Deep Learning for Computer Vision
Prof. Vineeth N Balasubramanian
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad
Lecture 14
Human Visual System

(Refer Slide Time: 0:15)

For the last lecture of this week, we are going to look at whatever we have seen from a very different perspective of the human visual system. So, we saw that processing images can be done to achieve several tasks, such as extract edges, extract blobs, corners, key points, extract representations around key points, segment images, so on and so forth. For many decades, these were used extensively in computer vision applications.

In particular, one of the topics that we covered in the lectures, which was a bank of filters using a Gabor filter bank or steerable filters was about using multiple different filters in different orientations and scales to extract content out of images. To some extent, we will see now how this approach is similar to how the human visual system processes images. It is not exactly an imitation, but there are similarities in how these methods were used to process images versus how things happen in the visual, human visual system. So to complete that, let us look at a slightly detailed view of the human visual system.

(Refer Slide Time: 1:47)

To start with an acknowledgement, most of this lecture slides are taken from Professor Rajesh Rao's slides at University of Washington, so unless stated explicitly the imagesources are also the same.
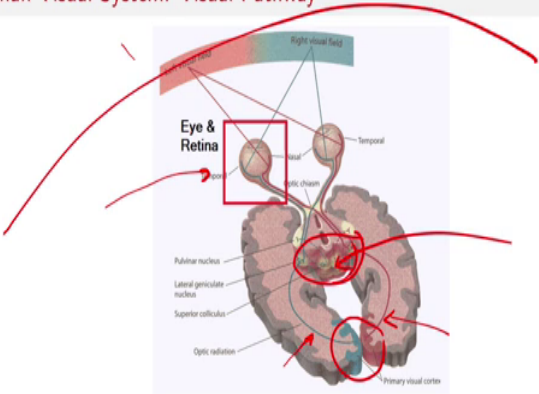
(Refer Slide Time: 2:03)

So the human visual system can be summarized in this diagram. There is a lot more detail than what you see in this diagram here, but what you see here is the eye and the retina, and the scene around you is here around the human and the left visual field and the right visual field fall on both the eyes and then you can see that the right eye goes to the left part of the brain, which is drawn in the blue color here and similarly, the input to the left eye goes to the right part of the brain drawn in red colors.

The primary visual cortex is located at the back and there are other components that the human visual system passes through, such as the pulvinar nucleus, LGN or the lateral geniculate nucleus, superior colliculus optic radiation, so on and so forth. So if you observe carefully here, among all the inputs that come in through the retina, most of it go to the visual cortex, but there is a slight deviation of some content here, which goes into the superior colliculus and the superior colliculus is what is responsible for feedback to moving the eye.

So the superior colliculus is what tells you to move your eyeballs to see something to get a better understanding, so on and so forth, while the visual cortex is what gives us understanding and perception of the scene around us itself. Let us see this in a bit more detail.

(Refer Slide Time: 3:48)
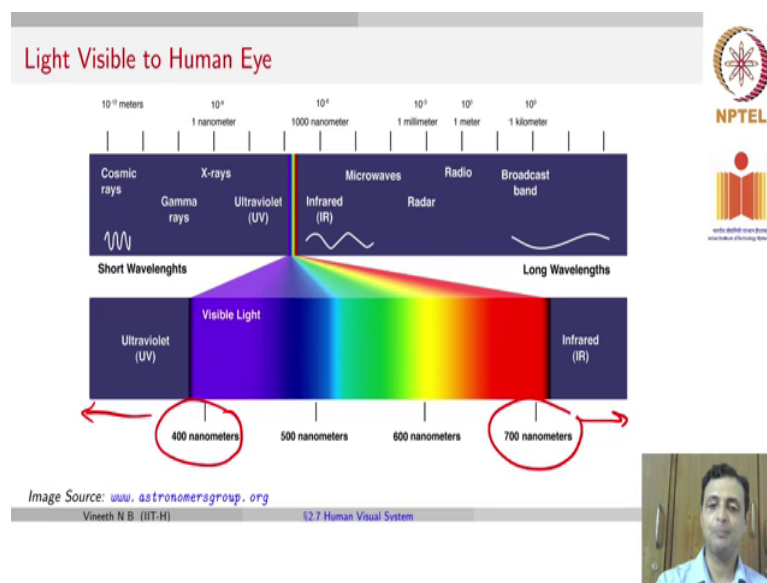
To start once again, we talked about this in an earlier lecture too that light visible to the human eye is restricted to one part of the electromagnetic spectrum, which goes from somewhere roughly between say a little less than 400 nanometers to a little over 700 nanometers going from violet to red. Obviously, the radiations that you have to the left of violet are called ultra violet and the radiations that you have to the right of red called infrared. So, this is known to us.

(Refer Slide Time: 4:24)
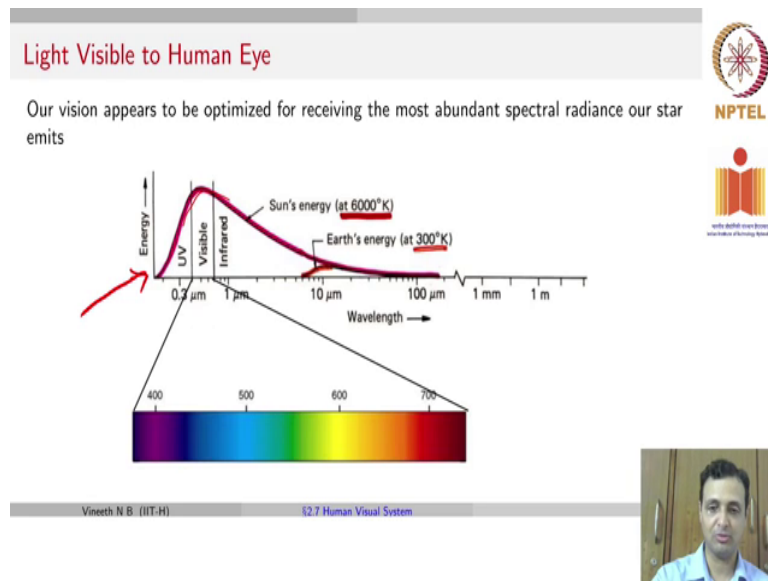
So, if you ask us, why is it that our eye receives only this light spectrum the most, it seems to be that as we have evolved our vision appears to be optimized for receiving the most abundant spectral radiance of our star, the sun. So, in this graph on top you see the energy of the various components in the electromagnetic spectrum, you can see that the sun's energy peaks in the visible spectrum and then falls off over the rest of the electromagnetic radiation. And so that is potentially a reason for why our eyes seem to have got used to that spectrum as the most useful spectrum from a vision perspective.

(Refer Slide Time: 5:18)
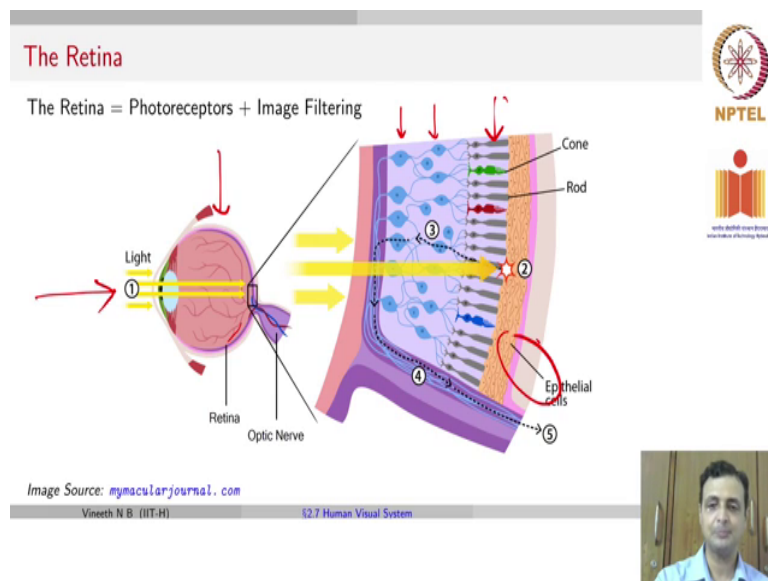
So the retina itself, which is the sensor of our human visual pathway, consists of photoreceptors, and also does a lot of image filtering, before it passes on information to the
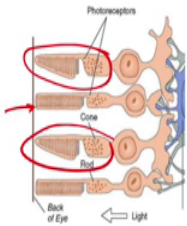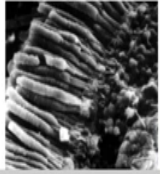
next phase of the human visual pathway. So if this was our retina and light fell from left to right here, so the back of the retina is blown up on the right side, so you can see that a bit more closer. So at the far end, it consists of, of course, epithelial cells, and just before the epithelial cells, the retina consists of what are known as the rods and cones, which you may have heard of. But before the photons fall on the rods and cones there are many other cells too, such as what are known as ganglion, bipolar cells, so on and so forth, which the information passes through before reaching the rods and cones. So, each of the rods and cones have specific properties.

(Refer Slide Time: 6:26)

The rods are sensitive to intensity, but not color and why are they called rods and cones they are shaped as you can see here, the rods are shaped like this, and the cones are shaped conically. So, the rods are sensitive to intensity, but are not sensitive to color, so they in some sense get a blurred image of what is happening around us. And cones are sensitive to color they form sharp images and require many more photons to absorb the information. Cones typically form three different types in humans each of these cones are sensitive to specific wavelengths.

(Refer Slide Time: 7:16)

Vineeth N B (IIT-H) §2.7 Human Visual System
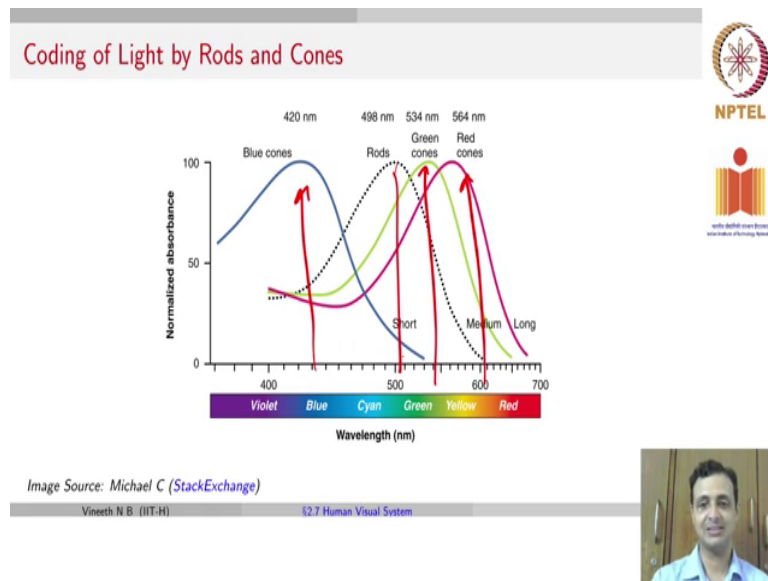
And what are these wavelengths? So you have a set of cones that respond very well to blue color, a set of cones that respond very well to green color, a set of cones that respond very well to red color. Clearly, rods are somewhere in between where they are not color sensitive, but just are sensitive to the intensity of the photons falling on the retina. I should also explain the RGB aspect of color that we choose because that seems to be where our cones are peaking in the VIBGYOR spectrum.

So this also explains why a person could be colorblind. So for example, if a person does not have green cones the person may not be able to see green color in the world around us.

(Refer Slide Time: 8:06)

Vineeth N B (IIT-H) §2.7 Human Visual System
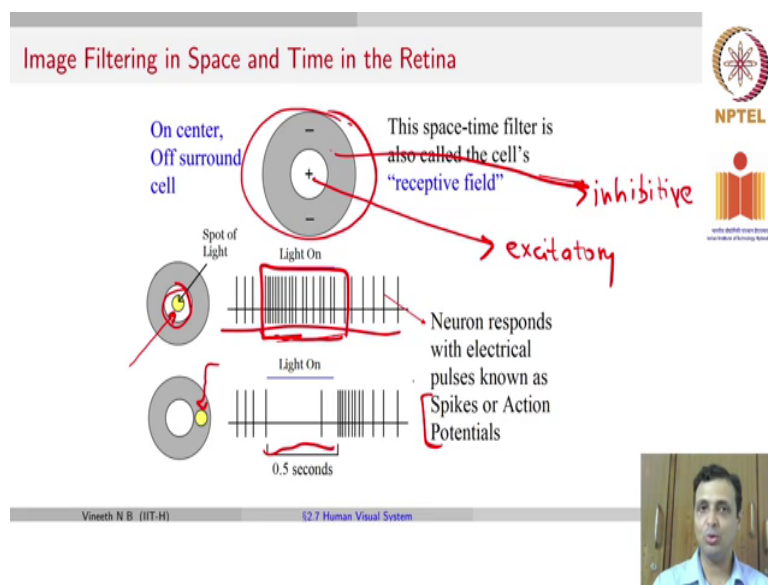
So, before the image the photons reach the rods and cones, there are what are known as ganglion cells or other cells in the retina, which typically operate in what is known as an excitatory manner or an inhibitory manner. So, in this diagram that you see on the slide, plus denotes an excitatory reaction and minus denotes an inhibitive reaction.

So, cells are organized this way, where there is a central cell, which is, which gets excited when a photon falls on it, and there are a set of cells around that gets suppressed when the photon falls on it. So what happens? Remember, at the end of the day, we will see this as we go through this lecture that even the eye access image filters and that is the reason we are talking about it now, having discussed image filters, edges, features, so on and so forth, it is perhaps a right time to be able to relate what we have discussed so far, to how things happen in the visual, human visual system.

One key difference between whatever we have studied so far to what we are going to talk about in the human visual system is that the human visual system does spatiotemporal filtering. It is not only spatial filters, which is largely what we've seen so far in this course, but it also does filtering over time. We will talk about this a bit more detail in the next few slides.

Before we go there, as we were saying, arrangements of cells in the retina have components of excitatory and inhibitory elements to them. So there could be an excitatory cell flanked by inhibitory cells on either side, so when a spot of light shines on that photon or the spot of light shines on the central cell, so when the light is on, you can see here these are just a set of impulses, remember that at the end of the day, the human brain or these cells release chemicals of spikes of electricity, as you can see, which are known as action potentials.

So, each of this is a spike and when the light is turned on, there seems to be an excitatory reaction, because the light is at the excitatory part those photons follow the excitatory part of the cell. On the other hand, if the light is on, and that part falls on the inhibitory part of the cells, you actually see that there is no response or spikes from the cells because those cells which are inhibitory, even when photons fall on them they actually suppress and do not throw out any potential. This idea of inhibitive an excitatory is extremely key to how our human visual system works.

(Refer Slide Time: 11:06)

So there are two kinds. The earlier kind is where we saw the excitatory to be in the middle, so it is called on-center off-surround cell. You also have the converse, where you have an off-center and an on-surround cell, in which case, you have an inhibitory cell in the middle and then an excitatory around it flanked on either side. In this case when the photon or the light is on and the photon falls on the middle cell, your action potentials or your spikes stop for some time this, so this particular set of spikes are spikes that you get over time.

So the light was on for that duration that you see there so that graph is a graph over time going from left to right. So when the light was on, you can see that there is no spike that comes out of that particular cells. Whereas, when in this case when the light is in the region outside the inhibitory cell those are the executive cells and you can actually see that they

throw a bunch of different spikes. So this idea of off-center and on-center where there are cells that inhibit and cells that excite are important components of how our visual system works.
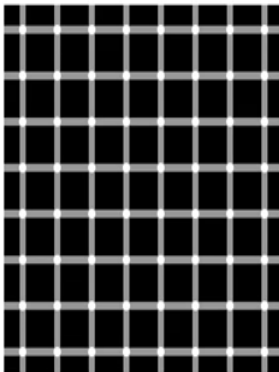
(Refer Slide Time: 12:30)
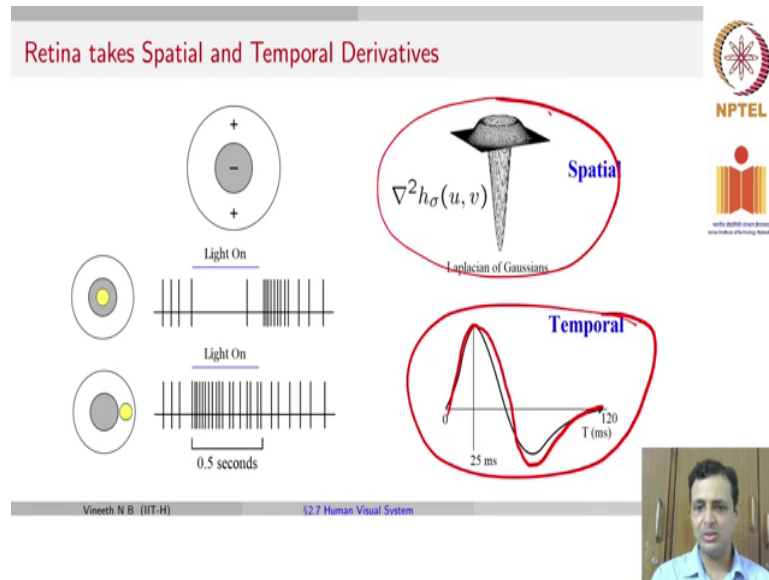
As I just mentioned, the human visual system is a spatiotemporal filter. So there's a filter on the spatial site, which largely resembles a blob detector or a Laplacian of Gaussian for large extent. So it could be either way. So you could have a Laplacian of Gaussian remember the other way, which can peak in the other direction. So you could look at for large part, they seem to assemble the Laplacian of Gaussians. But as I just mentioning, there is also temporal filter, which acts something like this graph here.

What does this graph mean? When the light is highest, you get the highest response. After that, you actually get a negative response before stabilizing, which means remember, again, that in a human visual system, it is a spatial temporal filter. So when you have a photon that shines, or you have an edge that falls on you, you are first going to detect the edge, then for a small few milliseconds, the reaction is going to be the opposite in terms of time, and then you revert back to a stable state. So that is what the temporal filter does.

Where can you see it taking effect? Why do you think this happens? Here is an example for that.
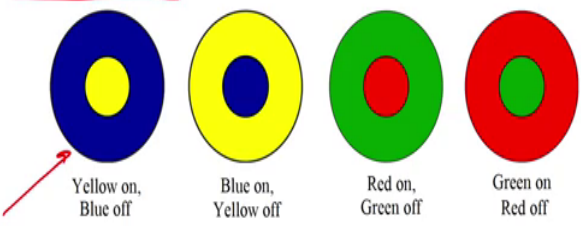
If you have seen this optical illusion, which is a common one, what do you think you see at the center, black dots are the intersections, black dots or white dots. This should explain to you what is happening in the eye. So if you see a white dot, when you move your eye from that, remember the response over time is to go back to the other side and make it look like a black dot before you recover and find out it is a white dot, and that is the reason such an illusion happens is because of how the temporal filter in the human visual system works.
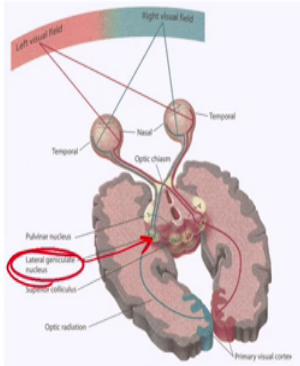
(Refer Slide Time: 14:30)
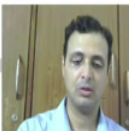
Another effect that you may have seen popularly is what is known as color-opponent processing. So, in this particular case, if you see a lot of these examples, these are also visual illusions, optical illusions but you can, you may have seen this has many other settings. When

you focus on some very strong colors, you typically have a negative after image. So you focus on the yellow and quickly move around you may find that to be a blue color and you get a negative afterimage, which again corresponds to the temporal filter that we are talking about, where you get an opposite response over time before stabilizing to an equilibrium.

(Refer Slide Time: 15:16)

As we mentioned, in the human visual system pathway, you also have a component called LGN, which lies somewhere in between. LGN also has very similar center-surround an on-off structure to the cells in that particular region, which means there are a set of cells while one cell could be inhibitory it may be surrounded by excitatory cells and the vice versa in that same region. So you have combinations of both kinds of cells, which together lead to perception the way we see things.

Originally, the LGN or the lateral geniculate nucleus was considered to be more a relay system that takes the input from the retina and passes it to the visual cortex, but it is now understood to receive a lot of feedback from various parts of the brain, which also come back into the LGN to make it get a more holistic picture of the scene. So there are other feedback that come in to make it get the perception that it actually sees.

(Refer Slide Time: 16:23)

So the visual cortex or the V1 cortex lies at the far end and let us talk about the visual pathway a bit more detail in the next few slides.

(Refer Slide Time: 16:35)

In the visual cortex the V1 cortex, we go back and recall the history of computer vision that we talked about last week, where we said that there were two researchers Hubel and Wiesel, who were the first to characterize the V1 and receptive fields by recording from a cat viewing stimuli from a screen. We also talked about them receiving the Nobel Prize in 1981 for this work.

(Refer Slide Time: 17:02)

And one of their largest contributions was to show that the V1 cortex has two kinds of cells, simple cells that simply detect oriented bars and edges. For example, you can see a bar detector, a bar is simply a white region flanked by two black regions or otherwise, and an edge detector is the edge detector that we already know those are simple cells. While complex cells may be invariant to position, but they are sensitive to orientation.

So if you have certain orientations of edges the complex cells are what pick up those kinds of orientations in their structure.

(Refer Slide Time: 17:46)

The cortical cells actually end up computing derivatives. Remember, again, that spatial derivative is orientation sensitive so you get depending on how you place your filter, you are going to detect different orientations of edges in your image. So, if you had such an edge in your, in the scene that fell on your eyes, the spatial receptor field would look something like this, which is a derivative in space and the derivative in time, as we already said, would peak and then fall off to the other extreme and then gradually go further.

To some extent the spatial derivative and this temporal derivative look similar, but the time derivative or the temporal derivative leads to illusions based on time when we are looking at an image.

(Refer Slide Time: 18:45)
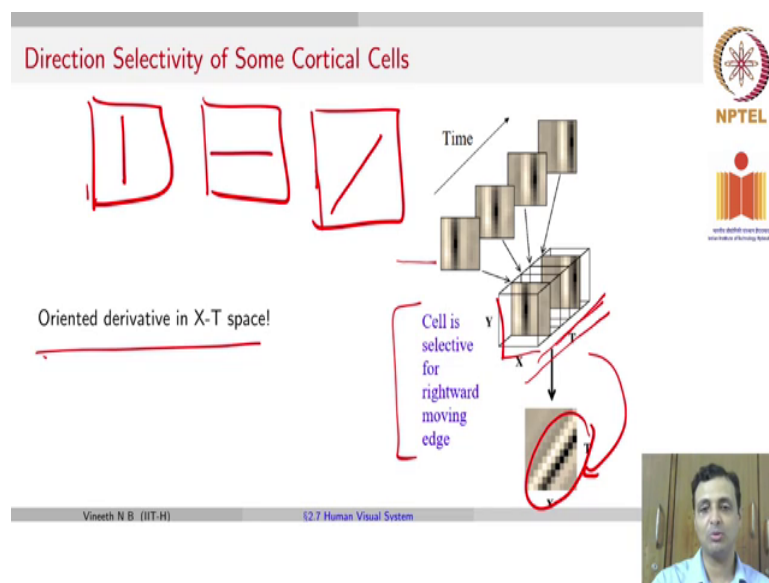
So also, some of these cortical cells have direction selectivity as we said, the complex cells respond to specific orientations, and the oriented derivative can actually be in an XT space rather than just in X space. So for example, with all the edge detectors that we saw so far, we saw that you could have an edge detector that detects a vertical edge, an edge detector that detects a horizontal edge, or an edge detector that detects an edge with a certain orientation. But because the brain is processing information in three dimensions, X, Y, and T, you could also have an edge that is moving.

For example, you could have a vertical edge that is actually moving that is what you see here. So you have a rightward moving edge, but as you keep moving the edge from left to right, you now have a cuboid of space X and Y and time T. And you would notice that because over time, the edges moving from left to right. Remember again, that unlike the simple cases

that we saw, so far with filters and masks, the human visual system is responding to stimuli that change over time. It is not a still image, but a changing image, so the human eye has to adapt to those changes in the image too.

So then it appears that over T, you are going to have an edge in a different direction, because the edge is actually moving from one part of the image to the other part of the image. So in X, T dimension, this particular cortical cell will end up having an edge along this direction. So where T comes from the movement in one direction, an X edge is the edge that it actually is, remember there is a vertical edge so you are going to have change along the X direction, and you will have changed along the T direction because it is also moving edge.

So an oriented derivative now need not be just an XY space which is what we have seen so far, but it can be in XT space YT space, so on and so forth. So, remember that the concept of an oriented edge detector is very different in the human visual system, because of the concept of time.

(Refer Slide Time: 21:13)

Why is oriented filters important? So even from the human visual system perspective, people have shown that, given natural images, and let us say we had to learn independent filters, whose linear combination would best represent natural images, it can be shown that the optimal set of such filters are actually oriented filters and are localized to different regions of the image.

Another way of saying this is a natural image simply becomes a positive response to a filter bank with several orientations, and each of these filters placed at different regions in the

image. This should perhaps, connect you to the discussion that we had with filter banks and Gabor wavelengths and Gabor filters and steerable filters, so on and so forth. So even at that time we mentioned that Gabor filters are known to be little similar to how the visual system, human visual system performs and this should perhaps be the context of why we made that statement.

(Refer Slide Time: 22:20)
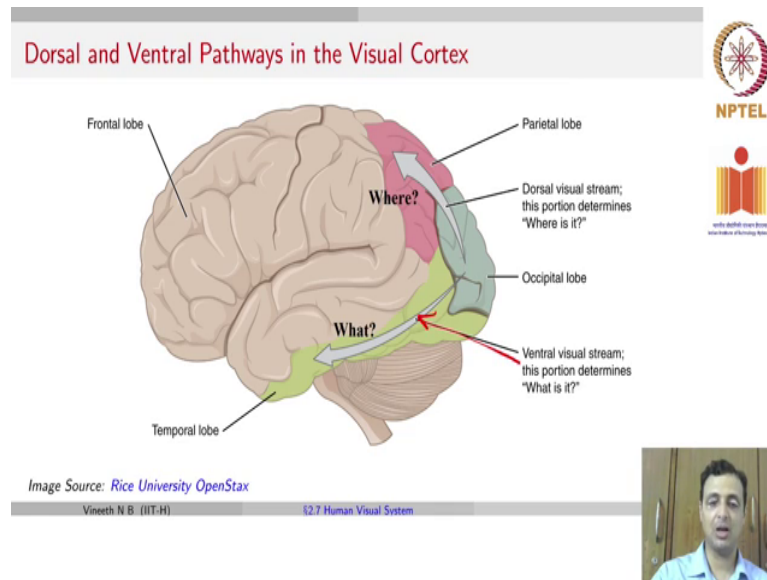
Also, at the visual cortex, the final processing also has two pathways called the dorsal and ventral pathways in the visual cortex. So the dorsal pathway is responsible for where information, so which part of the scene in front of you are you seeing what you are seeing and the ventral pathway is corresponds to the what information or what object are you seeing in front of you. So each of these parts lead to different aspects of perception that we see in the scene around us.

(Refer Slide Time: 23:02)

So the What pathway, so is what you see here. The What pathway goes from the V1 cortex to the V2 cortex to the V4 cortex to a couple of regions called the TEO and TE, we are not going to get into this today, there are going to be references at the end of this lecture if you would like to know more about this, but those are different parts of the brain as you can see here, which finally lead to understanding what the object is.
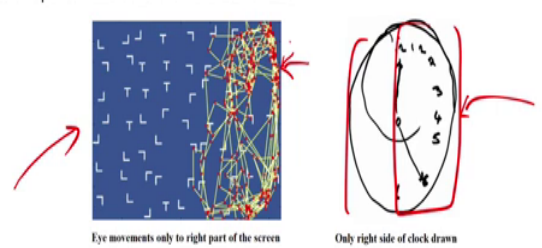
And as you go from each of these regions, as you go from the V1 cortex to the V2 cortex to the V4 cortex to TEO and TE, each region captures higher abstractions of the information around us. Remember again, that if rods and cones and other early processing in the human visual system are only responding to edges and textures there has to be later layers in the human visual system that make us understand the scene around us. Maybe a table, a desk, a wall, a water bottles so on and so forth. So the V4 gets higher levels of abstraction the TEO gets even higher level abstraction. And this is put together as you go deeper and deeper.

(Refer Slide Time: 24:25)

"Where" Pathway

V1 → V2 → MT → MST → Posterior Parietal Cortex
- Cells respond to more and more complex forms of motion and spatial relationships
- Damage to right parietal cortex may result in spatial hemi-neglect - patient behaves as if the left part of the visual world doesn't exist

Eye movements only to right part of the screen

Only right side of clock drawn

Image Source: Scholarpedia - Hemineglect

Vineeth N B (IIT-H) §2.7 Human Visual System

On the other hand, the Where pathway, you go from V1 to V2 two regions called MST, MT and what is known as the posterior parietal cortex. So, these cells respond to more and more complex forms of motion and spatial relationships and that is where the Where pathway comes into the picture, while the What pathway takes different features and puts them at higher levels of abstraction, the Where pathway response to more complex forms of motion and spatial relationships.

So in fact, it is shown that if there is damage to the right parietal cortex, it could lead to a condition called spatial hemi-neglect where a patient which is considered a disability where a patient cannot see one side of themselves all the time.

So, once again that relates to the Where pathway. So if one part of the parietal cortex is damaged, they really cannot see one side of the scene around them, and the patient behaves as if that left field does not does not exist at all. So there have been some experiments that have been conducted, where people have asked, so these are eye movements that have, that were tracked on the screen and you can see that the patient is only focusing on the right part of the screen or in another case where a patient is asked to draw a clock, the patient ends up drawing only the right side of the clock and does not draw a left side of the clock.

These are ways in which this condition is diagnosed and the condition is known as spatial hemi-neglect or hemi-spatial neglect.
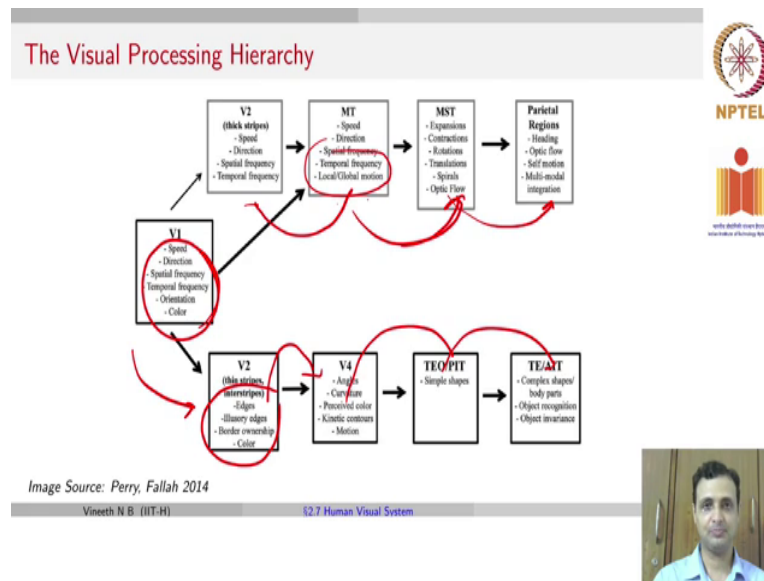
So to summarize the visual processing hierarchy, so you go from the retina to the LGN to the V1 cortex, and from the V1 cortex, there are two pathways the Where pathway and the What pathway, where the, What pathway goes from V1 to V2 to V4 where V1 gives you a certain set of attributes in your image, low level attributes in your image. V2 puts things together and gets things like edges, borders, colors, and so on and so forth. V4 gets angles, curvatures, kinetic contours, motion and so on and so forth and TEO gets simple shapes and TE gets the complex body parts or perceives the world around us as we see it.

The Where pathway, you go from V1 to V2 to MT which detects things like spatial frequency, temporal frequency, local and global motion so on and so forth. MST gets even higher levels of abstraction in terms of movement such as contractions, rotations, translation, optical flow so on and so forth. And finally, you have multimodal integration and a better understanding of the Where pathway and the parietal regions.

(Refer Slide Time: 27:15)

This set was primarily, intended to give you a parallel between what we have been discussing so far and the how the human visual system perceives. If you are further interested there is a nice summary of whatever we discussed so far in the lecture notes of Dr. Aditi Majumder at UCI on Visual Perception. And if you are further interested, there are many more links on the slide, which you can read to understand more and the lectures of Dr. Rajesh Rao from whom these slides were borrowed, is also there as one of these links if you want to read more.

(Refer Slide Time: 27:49)

Here are some references for you to read.