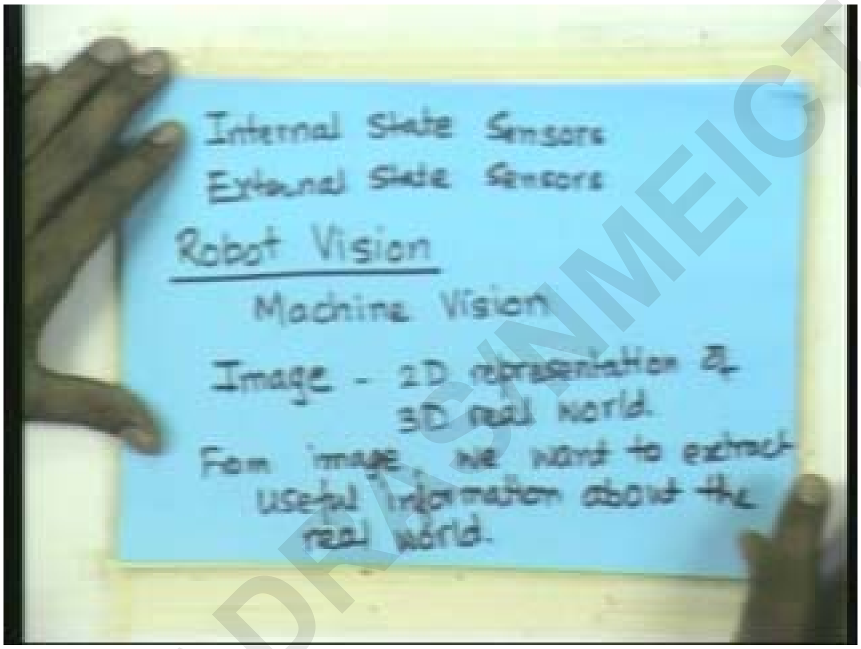
ROBOTICS

Prof. B. Seth
Dept of Mechanical Engineering
IIT Bombay

Lecture No – 25

Image Processing (Time:1:20)

So today I will continue on the issue of robot sensors. Recall that we had covered that topic earlier, and we had looked at the internal state sensors and external state sensors. (refer time : 05:05)



In the internal state sensors, we had covered sensors which were necessary for the basic movement of the robot, whereas in the external state sensors, these sensors augmented the capability of the robot, and we saw a few of those external state sensors like touch sensor, tactile sensor that is, and today, what we going to start out is what is called robot vision, ok?

Robot vision is the most powerful sensor you can think of in terms of augmenting the capability of a robot. This is also, of course, called machine vision because it is not only used for robotics, its also used for industrial vision applications for inspection, for other tasks which are not using robots but may be some other automation tasks, or as assistants to a human operator giving useful information.

Now, basically vision deals with issues of how you take images, and how these images can give you information about the real work. So,

it basically deals with images. And what is an image? An image is basically a reflection of the three-dimensional outside world on to a planar representation,

right? That is what we mean by an image, is that we take the outside world, we take a picture of it, and we have a two-dimensional representation of the three-dimensional world.

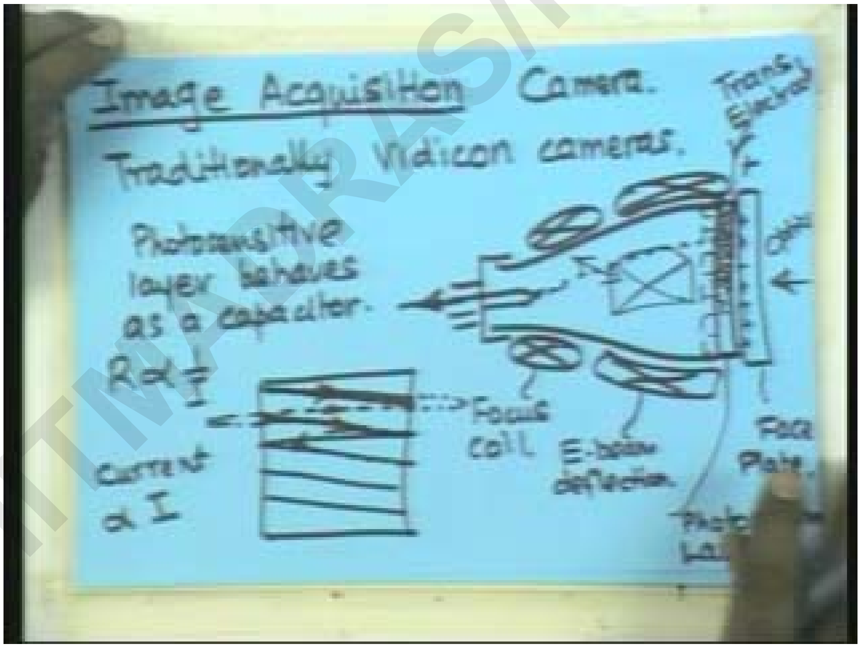
If we take, basically image is at one instant, representation of the work. So you can take images at intervals of time,
so you can have one image or multiple images, and from that image, if you try to extract information about the real world then we will call that image processing.
So, an image as I said, is basically a two-D representation or reflection of three-D real world, ok, and we are interested in taking the image.
From the image, we want to extract useful information about the real world.
Ok. So now if you look at image processing, (refer time : 09:23)

there are some essential steps which are required, and these will generally be classified as low-level tasks, medium-level, and high-level tasks.
So in the low-level task, if we say the first thing is image acquisition, ok, so this is one task. Let me enumerate all the tasks first and then we will look at each of them in more detail. This is, of course, very clear. This is basically sensing of image. It is the sensing part.
Next, once we have a image, then we will do some preprocessing. Now, preprocessing is usually used for reduction in noise, and noise when we say, we are talking about picture noise, you know?
Normally, we associate noise with audio signal. This is in picture. If there are defects, then it could be some due to some noise present in this thing, and of course, enhancement.
Enhancement refers to the fact that if you have a picture which is generally dark, and there is information in the dark portion which will become very difficult to see,
so sometimes we would like to transform the image, so that we are not adding any information, we are not changing the image, we are only transforming image intensity levels so that the dark image starts to look brighter, and features in the dark image become little more discernable. Once we have a proper image, which we have done by preprocessing, then we come to what is called segmentation, ok? Segmentation basically refers to separating different regions,
regions of the image which hopefully are connected with the different objects which are present in the scene.
Then we come to the next step which is description. That is, each of the segments, or the sub-area, or the sub-image that we have, what are the characteristics features that we need to look at so that we can distinguish different regions, ok?
We will describe a particular region by some properties; the property could be shape, could be size, could be general gray level, or brightness, or color, and so on and so forth.
Then we come to next step which is recognition, and recognition is basically referring to the fact that you have different regions available in the image, the image we have already described in terms of the different parts of the image, and then we are going to identify regions as regions belonging to the particular part, for example, ok, or an object. Then finally, you can say we have to interpret. So interpretation. Basically, we want to assign meaning to the scene that is in front of us.
Ok? Obviously, some of these tasks are fairly well defined, some of them are not so well defined, or,
so we would say that first one definitely is part of a low-level task; second one is also predominantly a low-level task. Low-level task is the task in which we don't have to, we

are not using that much of discerning, you know, we are doing something automatically. So image acquisition and preprocessing are going to be low-level tasks, interpretation is certainly a high-level task, where we are trying to assign meaning to what we are seeing: what are the different segments in the image, and what does it mean? Is the part good? Is it defective? Or, where do I hold it as far as the robot is concerned? All these are higher-level tasks, and basically, we will say that segmentation, description, part of recognition, are lower-level or medium-level tasks.

So let's start out by looking at the acquisition – image acquisition – and for acquisition, what do we use? We use a camera. So it will be a video camera. Traditionally, we have used a technology called vidicon, ok? Let's look at how a vidicon camera works. This is the, if you have seen television studios, you will see some big cameras, and those cameras are typically vidicon cameras.

So, a vidicon camera essentially consists of a picture tube with the face plate, and you have a way of generating a stream of electrons. So, this is my source of electrons here. I will have some coils which are going to focus, focus coil, this is what I call face plate, and by applying, suitably, heating up this element, we can generate electrons which are going to be attracted towards the face plate. The path of the electron is basically governed by other coils, so if I have some other coils here, this will be for deflection. So, electron E–beam deflection, ok? Now, there will be coils in the, along the vertical direction, as well as there will be coils along the horizontal direction, (refer time : 20:49)



so you will be able to scan the face plate by using these deflection coils. Now, the critical portion is near the face plate. At the back of the face plate we have a transparent electrode. Transparent electrode. So, it is basically a good conductor of electricity and it is transparent to light. So whatever light is coming through the face plate gets across this electrode.

Next to the electrode we have a deposition of photosensitive material, if they are in the form of small, small globules, ok? So this I will call the photosensitive layer, ok? These are photoresist material, which, when you have photons striking it, its resistance changes, ok?

Next to that we have a fine grid of another electrode which is connected to a positive charge,

oh I am sorry, it is connected to a negative charge.

So basically, the idea is that we are going to slow down the electrons by using this grid so that the electrons strike the back of this photosensitive layer with very low velocity, ok?

What we will do is, we will, this transparent electrode will be assigned a positive charge.

So we assign a positive charge here, what happens is, supposing there is no light coming. This portion here is optics, which means, your lens, etcetera, focusing lens, will be there in the camera, and behind that is all this stuff here.

So when there is no light coming at all, then this positive charge will be distributed. There is a good conductor here, so you will have a positive charge existing here, and we are depositing negative charge here, so negative charge will be existing on this side. This material is basically insulating, when there is no light falling, and this is basically working like a capacitor.

So, the photosensitive layer behaves as a capacitor, right? Now, what happens is the deflection coils are so controlled, that the electron beam scans the entire face of this. Behind the face plate, the pattern, if I look at the face plate now, the electron beam will start out somewhere, it will go through this, then it will return towards the starting point, and then again scan. So it will keep on scanning horizontally, and it will come back turned off. So we are basically having patterns like this existing, basically horizontal lines, and so the charge is deposited, and when there is no light falling, you will, basically it will equalize the two charges will be equal, and it will no longer, no more electrons will flow, because it cannot support any more charge.

Sorry? I have not yet come to that, so I am, you know. This is basically when I am looking from the front or from the back, then I am seeing this portion here, which is a rectangular plate. So this is the rectangular plate and I have electrons coming and striking at these points, meaning, all throughout, as it is being scanned, if I use horizontal deflection coils

Here, I ramp-up the voltage on these, then the electrons are going to move with some uniform velocity across.

It is actually the path of the electron would be that it will go further, then it will be slowed down, reversed in direction by changing the voltage, and then it will be, the electron beam will be turned off, it will come across, it will go beyond here, start accelerating when it achieves uniform velocity, then it will start to, it will be this part of the scan.

Yeah, it is, if I look from the front here. This is looking from the side, this is the side view. They are not coming outside the tube, they are remaining inside the tube. They are turning because of the magnetic field of the deflection coils,

Alright? So we now have a photosensitive layer behaving a capacitor. When the light falls on it, then the resistance of these globules changes.

Resistance, basically, is proportional to 1 over the intensity, so if the light is intense, then resistance decreases a lot, these positive and negative charges equalize, ok, there is a current flow takes place here in the neutralized, the positive and negative charges. Next time when the beams come to that particular spot, now there is a deficiency of electrons on that particular point, so the electrons will beam will deposit more electrons there,

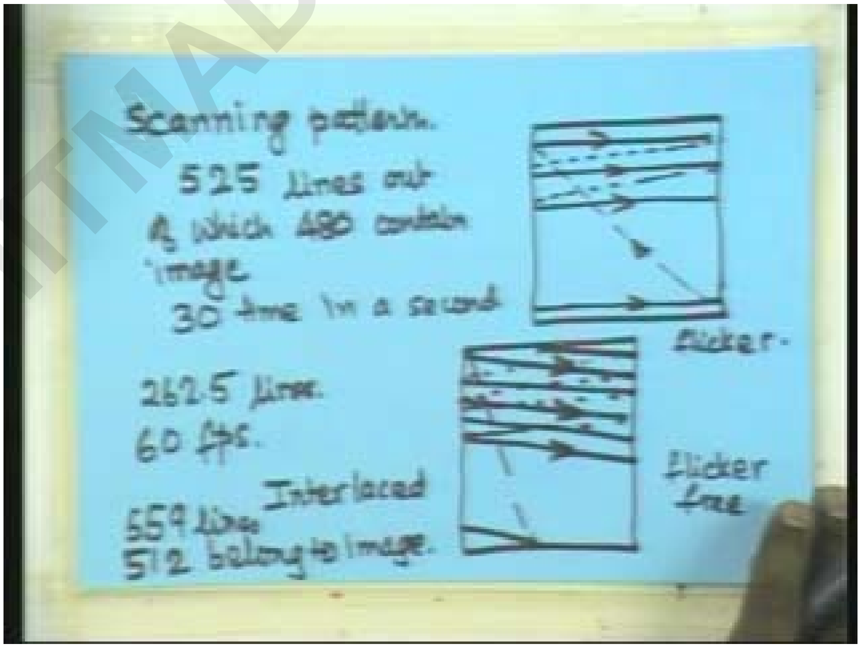
Ok? When that is done, there is basically a net flow current that will take place from this electrode to the point where the beam is being generated, and if we look at this current, this will give us an indication of the intensity of light for any particular given instant of time corresponding to that particular position on the image. So what we will now get is, we basically get a current which is proportional to the intensity, and this is what basically we want, we can generate a pattern of analog values of this current and send it out as a video signal.

Light is ending up here at this point, yeah, it is not turning 90 degrees, ok. Basically you have … Ok, this is your face plate, ok? This is the electron gun; so the electrons are injected into this; so they are going to strike here. Now, the direction of the beam is being controlled, so, instead of falling here it will start to fall here and here, then it will slow down. It will be turned off; it will come back here, and then again it will start and follow this. So the beam is moving in the direction perpendicular to the plane of the paper here.

No no no, the single beam, which is basically, you know, deflection electronics is for that only. It basically scans through the entire area by horizontal lines, and each time it is shifted down by a small amount, so you get a set of number of lines, and each line, the current flowing will give us the intensity corresponding to that particular point,

Right? The electronic deflection, then, you know you know the CRO and all that principle, it is basically similar. So there you are generating image, then you are here to capture the image,

Ok. Now, so what I talked about was what we get. Now, if I look at the face plate, so I have the beam traveling here; return part of the beam I am showing as a dotted line when it travels here again; returns; travels; ok? And finally it will come up here, it will travel here, last line, and then the vertical deflection control will take the beam, will turn it off, and bring it to the beginning of the scanning pattern again. (refer time : 25:53)



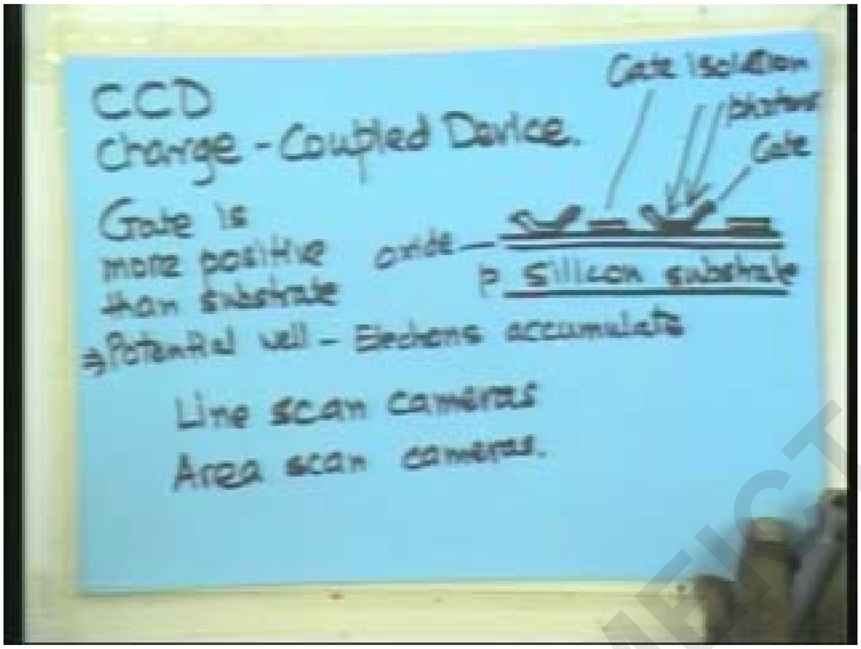
So, this is the scanning pattern typically, for US television signal. What they do is they use 525 lines, such lines, they are done 525 times, out of which 480 contain image, ok? Rest are lost because of the blanking circuitry which is necessary for the beam to travel back and so on and so forth.

Now, this whole exercise done 30 times in a second. So you get, if I call each a frame, ok, one scanning pattern and I'm calling it a frame, then there are 30 frames in a second. So, if you do 30 times, then you get more or less a continuous type of visual effect for the tv viewing. Of course, there is a problem that by the time the beam comes back to the beginning again, 1/30th of a second has elapsed, and then something has moved, so there is a problem of flicker in this type of arrangement. So a slightly different arrangement is also used in which, instead of having a single pattern, what we do is, instead of having 525 lines you will divide that by 2, you have 262.5 lines. And you do that 60 times second, so 60 fps.

And the pattern would be like this; that you will start out the beam here, then it will come back, then the next beam, it comes back, ok? This is done, it will end up at this point, let's say, it ends up here, then instead of pushing it back here to the beginning, it will be pushed back half a line away. So then, the scanning pattern will begin here, and return beam will bring it here, then scanning them the beam will be here. So we are basically filling up, you can say, the missing lines, so the whole image is created once in this set of lines, and other half line away from that, so that removes the flicker. So that is flicker-free, ok? This is called interlaced image. Because, well, you have to scan the whole area.

One could do that, I am sure, but that is not done. Typically, you have the electronics for bringing the beam back, turning the beam off at that time, and then taking it from the bottom to the top. So it is possible. I mean, if somebody wants to design a new camera, then you can start scanning backwards, upwards, and so on and so forth. Everything is possible, but these are the standards.

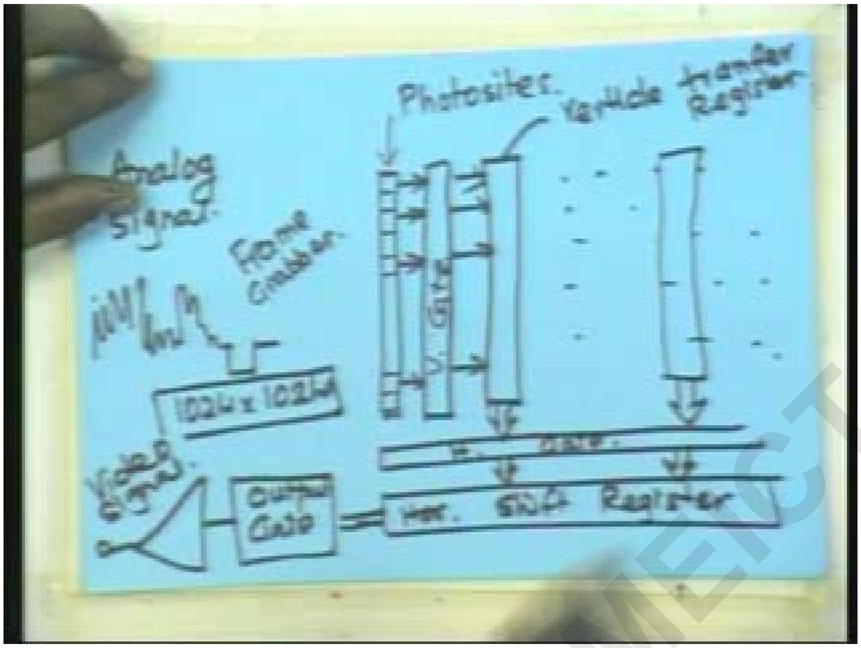
Typically, you would like for digital images, you would like to have, instead of 525 lines, one uses 559 lines, of which 512 belong to image, and this is useful because it is a power of 2, and storing the data, etcetera, becomes a little easier in digital images. One can have, this is 2 to the power of 9, value you can have 1024, or, we will come to that. (refer time : 26:05)



Ok, so now let us look at another technology that is used for image acquisition. In fact, that is what is more used, which is called CCD. This is Charge-Coupled Device, ok? The basic structure is it’s a solid state device, so you have silicon substrate, and basically use p type, on which you have insulating oxide layer, and then you form gate structures on this, and you have isolators, so you have, so the basic idea is that if a light, you have photons coming through this gate, and the gate is positively charged compared to the substrate, ok, if gate, so this is, first of all, this is gate, and this is gate isolation, ok, these are photons. So, when gate is more positive than substrate, then there is a potential well created, here this leads to a potential well where electrons will get stored, ok? So electrons accumulate, ok? So what is basically done is, every other gate, you make a matrix of it. Every other gate has a photo diode, so when the light falls on that, electrons are generated. They get strapped in the potential well, and then the gate voltages are such manipulated that the charge of the electrons which are deposited there, which are stored there, get transferred to the one which doesn’t have the photosensitive part.

So, it can be transferred from one side to another side by manipulating the voltages of these gates, Ok? Now, once that is done, you can shift it out on to a shift register, then the shift register you can manipulate and get a signal out from this. So you can have both types, CCD, and you can have line-scan cameras, and you have area-scan cameras,

Ok? Line-scan cameras are very common in, let’s say, Xerox machines. You have you will look at one line at a time, and then you scan through by moving the carriage which basically scans the whole of the page. It is basically used for that type, either inspection type of systems, or photocopying type of systems. Area scan is used for video images, and the area scan is basically, you will replicate the line-scan technology in the other direction also. So, basic structure is something like this. You have a photo site which will be like for a line-scan camera. You have, these are photosites, then you have a gate structure which can be used to transfer the electrons from each of these photosites by controlling the gate. So we will call this the vertical gate, ok? From the vertical gate you put it in a vertical this is transfer register, (refer time : 31:03)



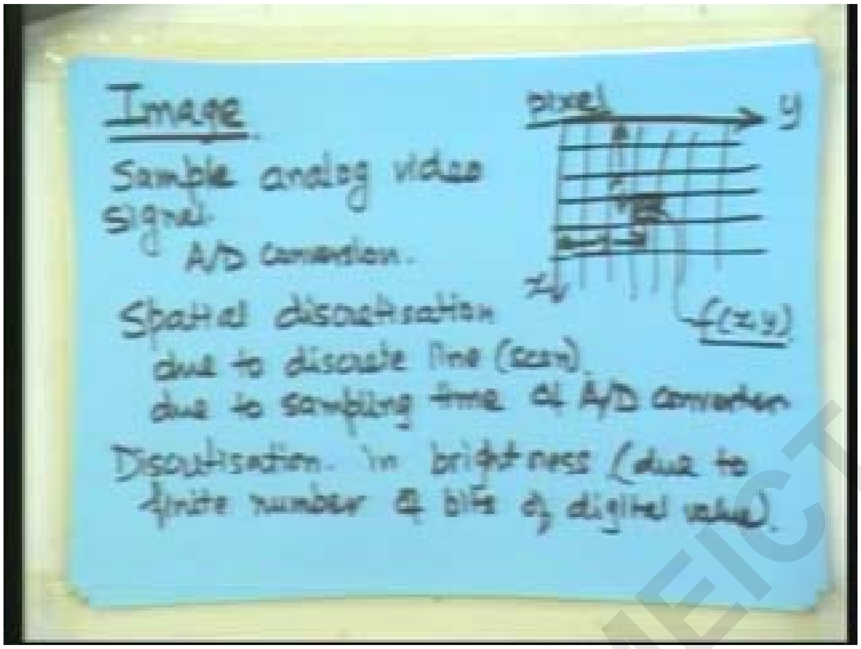
Ok? So this can be transferred to that. Now, from this whole shift register you transfer it through another set of gates, so these things are repeated here, what I've shown here is repeated along here, and each of these in turn will be, the voltages are again controlled, so that they go through the horizontal gate into the horizontal shift register, ok? This will transfer at this one time, then next one will transfer; so, in sequence, they can be taken out from here, through the output gate, and then suitable amplifier, etcetera, will be there, and your video signal will be their out here.

So this is the basic technology. There is, of course, a very simplified way of showing it, but um basically, you generate photons create electron accumulation and that whatever electrons are accumulated depends on the intensity of the image at that particular point. Then these are sequentially transferred by using proper electronics of the gates, two shift registers where they are sequentially transferred. All the vertical ones are transferred out to the horizontal ones, one after another, until you have a composite video signal.

Yeah, so each of them are sequentially put, and you can generate signal, So it is basically analog signal in both of these cases, ok? So typically, it will look like, whatever you know, depending on the image, and then you will have, after each frame there is some blanking time, and then the whole thing repeats, and this has to be interpreted with each of these representing a line. And then it will have to be, if you want to generate an image frame or a digital image matrix, then one will have to sample these and take out digital values and store them away, ok? That is the task of what is called a frame-grabber card.

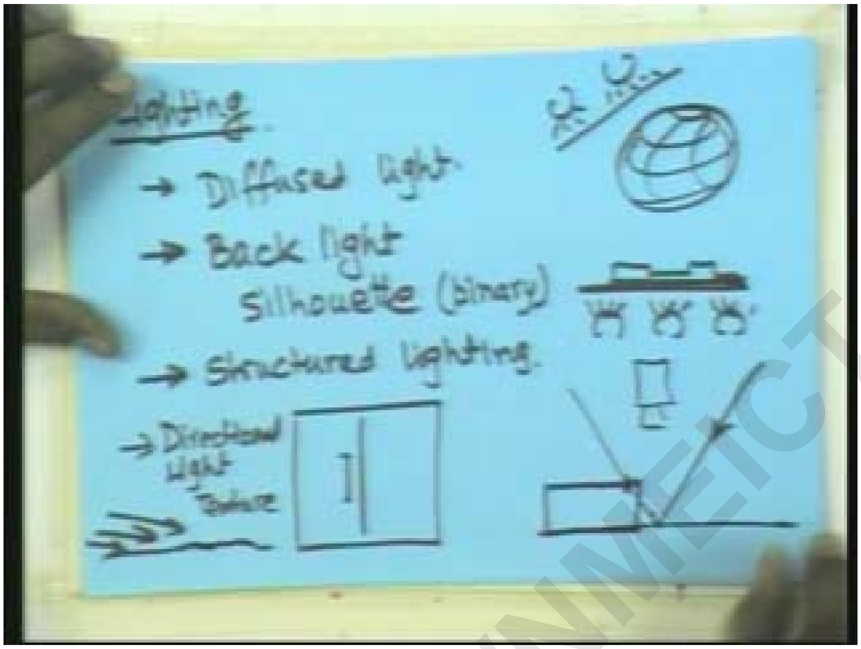
Ok? So typically, you have resolutions, used to be low in these cameras, but nowadays you can get pretty good resolution even in CCD. Earlier, vidicon was preferred because you could get resolutions of thousands of pixels,

whereas CCDs were restricted to 32 × 32 matrix, or 64 × 64 matrix, at most, maybe 256 × 256, but nowadays you can get easily 1024 × 1024 type of image from CCD. (refer time : 38:22)



Ok. Let's move on now, and let's see what an image then is, ok? Essentially now, you have from the composite signal, which is the audio signal, sorry, analog signal, you are going to sample it, video signal, ok, at some rate using the frame-grabber card, so each of the scan lines obviously is discretized, because you have only so many discrete lines.
On each of the lines, the analog voltage which is there is going to be sampled, and that will give rise to discrete digital values depending on the accuracy of the conversion process. So you require A to D conversion, right? And then, so you have then, spatial discretization,
first of all due to discrete lines, lines of the scan, and due to sampling time, sampling time of A to D converter,
ok? You also have discretization in the intensity values in grayness or brightness, whatever you want to call it, ok? This is due to finite number of bits of digital value,
ok? So now, once you convert this into a digital thing, so you have basically divided your, if this is my x and this is my y, when I have divided my x and y into discrete intervals, and each of these little squares will be having some intensity value, and that intensity value let's say x y, where it is x distance away here, and y distance away here,
ok? So this is basically called a picture element or a pixel, ok? Picture element is called a pixel, for short. Now, this pixel, depending on the kind of image that you have, if you have a color image, then you will have maybe several bytes of information. If you think of RGBS, three separate bytes, each having, you know 0 to 255 intensity values, then you can make so many million colors,
ok? Or, if that is too much memory, then you could restrict yourself to gray-levels image, which will be 1 byte image per pixel, or if your information is only that you require whether the part is there or not, theere, you could deal with a binary image. A binary image has pixels which are either 1 or 0, binary value, black or white. So that is depending on what levels of gray you have, you will have different looking pictures for that.

This is discretization and brightness, which is represented by this function, f x y. (refer time: 44: 15)



Ok, then, we will look at lighting which is a very important part of image acquisition, ok, unless you have proper lighting for the image, the image which will come out to be poor quality, then it becomes very difficult to process it.

So typically, you use different type of lightings. One is diffused lighting, ok? In diffused light, if you have some object, let's say its with some continuous curves and all that, so I have object, then I diffuse the light so I don't use the light source directly, I may have a ground glass here, and then I have light sources behind it, so that what I get here is a diffused light, then I will see continuous change in intensity because of the surface features, or the surface curves, which are existing here, and that will be useful for looking at objects like that when we want details of this or the overall surface of the curves of the surface.

Another useful type of lighting is what is called back-lighting, ok? In back-light, you will put, you, let's say you have source of light here, underneath, and then you have a ground glass semitransparent type of thing, and then you have a part line here, whatever the part may be looking like.

So, if I look at it from the top now, I will see a silhouette, ok, which means, wherever the object is there, I will see darkness, and wherever object is not there I will see bright, so it directly gives me a binary image.

The advantage of the binary image obviously is that you don't have to deal with so much information. The processing of the information becomes faster,

but then you lose the features which are there on the surface. You have either bright, I mean, 1 or the 0 there, so that is the only limitation that you have if you are only looking at the contours on the outside of the shape, and that is sufficient for you to know what the object is like, if you are dealing with basically flat objects of different types, then this will be a good technique to use, ok? So, there is one type of lighting called structured lighting, ok? In this, what is done is, instead of having a general diffused light, you have patterns of light which are generated. For example, if I have a sheet of light, now I am looking from one side, I have a sheet of light which is going into the paper, and there is another sheet of light, both of them coming at one line, basically, so this is my surface here, and if I have a camera looking down here, then the image that I will see will, if this is the image that I am representing which is seen in the camera you will see a bright light here,

Ok? Now, supposing we have an object which is, this may be a conveyor belt, and the object is moving on it comes and hits one of the lights, ok? Immediately, what will happen is, you will see that, depending on the width of the object, so if the object is only partly wide, you will see another light which will be corresponding to the reflection of a light beam from this point of the object.

So, you have this beam which is being reflected and seen as one big line, here across, and then you have a smaller line which is seen as the width of the object is the length of this line, and it is displaced from there, because now it is hitting at a different point.

So, if I wanted to detect an object in such conditions, then it will be very easy to detect; it will be easy to process this particular information, and therefore, structure lighting is very useful for simply find the task for image processing which we will see, is not so trivial,

Ok? They can be different structured lights. This is just one example I have shown. You have parallel sets of planes of light which will give you a stripe pattern on the surface of the object, and then you can discern the shape and orientation of the objects by looking at what kind of pattern is resulting. Lastly, I would say that you can have directional light, and this is basically if you have, if this is my object or surface, and I am looking at, I can have light falling at a very oblique angle, so what happens is all the features, the surface features of the object gets highlighted.

You will have micro shadows falling for the little ridges which may be there, and whatever features are there. So you can discern texture using this type of light,

Ok? So depending on the lighting, you will get different type of image, and it is very important to have the correct type of light for a given situation using image processing,

Ok? Having basically seen how the sensing part is done using a vidicon camera, or CCD camera, I would like to contrast this now with human vision, so that I would like to bring out the issues to you, what is important in terms of actually processing the images.

So if I look at the human eye, basically you have the front portion which is the cornea, ok, this, and then you have basically the eye which is spherical in shape, with different layers inside the eyeball; you have the retina ,which is the sensitive part, and you have all these things between the retina, this is the retina, ok, and between the retina and what is called the sclera, you have another layer which carries the blood to the eye and supplies it with the blood that is necessary to keep these cells alive; and then you have the iris and the lens; and the lens is supported on fibers, here, which are basically muscular; and also, the retina is connected to muscles, here, which are called ciliar muscles, ok? This part here is very very important. This is called the fovea; and you have the optic nerve going to the brain, (refer time : 46:38)

Ok? The optical axis passes through the center of the lens and on to the fovea. Now, the human eye is remarkable. You know this is, you will have basically a diameter of 2 centimeters, and the number of sensitive parts on the retina, including the fovea, are in millions and millions, ok?

So, we are talking about sensitivity of the camera. If you think about an image of 1000 × 1000, right, you have one million pixel elements, and if you have to do this, let's say, 30 times second as the frame rate, then you know that you have one over 13 millionth of a second available to process each of the pixels; process meaning, there is an analog value, there is an analog signal, which have to be converted into digital values. So that is the A to D conversion. The analog to digital conversion has to take place in 1/13 millionth of a second, and that is what basically you can see that number of pixels that we can deal with gets severely limited by that.

How much, because the amount of information that is there is truly remarkable.

If you look at the human eye, you have 6 to 7 million cone, what is called cone receptors. These are primarily limited in the area of the fovea, ok? Fovea itself is carrying something like 5 million cone receptors.

Each of these receptors are individually connected to an end of a nerve, ok? So you have all these millions of connections and these are good for bright light vision. So the remark is it is the fantastic kind of a resolution that we are talking about and this is only an area which is less than a centimeter, much less than a centimeter, few millimeters in size.

So, our, the way we look at the things, is that whatever we want to see, we bring it to this particular fovea by rotating the ball, eyeball, and bringing whatever we are interested in into the center part of the fovea part, so that we can discern the details there. How tear[???49:27] we have? Some cones, but we have mostly what are called rod receptors,

Ok? In most of the retina. And there are something like 75 to 150 million rod receptors. [???49:52] much much larger than any camera receptor. These are not good for colour vision. They basically work with dim light, ok? And they basically give a gray level, gray type of image.

Well, we, no, there are some cone receptors, but there are no rod receptors in the fovea region. Idea. That is the key word. It is the image which is constructed in your brain, not necessarily what you see, because I know Avinash is sitting here, I am looking at that person, I see a red shirt there, I don't see his face clearly at all, but I know it is Avinash, right? So that is the key in human image processing, the processors are very very complex. You have the whole experience, you have your memory which aids you.

Yeah, in the daylight, it will be, when the light is sufficient, it will be mostly cone receptors. So, if you see even in daytime you look at a picture which is, you know bright and colorful, you can see it, if you put that in moonlight.

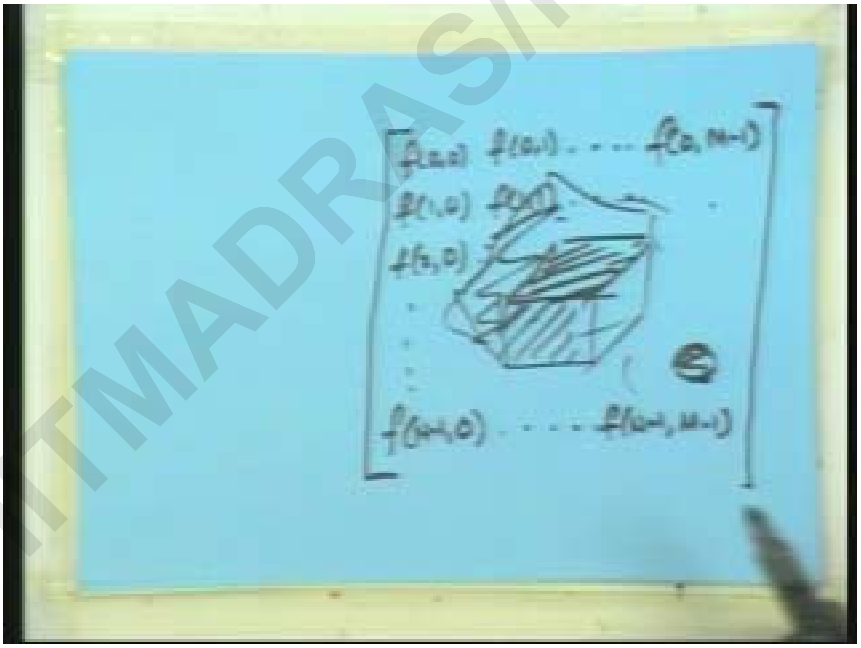
Yeah, then you will see the details of what is happening at that particular point, at the instant, right, but the image doesn't vanish away. Even if I close my eye for a moment, I can continue to see what I was seeing earlier.

So the brain keeps the image alive, ok, of course, that can also be used to fool the brain, ok, so you can have various optical illusions, and so on and so forth, which are basically reflecting in the complexity of the way the brain interprets what is seen by the eye. So sometimes, what is not there you will start seeing it, because of the way the brain functions, right, but by and large, it helps us to see huge amount of detail without as much processing as you will imagine will be required with these kind of number of pixel elements, talking in terms of hundreds of millions of pixels, you know, and we are doing all the time real time and we look at one glance and we know, we immediately have interpreted what is there,

Ok? That's why the danger is that we do not appreciate what is required in machine vision until you start to do things yourself and try to interpret, and even try to determine simple type of things that whether such and such type of part is present or not, or where it is present. Then you run into a huge problem because it is not easy, ok, and that is it is not easy and it feels easy because we are so used to it,

Because, if you imagine that you basically have a matrix of numbers, right, so you have intensity level at this particular pixel; you have intensity level at this particular pixel; and so on and so forth. Supposing I have m pixels along this direction, then you have some other intensity here, and so on and so forth. If I have n pixels in this direction, then I have, so many intensities are involved here, right? This is basically my image frame, digital image frame.

Now, each of these is a number. Finally, how do I know that there is actually a block sitting here, right? This picture is, finally it boils down to, in the computer, as a set of numbers in a matrix like this. They may be values from 0 to 255; they may be values, three values from 0 to 255, or they may be a single value 0 or 1, ok? A bit of information. Whatever it is. Finally, there are, these are the presenting somethings in the real world, (refer time : 56:55)



Then, that is why all these steps are required, that you need to enhance the image, that you need to remove the noise so that we are looking at basically, what is, should be represented, for the outside world, then we are, later on, trying to segment the image to see, ok, this part is different than the rest of it here, around here,

Ok, or something here is different than the surroundings, so we will segment it into different regions. Then we will try to say, ok, this region must be something, or what kind of region is it; if you have color information, and color provides a useful set of information that you can many times distinguish this blue sheet from the white surroundings, because, the property, blue, ok? So that is the description part that we describe a particular region with some attributes; it may be color, it may be size, it may be shape, it maybe variety of things. It could be like, how do we distinguish a letter from an object, or an alphabet from an object, because we look at text when we will use some properties which are going to distinguish it from other objects.

So, once we have done that, after that we will try to recognize and say that, ok, this is part of a top of the box, because this is green in colour, ok; this is the side of the box because it is red in color. These are simple ways of doing it and when you don't have such clear-cut color information available, then things are, you have to determine based on the intensity values, and not only intensity values, how the intensity values are changing suddenly. So, if you have the object here, which is[???56:23], and you have a light source, so this space is bright, but here you have a shadow coming which will of course look like a different region, ok, but there is no object corresponding to that, it is only the shadow of the object,

so we have to distinguish it away. We have to not use that. You have recognize that as a shadow. And so on and so forth. So, the task is not very straightforward. It is what I am trying to point out, and we will appreciate that as we go along, and look at different parts or tasks involved in image processing, which whould be required for that robot to use that information in some sensible way.

Typical way the robot will be required to use the vision information is that you have a scene in which you may have number of objects coming. You may have nuts coming, bolts coming, washers coming on a conveyor belt in some random order. Then we have to, we have only few of these given objects and we have to identify which is which, so which is one type of task, and then having identified it in the right order, the robot can take the gripper to that particular location, take it up and use it in the assembly, let's say.

The other type of task will be, the robot is required to do some continuous path motion, say in spray painting, could be the one, but spray painting you usually don't use vision, you will use vision in, say, welding application,

So you are seam tracking, you want to know where seam is, so that you could move the torch or the electrode for welding along that particular seam. Then you would like to use vision to see how the weldpool looks; am I moving too fast; am I too slow; the temperature is too high; too low; all these kinds of information you have to derive from a vision-based system. Or you may have to try to duplicate human beings too. You may have a bin full of parts in different random orientations, and you have a robot with vision system. You want to see, ok, this particular part is lying in this orientation, this is on the top of the bin,

Right? I want to know both these informations, so that I could go and pick it up.

So the variety of tasks, basically, vision will possibly give capability to the robot to duplicate some of the ways, some of the tasks, which humans do quite easily. So we will stop here.