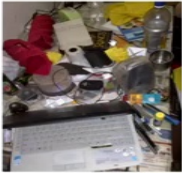Let us move on to the topic of this lecture.As most of you may know images are formed when a light source hits us the surface of an object and light is reflected and some of that light is reflected onto an image plane which is then captured through optics on to a sensor plane.So, that is the overall information and the factors that affect the image formation are the light source strength and direction, the surface geometry, material of the surface such as its texture as well as other nearby surfaces that, whose light could get reflected onto the surface, the sensor capture properties we will talk more about that as we go and the image representation and color space itself.We will talk about some of these as we go.

So, to study all of these one would probably need to study this from geometrical perspective, where you study 2D transformations, 3D transformations, camera calibration, distortion.So, we would not cover all of these but cover a few relevant topics from these in this particular lecture.If you are interested in a more detailed coverage of these topics please read chapters 1 to 5 of the book by Forsyth and Ponce.Starting with how light gets reflected off a surface the more typical morals of reflection state that when light hits a surface there are 3 simple reactions possible, there are more than 3 but 3 simple reactions to start with.Firstly, some light is absorbed and that depends on a factor called albedo rho and typically when you have a surface with low albedo more light gets absorbed.Some light is reflected diffusively.It scatters in multiple directions, so that happens independent of the viewing angle.Example of surfaces where lights scatters diffusively is brick, cloth, rough wood or any other texture material and in this scenario Lambert s cosine law states that the amount of reflected light is proportional to cosine of angle from which you are viewing the reflection.And finally, there are also phenomena such as fluorescence, where the output wavelength could be different from the input wavelength or other phenomena such as phosphorescence.

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

Vineeth N B (IIT-H)   §1.1 Introduction   5 / 21

And there are models to evaluate how bright the surface appears.So from a view point of colour itself we all know that visible light is 1 portion of the vast electromagnetic spectrum, so visible light is one small portion of the vast electromagnetic spectrum, so we know that infrared falls on one side, ultraviolet falls on the other side and there are many other forms of light across the electromagnetic spectrum.So, coloured light which arrives at a sensor typically involves two factors, colour of the light source and colour of the surface itself.So, an important development in sensing of colour in cameras is what is known as the Bayer Grid or the Bayer Filter.So, not every sensing element in a camera captures all three components of light you may be aware that typically we represent light as RGB at least coloured light as RGB; Red Green and Blue.We will talk a little bit more about other ways of representing coloured light a little later, but this is the typical way of representing coloured light and not every sensing element on the camera captures all three colours instead a person called Bayer proposed this method in a grid manner where you have 50 percent green sensors, 25 percent red sensors and 25 percent blue sensors which is inspired by human visual receptors.And this is how these sensors are checkered, so in a real camera device you would have a sensor array and there is a set of sensors that captures only red light, there is set of sensors that captures the green light, there is set of sensors that captures the blue light and to obtain the full colour image demosaicing algorithms are used where surrounding pixels are used to contribute the value of the exact colour at a given pixel.These are known as demosaicing algorithms.This is not the only kind of colour filter.

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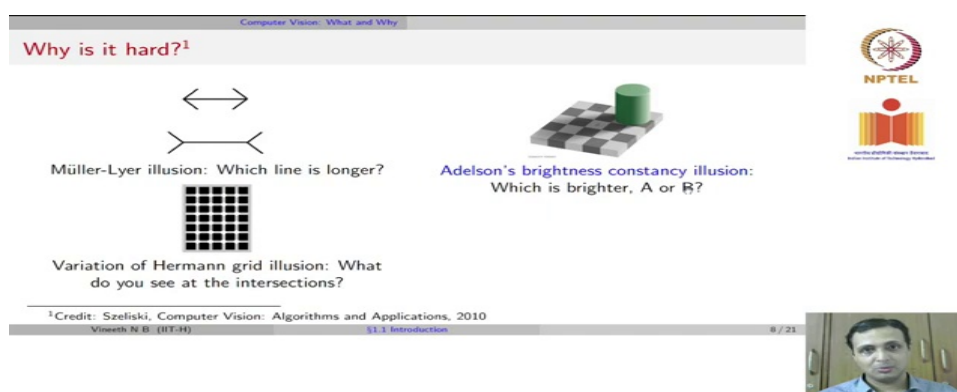
Vineeth N B (IIT-H)   §1.1 Introduction   6 / 21

So, this is the general pipeline of image capture.So, let us try to revisit, visit some of these components over the next few minutes.So, first thing is the camera sensor itself so you all must have heard of CCD and CMOS.This is often common decision to be made when you buy a camera these days a lesser issue but earlier days it used to be even more.What is the difference? So, the main difference between CCD and CMOS is that in CCD it stands for Charged Coupled Device.

Perspectives of Study
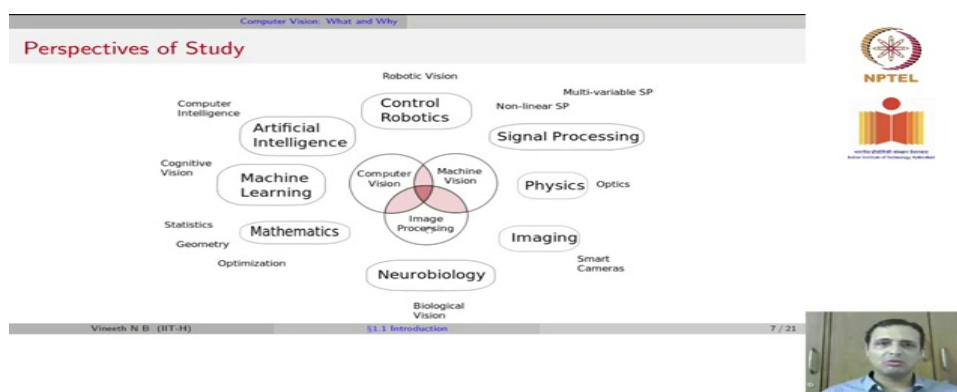
So, the many properties that you may actually see when you look at, when you take a picture on a camera.Sampling pitch, which defines a spacing between the sensor cells on the imaging chip.Fill factor or also known as active sensing area size, sorry, which is the ratio of the active sensing area size with respect to the theoretically available sensing area on the sensing element.Chip size which is the entire size of area of the chip itself.Analog gain which is the amplification of the sense signal using automatic gain control logic we would not going to the details of each of this once again if you are interested you can read the references provided at the end of this lecture to get more details of all of them.Typically, analog gain is what you control using your ISO setting on your camera, you can also have sensor noise that comes from various sources in the sensing process.Your resolution tells you how many bits is specified for each pixel which is also decided by an analog to digital conversion module in CCD or in case of CMOS in the sensing, in the sensing elements.Mirrored cameras such as DSLR also give you a physical shutter mechanism variable focal length and aperture so on and so forth.That is the reason there is value for DSLR cameras despite the advancement in smartphones cameras.So, the other factors that you need to understand when you talk about image formation is the concept of sampling and Aliasing, we will talk about this in more details bit later but a brief review now is Shannon Sampling Theorem states that if the maximum frequency of your data on your image is f_{max} you should at least sample at twice that frequency.

Why is it hard?[1]

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?

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

Why so, we will see a bit later but for the moment that frequency that you captured it is also called the Nyquist frequency and if you have frequencies about the Nyquist frequency in your image then the phenomenon called Aliasing happens.So, why is this bad and what impact can it have on image formation? This can often create issues when you up sample or down sample an image.Other colour spaces that are used in practice are XYZ, YUV, Lab, YCbCr, HSV so on and so forth.There is actually an organization call the CIE which establishes standards for colour spaces because this is an important, this is actually important for the printing and scanning industry, I think this is extremely important people working in that space.So, that is the reason there are standards establish for these kinds of spaces, we would not get into more details here once again if you are interested please go through these links below to know more about colour spaces what do you mean by additive, subtractive, so on and so forth, please look at these links.Finally, the last stage in image formation is image compression, because you have to store the image that you captured, so typically you convert the signal into a form called YCbCr where Y is luminance CbCr talks about chrome what is known as colour factor or the chrominance and the reason for this is that you typically try to compress luminance with a higher fidelity than chrominance.

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). $\implies$ 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

Because of the way humans or the human visual system perceives light, luminance is a bit more important than chrominance, so you ensure that luminance is actually compressed with a higher fidelity which means your reconstruction is better for luminance than for chrominance, so that is one reason why YCbCr is used as a popular colour space before storage, once again if you do not understand YCbCr, go back to the previous slide look at all of these links to understand YCbCr is one of the colour space representations that are available in practice.And as I just mentioned so the most common compression technique that used to store an image is called the Discreate Cosine Transform which is popularly used in standard such as MPEG and JPEG Discreate Cosine Transform is actually a variant of Discrete Fourier Transform and it is a you can call it as a reasonable approximation of an eigen decomposition of image patches.So, we would not get into in for the time now, videos this is how images are compressed using method call DCT, videos also use what is known as block level motion compensation, so you also divide images into frames and set of frames into block and then you store certain frames based on concepts from motion compensation, this is typically used in the MPEG standard which uses, which divides all frames into what are known as i frames, p frames and b frames and then uses strategies to decide how each frame should be coded, that is how videos are compressed.

And compression quality finally is measured through a metric called PSNR, apologies for the typo, it will be fixed before the slides are uploaded, which stands for Peak Signal to Noise Ratio, sorry for these typos.PSNR is defined as 10 log to the base 10 i_max square by MSE, where i_max is the maximum intensity and MSE is simply talks about the mean squared error between the original image and the compressed image, how much is the mean squared error pixel wise between these two images.And the numerator talks about the maximum intensity that you can have in an image, so this is typically called as PSNR which is used to measure the quality of image compression, there are other kinds of matrix which are based on human perception but this is the most popular statistical metric that is used.

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, Lightfields, Color Spaces, Shape-from-X, Reflection, Refraction, Polarization, Diffraction, Interference, etc