So, once again anything that is blue in the slides is clickable so, please do feel free to click on it and learn more about what AI-complete is, but in very simple words AI-complete means that a computer cannot solve the problem by itself and you need to bring a human to be able to solve the problem.

So, that is what we call as human AI-complete in very in very crude words, please be free to click on that link to understand more. And as you can see here on the right, if you simply want to check whether a photo that you click is in a particular National Park, you can probably do it easily but if you want to find out final details of a particular photo that you took, what bird is it? Which particular kind or a sub kind of a bird are you looking at. These kind of things are very hard computer vision problems.

(Refer Slide Time: 21:28)

Computer Vision: What and Why

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?

WHEN A USER TAKES A PHOTO, THE APP SHOULD CHECK WHETHER THERE IS A NATIONAL PARK...


SURE, BUT YOU'LL HAVE TO CHECK A FEW THINGS...

...AND CHECK WHETHER THE PHOTO IS OF A BIRD.


IT'LL NEED A RESEARCH TEAM AND FIVE YEARS.

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

Credit: Anish Chopra, Medium.com



NPTEL




NPTEL
National Programme on Technology Enhanced Learning

Vineeth N B (IIT-H)

1.1 Introduction

16 / 21



Further points here are that no complete models of the human visual system itself are known at this time. Most of the existing models of the human visual system relate to subsystems and are not holistic in their understanding in terms of the entire central nervous system and there are also questions that you can ask as what is perceived and what is cognized especially things like optical illusions, there is a significant difference especially if you took the example of the Hermann grid illusion that we saw, what you perceive and what you cognized are two different things.

And you may look at a piece of thread in the night and think it is a snake so once again what you perceive and what you cognized are different here so, that makes the problem even harder and you could also ask the question when is an object important for a task and when is the context around the object important for a task. These are difficult questions to answer and some of these have been answered very convincingly today and some of them still remain open research questions for the field of computer vision problem.

And also the verifiability of the mathematical or physical models for these kinds of systems is non-trivial. For example, if you found a good way to represent an image as a low dimensional vector per say, how should similarity or dissimilarity between representations be defined? Would that be a proper distance metric mathematically? Do all images in all settings, RGB images, thermal images, Lidar images, do all of them follow the same distance metric?

How do you change this distance metric between these images? Unfortunately, there is no universal answer, there is no clear verifiability for such. Even if you develop a method, how do you verify this in a in a in a very reliable way is also an open question. Finally, how would you, if you manipulate an image, what is typically called a counterfactual. A counterfactual is where you take an image and ask the question, what if I had a particular object in this imaged changed?

So, you had an image of the scene in front of you and let us say instead of a laptop, if you placed a table lamp there, how the scene looks like and what would change in your perception of that image? So, those are called counterfactuals. So, if you had a manipulation in a given environment, how would that change the behaviour of the perception, of the cognition of the image?

Can a physical model or a mathematical model of your environment actually capture those kinds of counterfactuals and how the counterfactuals affect your perception? These are open questions at this point and these are the reasons why computer vision stays a hard problem today.

(Refer Slide Time: 24:14)

This Course: Topics, Structure, Objectives

Outline

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

Vineeth N B (IIT-H) 11.1 Introduction 11 / 21

NPTEL

video@iitb.ac.in

So, with that overall briefly introduction to computer vision, why we need to study it, why is it hard, let us actually talk about this course and what we going to cover in this course and what this is going to be the structure of the topics.

(Refer Slide Time: 24:30)

This Course: Topics, Structure, Objectives

Computer Vision: Topics

- Learning-based Vision**
 - Visual Recognition, Detection, Segmentation, Tracking, Retrieval, etc
- Geometry-based Vision**
 - Feature-based Alignment, Image Stitching, Epipolar Geometry, Structure from Motion, 3D Reconstruction, etc
- Physics-based Vision**
 - Computational Photography, Photometry, Light-fields, Color Spaces, Shape-from-X, Reflection, Refraction, Polarization, Diffraction, Interference, etc

Focus of this course

Vineeth N B (IIT-H) 11.1 Introduction 12 / 21

NPTEL

video@iitb.ac.in

So, the way computer vision has evolved over the last few decades, we will talk about the history of computer vision in L02. You can broadly classify the topics into learning based vision, geometry-based vision and physics based vision. Once again these topics are fairly porous, so it is not that what topic does not flow into the other. Often to solve problems, you need to use transcripts from various sub-topics.

So, by learning-based vision we generally refer to problems in computer vision such as visual recognition such as object recognition, gesture recognition or emotion recognition from images, detecting your face in Facebook, in a Facebook photo or segmentation where you segment to seen into various parts at the pixel level or tracking an object or retrieving an image based on the search query so on and so forth.

Typically, all of these are solved using a learning based methods so, that is one segment of computer vision. Then there is geometry-based vision where you talk about feature based alignment, image stitching such as the Panorama that you see on your cell phone, epipolar geometry, structure from motion, 3D reconstructing or so on and so forth. Typically, you study that as part of geometry-based vision.

And finally, this physics-based vision where you talk about computational photography, photometry, light-fields, colour spaces, shape-from-X, where X should be shading or structure or any other any other topic, reflection, refraction, polarization, diffraction, interference, so on and so forth.

So, clearly you can see a lot of this has to deal with the physics of how an image is created. So, these are all different subfields of computer vision that have developed over the years. And the focus of this course is going to be learning-based vision and that is the reason this course is called deep learning for computer vision.

(Refer Slide Time: 26:25)

This Course: Topics, Structure, Objectives

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

Vineeth N B (IIT-H) 1.1 Introduction 13 / 21

NPTEL
National Programme on Technology Enhanced Learning

So, let us talk about how this course is structured. So, we have tried to divide the course into 5 different segments just for simplicity of covering topics. The first segment we will talk about the journey so far which covers traditional computer vision topics. We will talk about basis here including image formation, linear filtering, convolutions, correlation, edge detection, blob detection, feature detection, feature descriptors, feature matching so on and so forth.

That could be the traditional computer vision segment. Then we will go on to segment 2 where we will talk about the building blocks of deep learning for computer vision. We will quickly review neural networks, back propagation which we assume that you have knowledge of as I said this course is built on top of Machine Learning and deep learning courses. We will quickly review neural networks though because it is important to understand various other topics.

We will cover convolutional neural networks, various architectures and models that have evolved over the last few years and also try to understand how do you visualize and understand how CNNs work. In the third segment, we will talk about the various different applications and use cases and tasks in which CNNs are used, recognition, verification, retrieval, detection, segmentation, there are various models and loss functions and settings in which the CNNs are used here. We will talk about all of them in segment 3.

Then we will go to segment 4, we will add a new dimension there where we will talk about recurrent neural network, Spatio-Temporal models where we will talk about computer vision for video especially, action recognition, activity recognitions, so on and so forth. We will also talk about attention models here and vision language tasks such as image captioning, visual dialog or visual question answering so on and so forth.

And finally in the last segment, we will focus on staying contemporary where we will talk about deep generative models such as GANs and variational auto encoders, we will talk about learning with limited supervision such as Few-shot learning, One-shot learning, Zero-shot learning, continual learning, multitask learning so on and so forth. And we will also wrap up the course with some recent trends in the field. So, there will be some programming assignments on these areas, there will be weekly quizzes and we will obviously have an exam at the end for those of you who are trying to credit this course for in your own college.

(Refer Slide Time: 28:53)

This Course Topics Structure Objectives

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

Vineeth N B (IIT-H) 1.1 Introduction 14 / 21

NPTEL

NPTEL

NPTEL

NPTEL

What are the pre requisites? Completion of a basic course in Machine Learning we are going to assume that you know Machine Learning we are going to build on top of that. I would really hope that you have also done a course on deep learning before because we are not going to cover deep learning in it is entirety, but we will review concepts in deep learning to the to the extent you need for this course, but it will be highly be recommended if you cover a completely different deep learning course in a holistic way.

And obviously you need strong mathematical basics in probability, linear algebra and calculus. And this course is designed as advanced under graduate or post graduate electives. Through a programming perspective, we are going to hope that you are comfortable with programming in Python, knowledge of a deep learning framework such as PyTorch or TensorFlow will be highly recommended. You can pick it up as part of the course you want but we will highly recommend it if you already, if you already know that.

(Refer Slide Time: 29:54)

Resources and References

Traditional Computer Vision: References



Book website



Book website



Book website



NPTEL



Indian Institute of Technology Bombay

Vineeth N. B. (IIT-H) §1.1 Introduction 16 / 21



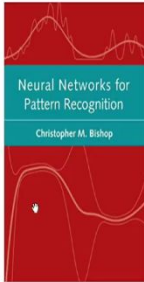

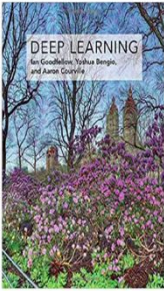
In terms of resources and references, the popular computer vision books that we will refer to, which have also follow widely across the world is Computer Vision A Model Approach by David Forsyth and John Ponce. The book website is linked here. Computer Vision Algorithms and Applications by Richard Seliski. It is a very popular book. It is there in the open public domain for you to download the pdf, we will follow this book for certain topics too.

And also Computer Vision models, learning and inference. Once again the book website has the book and you can download it if you like by Simon Prince. These are the popular computer vision books that we will follow.

(Refer Slide Time: 30:33)

Resources and References

Deep Learning: References






A nice, short online book by Michael Nielsen

Book website

Book website

Vineeth N B (IIT-H) 1.1 Introduction 17 / 21



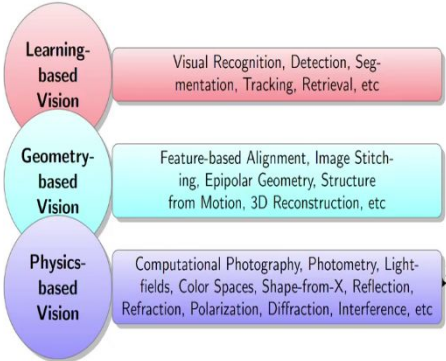
There are also the popular Deep Learning references. Deep Learning by Ian Goodfellow, Yoshua Bengio and Aaron Courville, the book website has the book. This also very nice short online book my Michael Nielsen which will refer to when required. And also the book by Christopher Bishop on Neural Networks for Pattern Recognition. While these are good references for you to follow.

For each lecture we will point you to where you can look at and sometimes we will just point you to good blogs and post that are online. These days you have some excellent blogs that explain concepts well. So, we will point you to those as and when required and possible.

(Refer Slide Time: 31:08)

Resources and References

Want to Learn Other Topics?






Learning-based Vision
Visual Recognition, Detection, Segmentation, Tracking, Retrieval, etc

Geometry-based Vision
Feature-based Alignment, Image Stitching, Epipolar Geometry, Structure from Motion, 3D Reconstruction, etc

Physics-based Vision
Computational Photography, Photometry, Light-fields, Color Spaces, Shape-from-X, Reflection, Refraction, Polarization, Diffraction, Interference, etc

Book Link:
Physics-Based Vision:
Principles and Practice
+ Relevant Course
Links: [Link 1](#)

Vineeth N B (IIT-H) 1.1 Introduction 19 / 21



So, if you wanted to learn geometry-based vision, so I said that the course is focused on learning based vision. So, if you want to learn geometry-based vision, what do you do? Is a very nice book called Multiple View Geometry in Computer Vision. It is again linked here. Please do take a look at that book if you want to learn and there is also a course on NPTEL for geometry-based vision. Please click on this link for these weeks in that particular course if you want to learn a little bit more about topics in geometry-based vision.

And if you wanted to learn physics-based vision, once again a very nice book called Physics-Based Vision Principles and Practice is available on this link. And there is also web based course which is linked here which you can follow if you want to learn more about physics-based vision.

(Refer Slide Time: 31:55)

The screenshot shows a presentation slide with a grey header bar containing the text "Resources and References". Below the header, the word "Homework!" is written in red. A bullet point follows: "Go through all links on the Applications of Computer Vision slide (Slide 6) - they are interesting views/reads!". To the right of the text are two NPTEL logos. At the bottom of the slide, a grey bar contains the text "Vineeth N B (IIT-H)" on the left, "1.1 Introduction" in the center, and "20 / 21" on the right. A small video inset in the bottom right corner shows a man with dark hair and a light blue shirt speaking.

So, one of the homework that we are going to let you take away from this inaugural introductory lecture is, go through all the links, all the applications of Computer Vision slide that was slide 6 for you. They are very very interesting views or reads to open your world to a lot of different computer vision applications which I think that some of you at least may not have come across. That is going to be the homework for your first lecture today.

(Refer Slide Time: 32:17)

The slide is titled "Acknowledgements and Disclaimers" in red text. It contains three bullet points: 1. 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. 2. Any errors in the material are our own. Please point out such issues, and we will be glad to rectify. 3. 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. The slide also features the NPTEL logo (National Programme on Technology Enhanced Learning) and the IIT-H logo (Indian Institute of Technology Hyderabad). At the bottom, there is a navigation bar with the text "Vineeth N B (IIT-H)", "1.1 Introduction", and "21 / 21". A small video inset shows a man speaking.

Resources and References

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.

Vineeth N B (IIT-H) 1.1 Introduction 21 / 21

So, to wrap-up the first lecture, I think I should thank many many people that have contributed to the creation of the slides and to my own knowledge in delivering this lecture. We are very grateful to the deep learning with computer vision courses and contents that are publicly available online. Wherever possible and relevant, these sources have been cited but if you notice an oversight, please let us know, and we will be glad to acknowledge.

I will thank all of the specific people that have contributed to creation of lectures. At the end of this course because there is a chance that I may add some contents, a little later down. I want to ensure that I thank all of them towards the end. We will do that towards the end. And obviously any errors in the material are our own, please do point out some issues, we will be glad to rectify, rectify and to the extent possible, the images used have been taken from free stock photos or public spaces to avoid copyright violations.

But if there is an oversight, please let us know and we will take it down or replace the image with something, something else. There are also teaching assistants that have been involved in creating the slides and helping make this course possible to offer to you. Once again, so of these teaching assistance contacts and names will be shared with you in the next few weeks. But we will also acknowledge all of them at the end of this course for I mean, there may be other people that contribute as the course progresses towards the end. So that is the end of first lecture, we will continue soon with the next topics. Thank you.