Lecture – 13-2
Tracking Systems (tracking with a camera)

And, if you want to now figure out, what is the displacement? You know notice that, if I if I were to let us say keep my collarbone fix and I just start looking around, then my pupils are clearly translating around the space right. Even though, I am just rotating my head. So, if I estimate the rotation of your head I should be able to additionally estimate how much your pupils have been transformed? Assuming that you kept your torso fix say your collarbone was screwed down.

Let us suppose and I should be able to estimate you know make a reasonable average human being model of, how much translation occurred? And, that will get you pretty far that is better than just assuming that your eyeballs are individually just rotating in place right really it s your head that is turning. So, so, this is a very helpful approximation try to compensate for some of the position change, but if the user decides to go back and forth like this and you might remember one of the motion cues called parallax, where you see nearby objects moving at a faster rate than faraway objects, if you would like to do these kinds of motions and it does not work for you, because you are not tracking the position back and forth then it ends up being uncomfortable.

So, that causes motivation to bring in more sensors right. So, if you want to estimate position; what more do we need is a question? In addition, to orientation again you can fake some of the position.
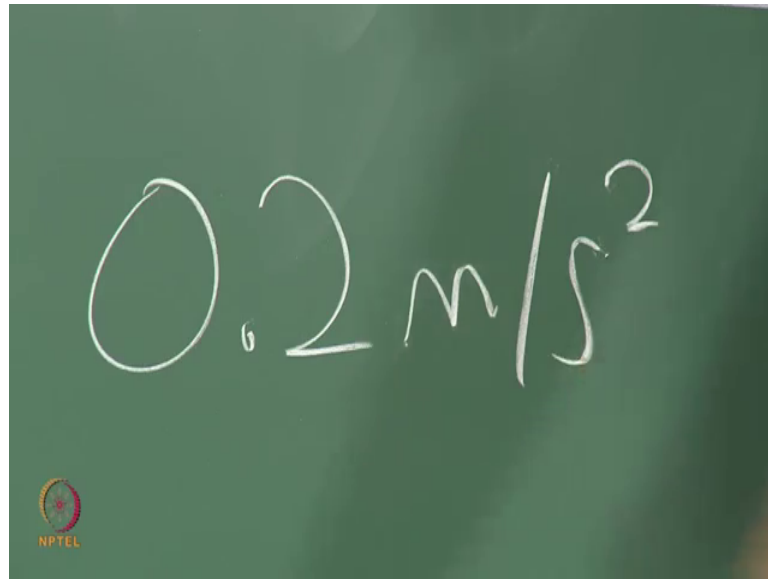
You might look at this is the greatest hopeful, if you have just study it purely in textbooks you may say well I figured out how to separate up from linear acceleration of the body, while I just take this linear acceleration component instead of throwing it away, why? I just integrate it once I get velocity integrate it twice and I get position. So, I can just figure out what your position is from that right any difficulties with that; let me explain some difficulties with that and this will you know motivate why we might one might want to use cameras and other more sophisticated sensors so.

If we just take this residual acceleration here left over after we got rid of the up compound at first of all we might have separated these two incorrectly another thing that might happen is remember that we are always trying to estimate the orientation. So, if we have some errors in the estimation of the orientation, then we will not know exactly which way up is when we do the subtraction of up to get this residual component. So, that will lead to some additional error that does not look like noise it just looks like an offset, because we do not know truly which way up is we cannot measure the we cannot estimate the global frame perfectly. So, that is going to cause additional errors.

And, now suppose you have a small amount of error say it is you know some kind of bias like maybe 0.2.

Meters per second squared you know. So, amount that is that is left over here and now you integrate that one to integrate it twice and you take a look at how far your estimate is from reality after 5 or 10 seconds, it turns out this grows very fast. You may remember that when you integrate a constant once you get a linear rate and if, what happens you get a linear drift rate that looks very much like the gyro drift rate, but if you integrate twice to get up to position your two derivatives away right are two integrals away from the position, then it ends up with a fast quadratic growth rate you know as a simple offset like this a simple fixed air will help will cause a quadratic growth rate.

So, it is very much as if your head has a fictitious rocket thruster attached to it that is thrusting at this rate and it makes you shoot away and; so, that is what gets very very bad about this and if you think you can compensate for this perfectly then there is still the problem of double integrating noise which will also be a very fast rate, but this quadratic term tends to dominate, because usually we have a very simple kind of bias or offset here, because we cannot accurately estimate the orientation and we cannot perfectly calibrate the accelerometer.

So, these are the difficulties we get into and so, we did a lot of experiments trying to do whatever we could also adding kinematic constraints based on the human body trying to rule out absurd possibilities that would cause a faraway drift, because we know that your heads not detaching and floating away into space; however, it does not take long like

maybe a half a second or less for a very large amount of drift error to occur and then you cannot correct that very naturally you have to detect it and correct it without making the person nauseated.

So, it ends up being very very difficult. I think essentially impossible with this generation of sensing technology and for the reasons; I have explained there are some fundamental limitations no matter what there is going to be some kind of bad drift unless you have some additional global reference for position like a camera right. So, that is what we are going to get to next. So, that leads us to estimating position and orientation.

(Refer Slide Time: 04:51)
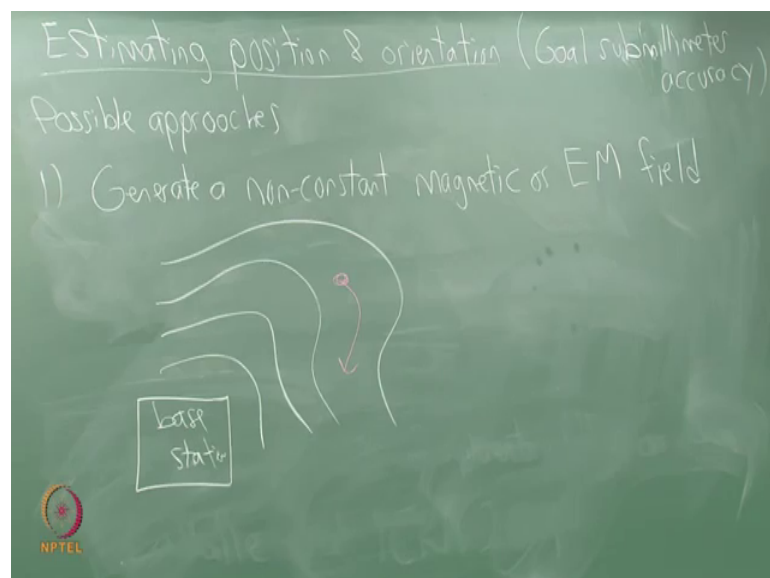
So, estimating position and orientation. So, in other words the full 6 degrees of freedom, now this acceleration information may very well be useful this linear acceleration information we can use this in a system to perform position and orientation tracking; however, we also want to have an extra signal that provides positional drift correction all right.

Couple of solutions will possible general approaches one is you could generate a non constant magnetic or electromagnetic field. In other words, you may want to have some field lines that are intentionally bending right or maybe you want the amplitude of the field to be intentionally changing in some way and you have engineered a technology based on some base stations that you put around.

So, you may put a base station and there is some kind of field lines that you are generating from that base station and based on that you may have a second base station as well; if this does not provide enough information. So, based on the information coming from one or more base stations they could use ultra wideband radio as an em example or you could be generating magnetic fields. So, I am based on the information obtained; when you place a sensor in here that is moving around and designed exactly to sense your non constant vector fields, then it can recover the position.

It is still with this type of technology is still very hard to get a high degree of accuracy it is very easy to walk magnetic fields by moving metal through the field other kinds of problems like that and the radio based message I think still can knock it down the level of accuracy that we would like which I should point out the goal is sub millimeter accuracy sub millimeter accuracy, why is the goal sub millimeter accuracy?

Well, if you have very large deviations it is quite perceptible you know if I put on a virtual reality headset; and I start seeing the world shaking and jittering on becomes very perceptible. So, what might be a reasonable tracking solution for looking at your hands or just a tracking viewpoint from a scene, but you are watching it on a screen maybe up to one centimeter of air is fine, but we need to be sub millimeter level of accuracy becomes very clear once you put on a virtual reality headset and you look at tracking errors that are a few millimeters or greater you start to see the problems with it.

So, that is one thing you need to point out we need to be extremely accurate. So, this is one technology that has some hope and there is there certainly several companies developing methods and products in this arena and there have been for many years the other category which I will spend more time.

(Refer Slide Time: 08:21)
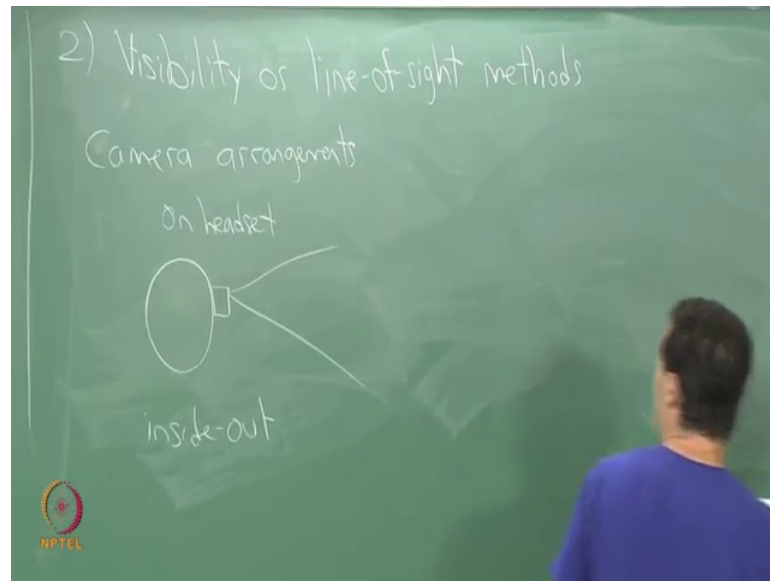
On is visibility or another name for is line of sight methods and one of the; most common implementations is based on cameras. So, I can talk about various camera arrangements and, I believe I mentioned this last time you could have human head wearing a headset and the camera is looking outward. So, this is where the camera is on the headset and this is often referred to in the industry is inside out tracking.

(Refer Slide Time: 09:20)
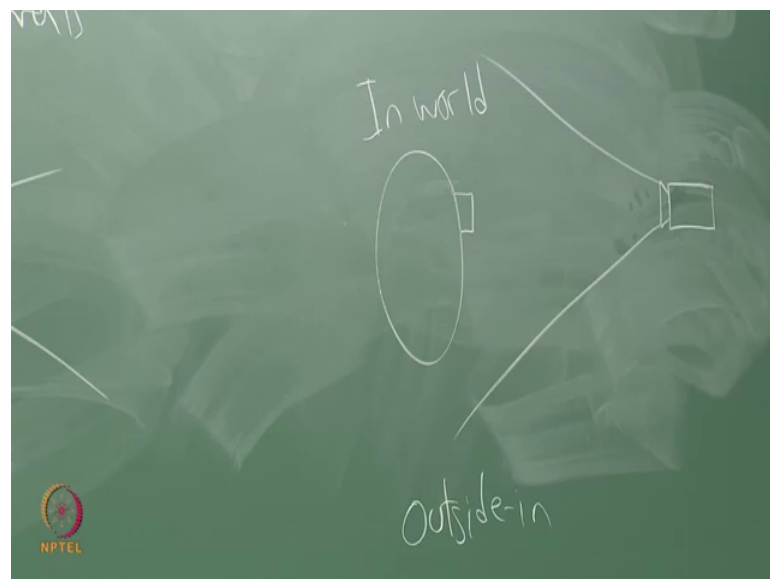
And the other case I have the person again wearing the headset and you have the camera out here this is what you have on the lab on the oculus rift dk twos, and the camera is

facing the user and this is referred to as outside in or I would put over here in world camera. So, either the cameras on the headset are the cameras in the world questions about that so far.

Student: what is the paint of marker that is we using is it an actual marker or (Refer Time: 10:03) marker.

That is very good question that is exactly, what I am covering? Next, that is that is the thing to be asking what is the camera are going to be looking at right yeah. So, let me go over that, what does the camera need to see for this to work, we need to have line of sight or visibility towards something to make these methods work it also helps to have an accurate model of how the camera projection works as well right. How does the camera transform points that are in the world into image points? So, that is the first thing, I will I will mention before completely answering your question is we have a let us just make a simple pinhole camera.

(Refer Slide Time: 10:45)
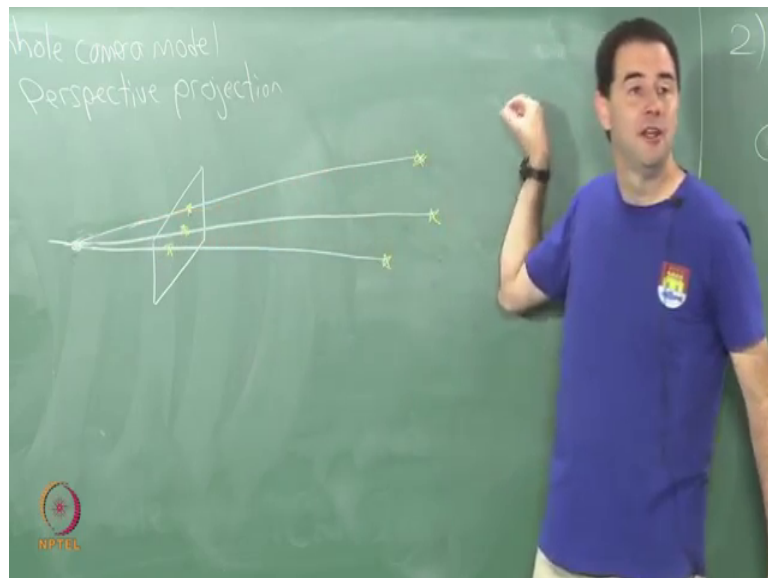
Model which uses perspective projection.

So, I have some features out in space we will just call them features markers features whatever you like out in space you draw three of them here and then imagine I have a camera that is pointed at these. So, I have an image plane for the camera like this, and then based on some projection model I have all of these features treated as rays that all

meet at a point somewhere using a perspective projection model and so, on. All these two come together at some particular point.

And they pierce the image plane somewhere I can get the perspective right here probably not. So, I have some simple model of the imaging that looks like this right. So, I imagine there is an image plane and you know the particular coordinates of the points in this image plane it might depend on where I stick this plane and if I move the plane all the way past this focal point of the lines here then the image will turn upside down.

So, let us suppose I have some model of the camera parameters. So, that I know where this plane is at. So, that I can then understand based on this focal point here and where the points are, but for each one of these points that I see in the image each one of these feature points I just see in the image I know there is a corresponding ray in 3 dimensional space ok; that is the thing that is the assumption that I am going to end up using right. So, I see features in the image plane I am going to know that they lie on some particular ray based on, how I have calibrated and set up the camera right.

So, that is the model and that is, the information I am going to be using; what exactly are these yellow stars? I have drawn what I want to talk about next which I believe was the question that was just asked so.

(Refer Slide Time: 13:25)
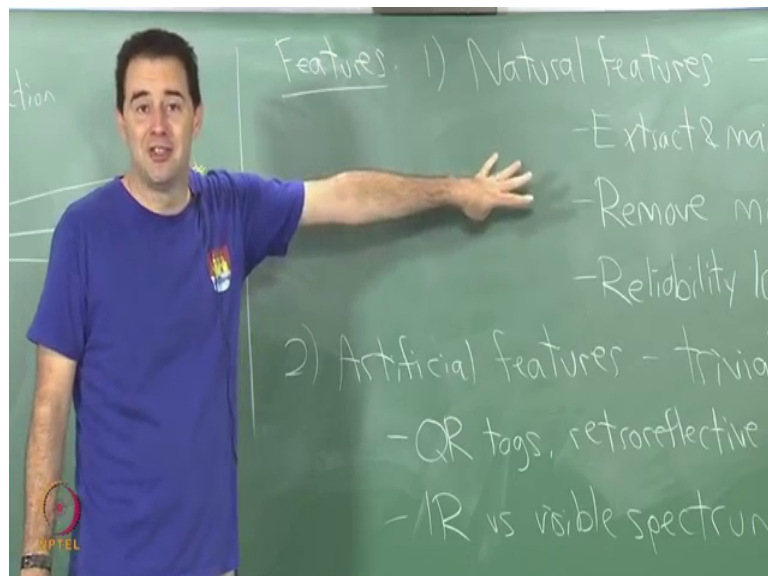
Feature and what are our choices? Well we could use natural features right and this is what I would call? Let us say hard computer vision. So, imagine for example, you have the inside out camera you have two camera on the headset; it just looks for whatever features that can find and works with those it may as the camera turns it may grab some new features and some that were previously in the field of view just fall out of the field of view they disappear that is fine.

So, it just robustly and automatically tries to use features; however, it can. So, extract and maintain from natural scenes this might not be such a bad room for this, if as I look around I see that the walls have nice grid patterns on them if I were to take the headset and look directly at a with the camera and look directly at a perfectly white walls there may be no features that I can use based on whatever feature detection algorithm I have made right we have to be careful about that, but it may be that for whatever automatic feature extraction and maintenance and algorithm you have made we may be able to find a natural scene that it might not work so well in.

So, if your engineering a product you have to think about that is it going to cover all the cases or not or is going to cover enough of the cases that the users will not complain, because if you get it wrong and you make gross errors in your estimate of your position then this will have a dramatic effect in virtual reality, because it may make people nauseated very quickly you will feel all of a sudden like you are moving. When you are not so, very bad problem one thing you have to pay attention to here there are some tricks for doing this, but it may be possible to fool them as well you have to remove any moving objects from the scene.

So, the headset may be moving the camera may be on the headset and moving you have to remove objects from the scene that you conclude are not moving just, because the headsets moving. Now, you have a; the gyro and accelerometer on the headset. So, you may be able to figure out what motions are corresponding in the image plane you may be able to figure out, what motions correspond to your own motion? And which motions are not explained by that and then remove the ones that are not explained by that. So, it is not bad there is some information you can use there, but generally is difficult.

So, I would say that overall using current technology the reliability is low you will see good demonstrations, but making us work very robustly in a very large number of studies

and have sub millimeter level accuracy I have not seen anything that can accomplish this yet and. So, this is this is the kind of the state of the art this is the state of the art for natural features this then motivates people to go to artificial features.

And, I would call this instead of hard computer vision I would call this trivial computer vision, because most of it boils down to something called blob detection as far as the vision park right you are not doing very much work other than I am trying to identify markers that you have put into the environment. So, it boils down to some kind of blob detection or variations of that whatever patterns you have decided to put into the environment you just have to go and quickly scan those back out of an image.

So, you are not really doing the type of computer vision research that would make your computer vision professors proud let us say right this is more like a vision in 1971 or something you know computer vision you know challenges of that error let us say. So, so just simply pulling blobs out of an image, but you have designed it. So, that you have easy to find features.

So, some examples you could use what are called QR tags you could use what are called retro reflective markers.

(Refer Slide Time: 18:00)
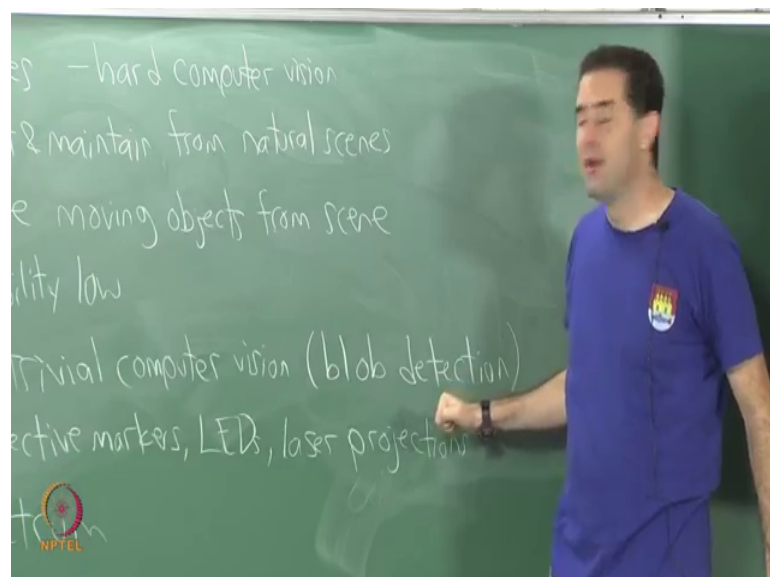
You could use LEDs laser projections some of these could be designed to work in the infrared spectrum. So, that they do not interfere with visible lights. So, that human tend

not to notice them and there is not a kind of interference that may cause other troubles. So, so, that is. So, choice you have you have IR versus visible spectrum.

So, so that is the idea either you go the hard way or you go the easy way when you go the easy way for artificial features you end up with a more controlled environment and you can guarantee much more robust performance right it works in a larger variety of settings and with much higher accuracy. So, that is one thing that we went with an oculus on that stage for dk 2, because of this high level of reliability and accuracy it was better than trying to make the computer vision the most impressive computer vision ever.

For example, with a high level of robust this was very very hard to achieve, that it would be more impressive if any of you can find a solution that works with natural features; it has the same level of accuracy and robustness as this current a dk 2 tracker. For example, then I can assure you many companies would be very interested in that. So, it is quite a difficult challenge.

So, let us take a look at kinds of markers let me say something about retro reflective markers is a very common technique; that is used in motion capture systems they have been around for quite a while the motion picture industry has used them before for capturing dancers and digital actors capturing real actors and turning them into digital actors. Let us say when you use retro reflective markers the idea is you take the camera.

(Refer Slide Time: 20:09)

And then around the camera you have a ring of lights[n] a ring of bright LEDs all right. So, you know there is lights facing me today right. So, so you are trying to get a good image of me and the camera. So, same thing if you are trying to image something out here shine some bright lights on it with LEDs except we can move into the infrared spectrum. So, that the actor or artist who is moving around in this motion captures studios does not feel the interference. So, it is not blinded by the lights. So, you can make these infrared LEDs. So, just exploit the fact that we cannot see in that range.

(Refer Slide Time: 21:09)
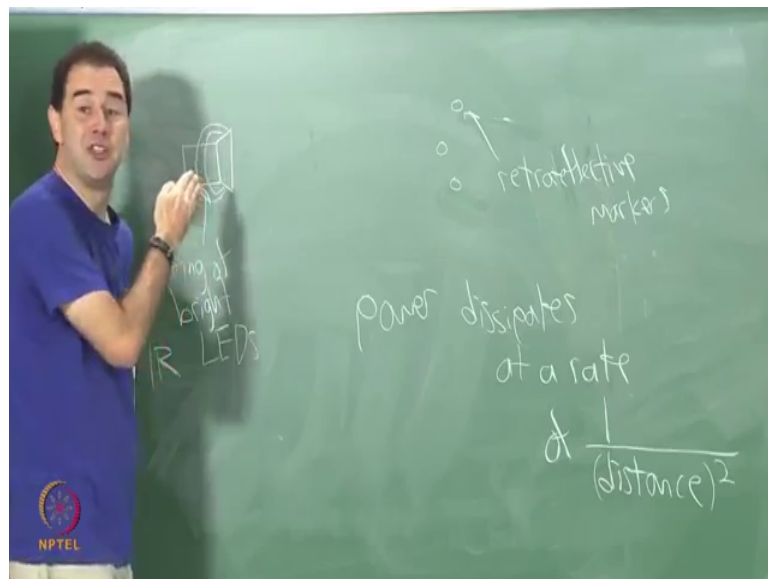
And then, out here in the environment if you put the features which are called retro reflective markers retro reflective markers you can buy special paint and you can paint tiny objects you may have seen ah, what looked like deer antlers sticking out of objects before in these kinds of systems and you have grave tiny painted spheres this is retro reflective paint, what will happen is that the amount of light coming back to the source.

So, if I shine a light and IR light onto the source and then I look at the reflected light back it will be independent of the angle at which the light hits this object. So, there will tend to be a very bright spot coming back. So, very nice for that. So, that is retro-reflective it is as if you are shining into a mirror regardless of the orientation which is a very unusual property right normal; if you want to have the light bounce back you have to have the mirror perfectly aligned this is as if you have all orientations of mirrors when

you shine the light. So, it is guaranteed to come back and hit you very bright hit the source very brightly.

So, this is one possibility one of the things you have to pay attention to those if you would like to extend the range as much as possible of these sensors then remember that the power of your light dissipates at a rate of 1 over d squared I will put 1 over distance squared; 1 over distance square right. So, just usual power dissipation of a of a propagating wave in 3 dimensions the amount of energy or power dissipation is one over the distance squared and. So, the further away you move these you know you get a lot of power up very close the further away you move the lower it gets and you have to get the light to essentially travel twice the distance right it has to travel from the ring to the marker and then from the marker to the camera.

So, you have double the distance. So, one possible improvement is to just put the light directly on the markers right if I can directly like the markers I have cut the distance in half right so; that means, I should be able to move the markers further away and get the same amount of brightness in the camera; if I do not mind putting batteries or whatever I need or this has to be plugged into a separate source if it is on a headset, then I have to power these lights. So, that is another alternative that is what is used in the lab which is I take the camera.

(Refer Slide Time: 23:51)
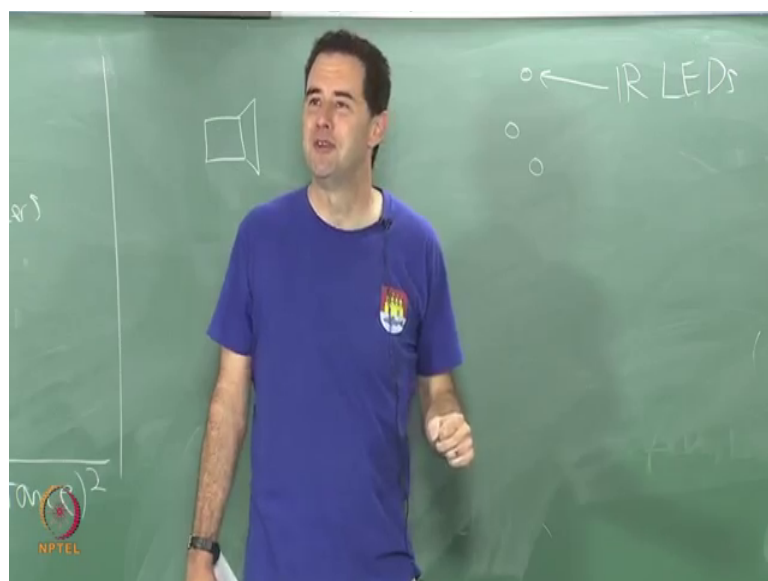
And instead of having retro reflective markers I have these special markers that I have designed and these are just infrared LEDs.

And we are also taking advantage of the fact of LEDs becoming increasingly more powerful and low cost and very compact as well. Earlier LEDs were larger they had a huge lens across them. So, you could see them looking bright from different angles, now they emit a much greater amount of light infrared light in this case very low cost very small footprint.

So, now it is essentially if you want to set up this visibility situation and think about the amount of polities LEDs have to have you have cut it in half in order to be able to see them in the image with the same in the same way and also there is the efficiency of the retro reflective material which I have uncovered as well. So, so this ends up being more efficient if you have the ability to power these markers it is cheating even more right.

So, so in terms of computer vision real computer vision you know if you were to make something powerful should be; using natural features if you put the features in the environment yourself as special markers that you have painted you have cheated right you are helping the vision out. So, that is fine that gives much better performance, but not as impressive from a pure computer vision standpoint. You want to go even further put some bright lights in the environment bright and infrared in this case and then just track those and if they are small very well separated you get these points or these features that we need to get to the next level.