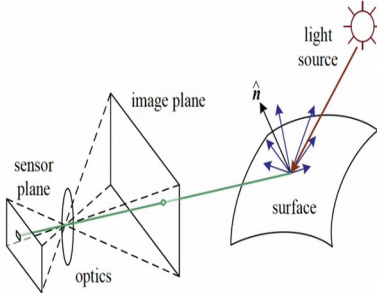


Deep Learning for Computer Vision
Professor Vineeth N Balasubramanian
Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad
Image Formation

Moving on from the history of computer vision, in this lecture we will talk about the formation of images. Before we go there, did you have a chance to check on the answer for the trivia question that we had last class? What was Lawrence Roberts known for? Besides his contribution to computer vision he is more well known for being one of the founders of the internet. Infact he was the project leader of the ARPANET project that was the precursor to the internet to the US defense organization - DARPA.

(Refer Slide Time: 00:57)



Image Formation: An Overview¹



The diagram illustrates the process of image formation. A light source (represented by a sun icon) emits rays that hit a curved surface. Some rays are reflected away, while others are reflected towards an image plane. The image plane is connected to a sensor plane via an optics component. A normal vector \hat{n} is shown at the point of incidence on the surface.

Factors


- Light source strength and direction
- Surface geometry, material and nearby surfaces
- Sensor capture properties
- Image representation and colour

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

Vineeth N B. (IIT-H)

§1.3 Image Formation



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.

(Refer Slide Time: 01:54)

Related Topics

- Geometry**: 2D Transformations, 3D Transformations, Camera Calibration, Distortion
- Photometry**: Lighting, Reflectance and Shading, Optics
- Colour**: Physics of Colour, Human Colour, Colour Representation
- Sensor**: Human Perception, Camera Design, Sampling and Aliasing, Compression

We will cover a few relevant topics from these

For a detailed understanding, read Chapters 1-5 of the book, *Computer Vision: A Modern Approach* by Forsyth and Ponce

Vineeth N B. (IIT-H) §1.3 Image Formation 3 / 18

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. From a photometric perspective where you study lighting, reflectance, shading, optics, so on. From a color perspective where you study the physics of colour, human colour, colour representation and from a sensor perspective looking at it from human perceptions, camera design, sampling and Aliasing, compression so on and so forth. 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.

(Refer Slide Time: 02:48)

Models of Reflection

When light hits a surface:

- Some light is **absorbed** ($1 - \rho$, ρ = albedo);
 - More absorbed for low albedos
- Some light is **reflected diffusively**, independent of viewing direction
 - E.g.: Brick, cloth, rough wood
 - Lambert's cosine law**: Amount of reflected light proportional to $\cos(\theta)$
- Some light is **reflected specularly**, depends on viewing direction
 - E.g.: Mirror



Credit: Derek Hoiem, UIUC

Vineeth N B. (IIT-H) §1.3 Image Formation

Starting with how light gets reflected off a surface the more typical models 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 (ρ) and typically when you have a surface with low albedo more light gets absorbed. So that is why you say it is the factor $1 - \rho$ for absorption. Some light is reflected diffusively. It scatters in multiple directions, so that happens independent of the viewing angle.

Example of surfaces where light 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 thirdly some light is also reflected specularly where the reflected light depends on the viewing direction. So, an example of a surface where this happens is a mirror where we all know that the reflected light follows the same angle as the incident light.

(Refer Slide Time: 04:15)




Models of Reflection

- Most surfaces have both specular and diffuse components
- Intensity depends on illumination angle because less light comes in at oblique angles
- Other possible effects:
 - Transparency
 - Refraction
 - Subsurface scattering
 - Fluorescence, phosphorescence
- **BRDF - Bidirectional Reflectance Distribution Function:** Model of local reflection that tells how bright a surface appears when viewed from one direction when light falls on it from another

Credit: Derek Hoiem, UIUC

Vineeth N B (IIT-H) 51.3 Image Formation

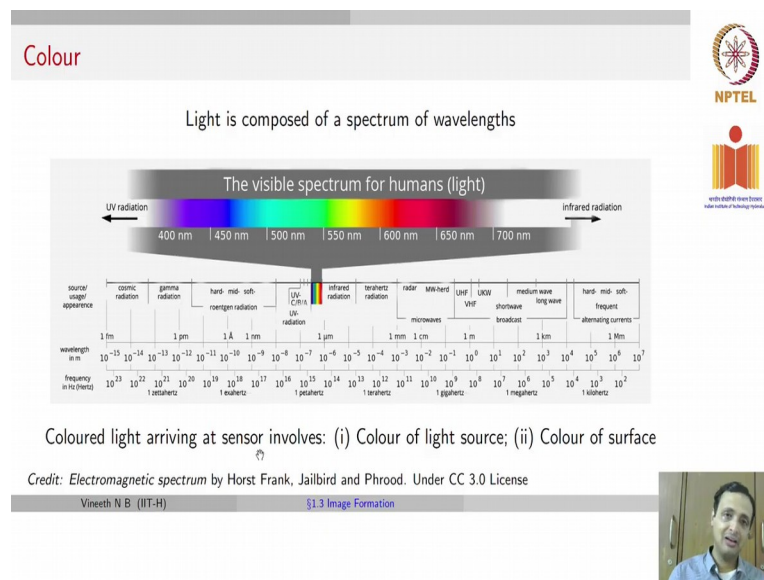


Generally, in the real world most surfaces have both specular and diffuse components and the intensity that you receive at the output depends also on the illumination angle because when you have oblique angle, lesser light comes through. And in addition to absorption, diffuse reflection and specular reflection, there are other actions possible like there is transparency, where light could pass through the surface, there is refraction such as a prism where light could get refracted there is also sub surface scattering, where multiple layers of the surface could result in certain levels of scattering. 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.

An important concept that is also studied here is called the BRDF or the Bidirectional Reflectance Distribution Function which is a model of local reflection that tells us how bright a surface appears from one direction when light falls on it from another direction, another prespecified direction. And there are models to evaluate how bright the surface appears.

(Refer Slide Time: 05:46)



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.

(Refer Slide Time: 06:26)

Bayer Grid/Filter

- Bayer arrangement of color filters on a camera sensor
- Filter pattern is 50% green, 25% red and 25% blue
- To obtain full-colour image, **demosaicing** algorithms used - surrounding pixels used to estimate values for a particular pixel.

Credit: https://en.wikipedia.org/wiki/Bayer_filter

Vineeth N B (IIT-H) §1.3 Image Formation

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

So, an important development in sensing of colour in cameras is what is known as the Bayer Grid or the Bayer Filter. The Bayer Grid talks about the arrangement of colour filters in a camera sensor. 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.

So, that particular sensing element will have its own colour that you also use the surrounding elements to find out to assign a colour at that particular sensing element. These are known as demosaicing algorithms. This is not the only kind of colour filter. Bayer Filter is one filter that is more popular especially in single sensors cameras, but there has been other kinds of filters, other kind of colour grading mechanism that have been developed over the years too.

So, you can also read a little bit more about this on the Wikipedia entries of Bayer Filter which also talks about other kinds of mechanisms that are used.





(Refer Slide Time: 08:41)

Question

On Colour
If visible light spectrum is VIBGYOR, why RGB colour representation?

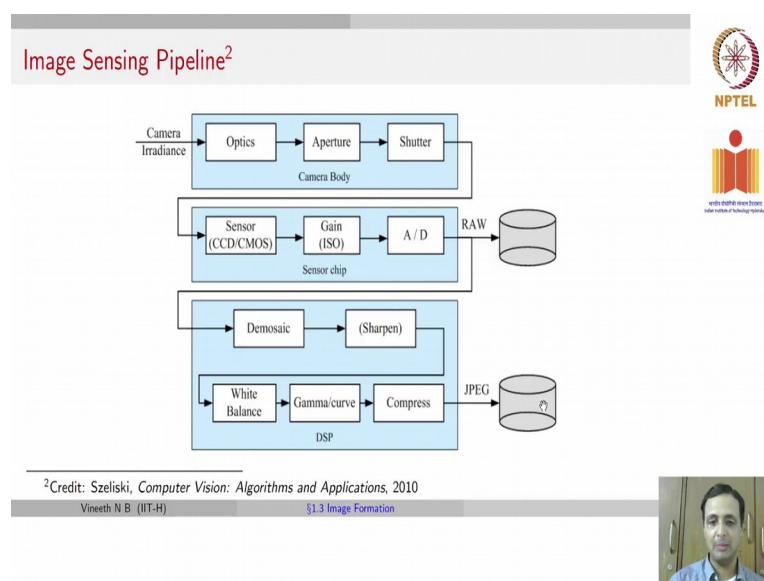
Vineeth N B (IIT-H) §1.3 Image Formation

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



So, let us ask one question for you to think, if visible lights spectrum is VIBGYOR or Violet, Indigo, Blue, Green, Yellow, Orange, Red, why do we use an RGB wave representing colour? There is something for you to think about, we will answer it in the next class at least try to find this out yourself if you can.

(Refer Slide Time: 09:01)



So, the image sensing pipeline in a camera follows a flow chart such as this, where you have the optics such as the lens. Ofcourse light falls in through that. You have an aperture and

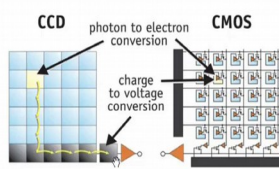
shutter parameters that you can specify or adjust and from there light falls onto the sensor. Sensor can be CCD or CMOS, we will talk about these variants very soon.

Then there is a gain factor, we will talk about that also soon. Then the image is obtained in an analog or digital form which represents the raw image that you get, the typically cameras do not stop there, you then use demosaicing algorithms which we just talked about, we could, you could sharpen the image if you like or any other important image processing algorithms. Some white balancing, some other digital signal processing methods to improve the quality of the image and finally you compress the image into a suitable format to store the image. So, this is the general pipeline of image capture.

(Refer Slide Time: 10:12)

Digital Image Sensing




CCD vs CMOS: What's the difference?



CCD

- Move photogenerated charge from pixel to pixel, and convert it to voltage at output node
- An analog-to-digital converter (ADC) then turns each pixel's value into a digital value

Photo Credit: Litwiller, CMOS vs. CCD: Maturing Technologies, Maturing Markets, 2005
Vineeth N B (IIT-H) §1.3 Image Formation



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.

You typically generate a charge at each sensing element and then you move that photogenerated charge, so the charge generated by a photons striking that sensing elements from pixel to pixel and you convert it to a voltage at an output node on that particular column. Then typically an ADC or an analog to digital converter converts each pixel's value into a digital value. This is how the CCD sensors work.

(Refer Slide Time: 11:15)

Digital Image Sensing

CCD vs CMOS: What's the difference?

CCD photon to electron conversion
charge to voltage conversion

CMOS

CMOS

- CMOS convert charge to voltage inside each element
- Uses several transistors at each pixel to amplify and move the charge using more traditional wires
- CMOS signal is digital, so it needs no ADC

NPTEL

Indian Institute of Technology Bombay

Vineeth N B (IIT-H)
§1.3 Image Formation

On the other hand, the CMOS sensors, complementary metal oxide semiconductors, they work by converting charge to voltage inside each element. So, CCD accumulates there is CMOS converts at each element it uses transistors at each pixel to amplify and move the charge using more traditional wires.

So, the CMOS signal is digital so it does not need any ADC at a later point of time. So, today CMOS, originally CMOS technologies had some limitations but today CMOS technologies are fairly well developed and most of the cameras that we use today are actually CMOS or CMOS devices.

(Refer Slide Time: 11:59)

Digital Image Sensor Properties

| | |
|-------------------------|--|
| Shutter speed: | Controls the amount of light reaching the sensor (also called <i>exposure time</i>) |
| Sampling pitch: | Physical spacing between adjacent sensor cells on the imaging chip |
| Fill factor: | Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches) |
| Chip size: | Size/area of the chip |
| Analog gain: | Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras) |
| Sensor noise: | Noise from various sources in the sensing process |
| Resolution: | How many bits for each pixel, decided by analog-to-digital conversion module |
| Post-processing: | Digital image enhancement methods often used before compression and storage of captured image |

NPTEL

Indian Institute of Technology Bombay

Vineeth N B (IIT-H)
§1.3 Image Formation

So, the many properties that you may actually see when you look at, when you take a picture on a camera. Shutter speed which controls the amount of light reaching a sensor or also called exposure time. 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.

So, which means if you use 8 bits to represent each pixel, so you could get a value going from 0 to 255 for each pixel that gives you the sensing resolution for that particular pixel, and finally there are also post processing elements as we already briefly mentioned such as digital image enhancement methods used before compression and storage of the captured image.




(Refer Slide Time: 13:48)

Popular Question: "With smartphones, do you need DSLR cameras?"³

- **DSLR - Digital Single Lens Reflex camera:** Uses a mirror mechanism to reflect light from lens to a viewfinder, or let light fully pass onto image sensor by moving the mirror out of the way
- Essentially a comparison between mirror and mirrorless cameras
- *Pros of mirrorless cameras:* Accessibility, portability, Low cost
- *Pros of DSLRs:* Picture quality³, Versatility and functionality, Physical shutter, Variable focal length/aperture


³Source: VSBytes.com

Credit: <http://www.pixelrajeev.com/>



Vineeth N B (IIT-H)

§1.3 Image Formation




So, one popular question that often can be asked here is, these days smartphones seem to be so good, you have very high-resolution cameras in smartphones, do you really need what are known as DSLR cameras. So, what are DSLR cameras? DSLR camera stand for Digital Single Lens Reflex camera and the main difference between a DSLR camera or any other point and shoot camera or a cell phone camera is the use of mirrors. DSLR camera uses a mirror mechanism to reflect light to a view finder or also can turn off the mirror, moving the mirror out of the way to actually reflect the light on the image sensor.

So, affectively the comparison here becomes between mirrored cameras and mirrorless cameras. So, mirrorless cameras such as what you see in your smartphones are more accessible, portable, low cost, whereas when you have a mirror, the picture quality tends to be better, you have more functionality possible, again we will not step into more details here but please do read down the sources of the links given under each slide if you want to know more. 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.


(Refer Slide Time: 15:22)

Sampling and Aliasing



Credit: Wikimedia Commons

- **Shannon's Sampling Theorem:** $f_s \geq 2f_{\max}$, where f_s is sampling rate, and f_{\max} is maximum frequency in signal, also called **Nyquist frequency**
- Frequencies above Nyquist frequency or when Shannon's sampling rate is not met \Rightarrow *aliasing* happens
- Why is aliasing bad? Creates issues while downsampling and upsampling an image
- More coming in later lectures

Vineeth N B. (IIT-H)
§1.3 Image Formation


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 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. If you capture an image at a particular resolution say 256 cross 256. If you choose to up sample or down sample Aliasing can be bad in those settings, we will see this in more detail a bit later in a lecture which will come in sometime.

(Refer Slide Time: 16:37)

Colour Space Representations

- Popular colour spaces: RGB, CMYK
- Additive colours: R, G, B
- Subtractive colours: C, M, Y
- Other colour spaces: XYZ, YUV, Lab, YCbCr, HSV
- Standards established by Commission Internationale d'Eclairage (CIE)
- Understanding of colour spaces important in printing industry

For more information:

- https://www.tutorialspoint.com/dip/introduction_to_color_spaces.htm
- <https://ciechanow.ski/color-spaces/>

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

Vineeth N B (IIT-H) §1.3 Image Formation

The slide features two Venn diagrams: the left one shows three overlapping circles of red, green, and blue (additive colors) on a black background, and the right one shows three overlapping circles of cyan, magenta, and yellow (subtractive colors) on a white background. Logos for NPTEL and the Indian Institute of Technology (IIT) are visible in the top right corner. A small video inset in the bottom right corner shows a man speaking.

Also, in terms of representing the image itself there are multiple colour spaces possible, while RGB is the most common one, people today use various other kinds of colour spaces not necessarily in a camera but in other kind of devices we will see that. I will mention that briefly now. Popular colour spaces are RGB and CMYK, CMYK stands for cyan, magenta, yellow and black that is what you see here. So, they are supposed to be; so additive colours are RGB, R, G and B; subtractive colours are C, M and Y particular application where CMYK is used in practice is in printers.

So, it happens that it is a lot easier to control colours using CMYK in printers, you can read more about these on these links provided below. 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.

(Refer Slide Time: 18:19)

Image Compression



- Last stage in a camera's processing pipeline
- Convert signal into YCbCr (or variant), compress luminance with higher fidelity than chrominance
- Most common compression technique: **Discrete Cosine Transform (DCT)**, used in MPEG and JPEG
- DCT, a variant of Discrete Fourier Transform - a reasonable approximation of eigendecomposition of image patches

- Videos also use block-level motion compensation
- Compression quality measured using **Peak Signal-to-Noise Ratio (PSNR)**:


$$PSNR = 10 \log_{10} \frac{I_{max}^2}{MSE}$$

where $MSE = \frac{1}{n} \sum_x [I(x) - \hat{I}(x)]^2$

where \hat{I} is compressed version of I

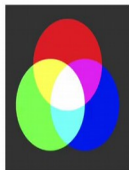






Vineeth N B (IIT-H) §1.3 Image Formation



Colour Space Representations

- Popular colour spaces: RGB, CMYK
- Additive colours: R, G, B
- Subtractive colours: C, M, Y
- Other colour spaces: XYZ, YUV, Lab, YCbCr, HSV
- Standards established by Commission Internationale d'Eclairage (CIE)
- Understanding of colour spaces important in printing industry





For more information:

- https://www.tutorialspoint.com/dip/introduction_to_color_spaces.htm
- <https://ciechanow.ski/color-spaces/>

Vineeth N B (IIT-H) §1.3 Image Formation

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



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.

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 Discrete Cosine Transform which is popularly used in standard such as MPEG and JPEG Discrete Cosine Transform is actually a variant of Discrete Fourier Transform and it is 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 $10\log_{10} \frac{I_{max}^2}{MSE}$, 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.

(Refer Slide Time: 21:40)

The screenshot shows a presentation slide with a light gray background. At the top, there is a header bar with the text 'Homework Readings' in small blue font. Below this, the word 'Homework' is written in a larger, bold, red font. In the center, there is a gray box with the title 'Readings' in bold. Inside this box, there are three bullet points: 'Chapter 2, Szeliski, *Computer Vision: Algorithms and Applications*', 'Other links provided on respective slides', and '(Optional, if you want to know about geometric formation of images) Chapters 1-5, Forsyth and Ponce, *Computer Vision: A Modern Approach*'. To the right of the slide, there are two logos: the NPTEL logo (a circular emblem with a star) and the IIT-Hyderabad logo (a stylized orange and yellow emblem). In the bottom right corner, there is a small video feed showing a man with dark hair, wearing a blue shirt, speaking. At the very bottom of the slide, there is a footer bar with the text 'Vineeth N B (IIT-H)' on the left and '§1.3 Image Formation' in the center.

Homework Readings

Homework

Readings

- Chapter 2, Szeliski, *Computer Vision: Algorithms and Applications*
- Other links provided on respective slides
- (Optional, if you want to know about geometric formation of images) Chapters 1-5, Forsyth and Ponce, *Computer Vision: A Modern Approach*

Vineeth N B (IIT-H) §1.3 Image Formation

That is about this lecture on image formation so if you need to read more please read chapter 2 of Szeliski's book, please also read the links provided on some of the slides specially one of those topics interest you or you are left with some questions please read those links. If you want to know in a more detail form about how images are captured including the geometric aspects of it and the photometric aspect of it please read chapters 1 to 5 of Forsyth and Ponce.