

Virtual Reality Engineering
Prof. Steve Lavalle
Department of Multidisciplinary
Indian Institute of Technology, Madras

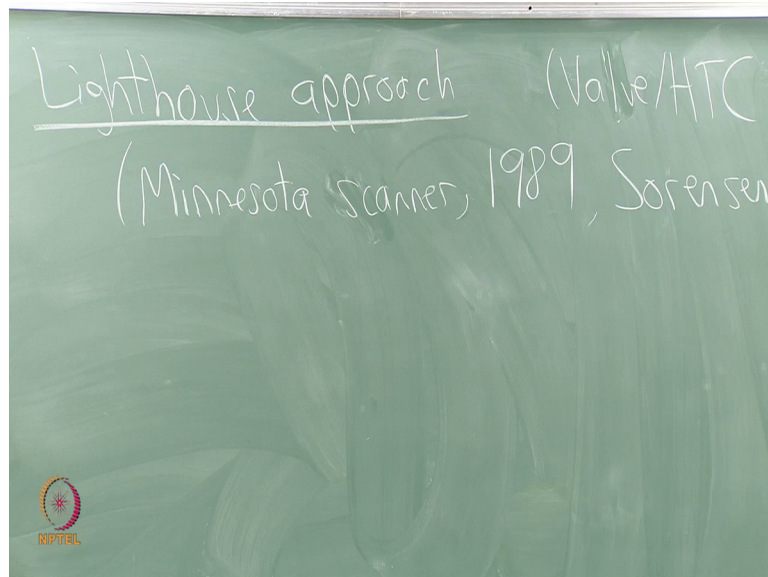
Lecture 14 – 2
Tracking Systems (lighthouse approach)

What I would like to explain next is if we were just doing, what I called how it did I called it, I said I had hard vision and I called it I think trivial computer vision for the case of trivial computer vision for the case of creating your own markers or features in the environment. So, if we were going to be doing trivial computer vision, why not just bypass the entire imaging process all together.

And just come up with some direct line of sight method in the physical world, where a sensor can detect this connection, this perfect line of sight without relying on cameras; cameras are cheap and abundant right. So, it is very easy to just grab a camera, grab your favourite library, such as open c v and start doing image processing, but it may be that we end up with something that is more reliable and better engineering by just bypassing the camera all together and, that is the next method that I want to cover which I think is a is a is a great idea.

This was implemented by valve recently, the game company and named the lighthouse approach. And it is based on the same principles it is based on line of sight visibility it just bypasses, the camera it is like making a kind of virtual camera, if you like it is another way to interpret it.

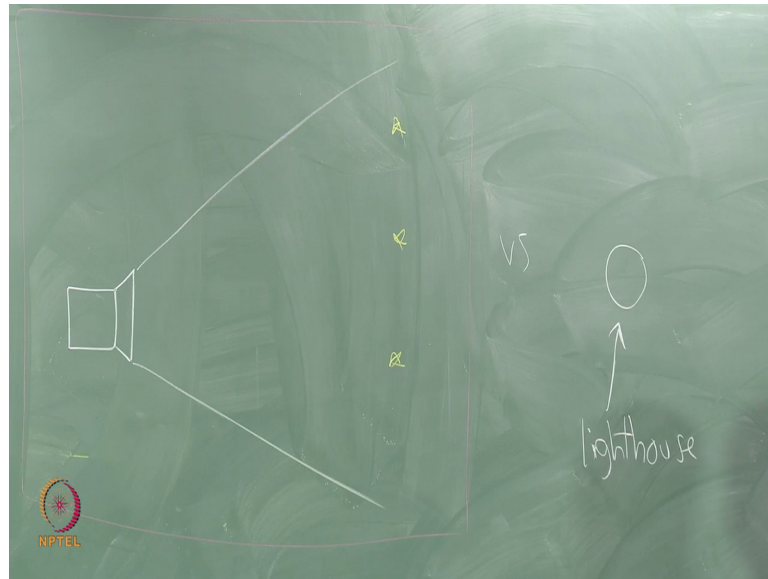
(Refer Slide Time: 01:34)



So, lighthouse approach like I said you can find this in the current valve HTC headset, it is been in the press lately it is also this approach was called the Minnesota scanner, or 1 implementation of this which you can find in a paper came up in the robotics literature in 1989 by Sorensen et al and so, most of the ideas the engineering systems that are coming up in industry and they have been around in earlier implementations in previous decades it is just that, they were not feasible in the consumer product realm because, the components often were not were not good enough or there was not enough motivation based on the market.

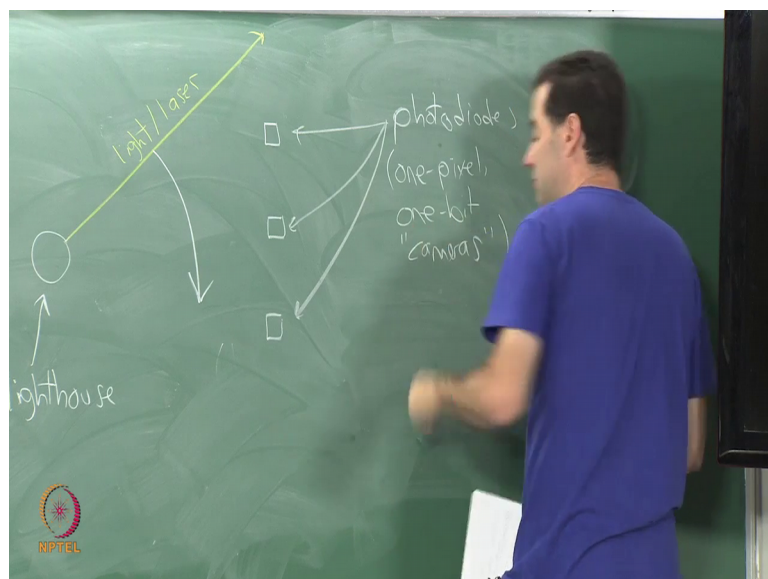
All right so, here is the idea, let us think about the visibility or mathematical equivalence between 2 scenarios. So, let me let me draw it over here. So, let us look at 2 scenarios, 1 of them we have already seen we have a camera.

(Refer Slide Time: 02:58)



And it is looking at some LEDs. So, it images some LEDs and that is all we have just studied. So, that is 1 scenario, we consider this versus I want to have what I call a lighthouse and yes I mean the same principle as a spinning light that is used for the navigation of ships along the coast. And so, I have a light that is spinning I will just draw the beam of light coming out.

(Refer Slide Time: 04:00)



So, there is some beam of light coming out perhaps, it is coherent light so, might be a laser spinning along in some way. So, it is rotating around. So, it is just undergoing a

yaw rotation, if you like just keep spinning around at some rate and now out here, I have sensors these are my sensors, which maybe are just as simple as photodiodes, maybe I can even call them 1 pixel cameras or 1 pixel. In fact, 1 pixel with maybe just on off right, there 1 pixel cameras and they do not even have 0 to 255 values they are just a pixel on or the pixels off just that simple, that is what I will call a photodiode 1 pixel 1 bit cameras, and I will put cameras he quotes you as it is being a bit silly right I am just trying to point out the simplicity of it. So, are not these two systems very similar, if I have a camera and it takes an image I reason about the angle to each one of these features based on the parameters of the camera correct.

If I have a beam of light that is sweeping along here, each one of these sensors will receive a blip right there will be it will go from 0 to 1 here, and then 0 to 1 here, and then 0 to 1 here, if I know the rate of rotation of this then I should be able to recover the angle, I need the rate of rotation and I need to know the offset like I need to know at exactly what time is this beam let us say pointing straight up.

If I know that and I know it is rate of rotation, I should have the same information here because I know the angles at which these photodiodes are located I, will not know how far away they are will I right did I know how far away they were here, no same thing in this model right. So, I did not know how far away they were you could use the size of the blob and a bunch of other you know maybe image processing reasoning, but based on this PNP model that we used there is really no difference between the 2 all I am the only information that really have here is what is the orientation.

Now, I am drawing a two dimensional picture here, instead of a three dimensional picture. So, we have to be a little bit careful I am cheating a bit, but let us just suppose the world were 2 dimensional here, just for a bit do we agree on the mathematical equivalence between the 2 of these let us see them all right. So, the way you put them in boxes here. So, we have exhibit A, using cameras exhibit B, using a lighthouse and I just want to argue that the information provided by these is equivalent assuming I know the camera parameters here, and here I know let us say when the laser is vertical and, where the laser is located in the world and I also know the rate of rotation of the laser to some reasonable level of accuracy. And if I spin this laser really fast say I spin this laser at 100 hertz that should be equivalent to a frame rate of a 100 hertz over here, do you agree and notice how I think I have avoided over here, the problem if I have a spinning laser I have

avoided the problem of blob detection right. So, over here in a single frame I have to scan an entire image looking for blobs and, I have to deal with all of the artifacts of computer vision here, I just have to deal with can the laser see the photodiode, if yes it lights up the photodiode. So, very simple circuit the physical world it is as if the physical world is doing the image processing for you right. So, why make an entire imaging system, when all you are going to do is detect a blob.

So, why not have a spinning laser and do it this way. Now there is some disadvantage in the sense of that I do have to build a mechanical system here with a spinning laser. So, you can put a mirror inside of here, and use some lenses to spread the laser apart into a vertical stripe and then spin it around. So, you have something mechanical moving here, here we seem to have no additional moving parts. So, that is one thing that is going on.

If I want to make this work over here is it better to do an inside out or outside in. So, it is better to attach the lighthouse to the headset or put it in the world, what do you think probably better to put it in the world and say I do not have a spinning thing on my head. So, so, valve did in a recent show for example, is they put it inside of a room in the in the corners of the room up on the ceiling they, put it in opposing corners you just need one of these devices, but they put 2 of them to handle the case of occlusion what if you are blocking one of these devices right.

So, this goes spinning along providing this information and I made a simplification by making it 2 dimensional how do I make the third dimensional work any ideas, we agree on the equivalents here in 2 dimensions right that, I can I can make a virtual camera here right so; that means, that if I am let us see if I am doing this with the if I am doing this with the lighthouse fixed in the world; that means, I have to put these photo diodes on my headset correct, or if I want a track of control, or I have to put them on the controller whatever it is I have to put these around and I guess ID better have some kind of circuitry, that can also take into account the timings right. So, that I know exactly how many milliseconds later this photodiode received the light after this one correct. So, I need some circuitry for that should not be too hard, oh what about the identification problem trying to figure out which LEDs which is that solved here.

Because I know based on the circuit, I have built which one of the LEDs is reporting that it is been to lit up. So, beautiful it solves the identification problem in a very simple way.

(Refer Slide Time: 10:37)



So if I have these photodiodes I am just drawing let us say the headset here, and I have now the photodiode placed on it, let us draw some photo diodes placed around I can put a bunch of them on here, it is pretty easy photo diodes are cheap. So, I put some photo diodes around on this. And now I have this beam of light sweeping. So, let us say I take this two dimensional picture here, and I extrude it outward this way. So, that all along this was just a vertical beam sweeping by ok.

So, if that is a vertical beam sweeping by that is what we agreed is happening here, then this looks like I have a vertical beam like this and, it is sweeping across right it goes around the room and comes back goes around the room it goes back. So, that will give me the horizontal coordinate right, we will give me the x coordinates in the in the world how do I do the other part why do not I just make a sweeper that goes the other way. So, I can use the exact same idea and put them both inside the same base station and they are just spinning 90 degrees offset with respect to each other.

So, I have another spinning lighthouse signal going like this providing the y-coordinates. So, very nice right so, that is how I get the other coordinate and now there is another problem you might expect to happen there should be a drift error that happens, which is going to be that I might when I turn on the lighthouse I may see where the headset is at and I may declare that to be the origin in the world perhaps and, then as I keep going I

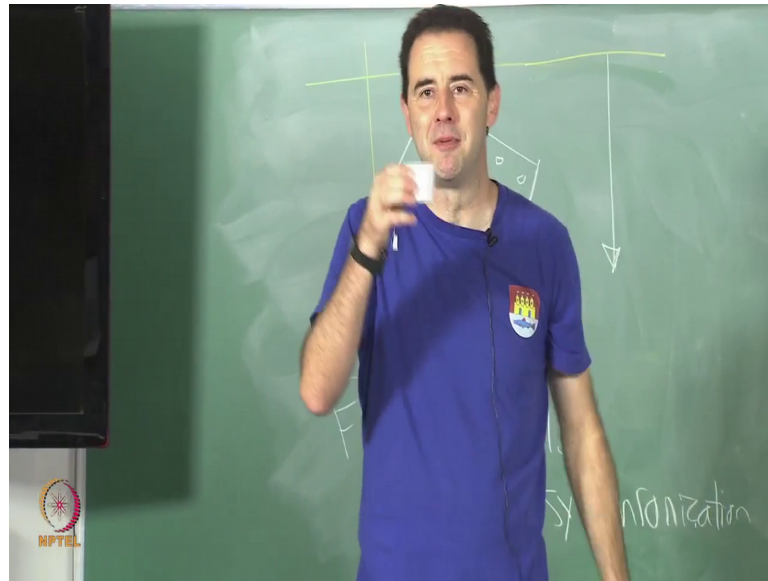
know the rotation rate of this laser, but I should expect some kind of drift over time there should be some angular drift.

So, in order to correct for that what they do is they put a bunch of lights all around and flash them all at once. So, I could make a floodlight in other words right. So, let us call it a floodlight. So, I could have a floodlight pulse and that is for synchronization. So, for example, whenever this beam is vertical at the same time I just flash in all directions. So, I have light emanating in all directions that will cause all of the photodiodes to light up maybe not as brightly as if they were hit directly by the laser, but at least some pulse will be observed and we can again do this in the IR spectrum. So, we do not make it look like you have had a camera flashed in your face right.

So, we provide a flash pulse if all of these light at the same time; that means, the beam is vertical. So, that gives me a signal that let us me avoid compensate for drift. So, based on the rate of drift that you might have that will tell you what frequency to do the flashing at, I do not know what they use I have not had a chance to even try the device, but I would say that you know no more than once every 10 seconds or so, maybe once in a minute it would be just fine just.

I would have to you know I had to be one of the engineers experimenting with it to know exactly, but you do not have to do that very often does that make sense questions about that. So, this is a great technology I think they are making this available in some kind of open-source form. So, you should be able to play with lighthouse systems fairly soon and do your own engineering of those should be very convenient for making tracked objects in 3 dimensional space, you could just hook up your photodiodes to whatever objects you want to track now of course, that there is a downside if I just want to suppose this is this were my well let us actually pick take, my little coffee cup here, if I want to track this in the virtual world.

(Refer Slide Time: 14:23)



If I just wanted to do it with retro reflective markers or q r tags I just stick them on my cup if I want to use this technology with photo diodes, I now have to have a circuit on my cup and, it has to detect the light that is hitting them and then transmit that information somewhere right. So, there is some differences based on where the computations happening where the sensor is placed where is the light source placed. So, something to pay attention to, but nevertheless this is a very nice and elegant solution right questions about that.

So, that gives you some idea of how tracking systems work, I have shown you some examples of the technologies giving you some high level overview, the most critical and sensitive tracking part is the tracking of the viewpoint of your eyes. If there is just a little bit of jitter or shakiness in the tracking it, may lead to a very uncomfortable experience it is certainly an annoying experience.

And, if there is some lag or latency in the system then it leads to discomfort and even nausea. There is also another part which I did not talk about I may come back to that after I talk about rendering, which is you can use prediction to overcome additional latencies that are in the system because, the next part I am going to cover is graphical rendering, it may take time to produce the appropriate images that are going to go on to the display and then get to display updated accordingly.

During that time what happens, I may know the orientation in position, but by the time I finally, put the right information there it could be 20 or 40 milliseconds later. So, I can use further prediction techniques to estimate where the position and orientation is going to be to compensate for that, but if I have to predict too far then I may add more jitter to the system again, and it might become let us say perceptible artifacts that the prediction introduces and this may be unwanted.

So, these are some of the challenges that we that we have to deal with in these systems. So, so even if the tracking system itself has let us say effectively no latency, or very small amount of latency that can be easily compensated for very simple prediction techniques, we still need to predict further than that to compensate for all of the latency introduced by the next phase, which is called rendering in this case this will be the visual rendering part. So, that is what I want to cover next do we need one more.

Excuse me a minute. So, it seems like it seems like we do not have enough information right. So, we have two different spinning directions that gives us, 2 pieces of information where is the third dimension right. Let us go back to the image processing case when we had when we were using images, remember that we do not know how far away an object is due to the perspective projection of the camera.

And we agree that if we see enough points, then we get enough information from this p n p solution right, I went and argued the degrees of freedom coming down. So, you are right we only lose two degrees of freedom each time when I observe a point rather than figuring out exactly where the point is with a single image. So, all I am arguing is that with the lighthouse solution the same scenario exists. So, you still have the depth part is missing you do not know how far away it is that is the part that is missing, but it is the same scenario as you had with image processing.

So, you again have this p n p problem to solve. So, were so it is no worse than that hmm can you spin a laser some other way and maybe make it. So, you get all three coordinates in one shot I would have to think about that if there are certain geometries I might allow that, but even without that were fine were at least back to the same scenario that we were with cameras, but we have avoided all the problems of imaging systems yes.

Student: Sir in the Google self driving car do not they use a spinning laser to also find the distance from other cars and obstacles lidar lidar right.

That is right you can use you can use lidars and other kinds of depth measurement sensors and that is something I have not covered here, and I think that if you take the kinect and other types of what are called RGBD sensors right to provide both color information and depth information together in a calibrated coordinate way you can get significantly more information and I think that, one thing we will see in the marketplace in the coming couple of years is that RGBD sensors will become cheaper more abundant.

And we can experiment with using those in the real world and we may be able to have highly accurate featureless tracking like, where you do not have to engineer the features, but it remains to be seen because you still have the trouble of extracting away objects that are moving in your scene, but they do not correspond to your motion they are moving by themselves and so, forth. But this may provide another way of doing it; however, I have to say that I think that there is always a temptation to rely on a very sophisticated sensor that is gathering a lot more information than you need and, this is why I like this lighthouse solution very much because, it is actually not leveraging the fact that you have cameras lying around that are low cost, but re engineered.

Let us say not reengineered, but engineered a solution that is particular to this problem that I think is very elegant and. So, so this is a kind of depth imaging system, if you like and it is designed exactly for extracting the pose of the objects that are placed into the environment. Now how much more needs to be added to this kind of system.

So, that maybe you can do tracking of objects without having to go to a full RGBD scanner, or coming from the other direction will RGBD scanner RGBD scanners become cheap and abundant enough and will the computations become simple enough. So, that we could use them to get very highly reliable tracking of 3 d objects in the scene, which one of those directions will ultimately win out it is hard to speculate, but it is a good question to ask. So, so I did not cover imaging sensors that produce depth information directly, but I think those will play an important role in this space as well.

Student: Sir.

Yep.

Student: Sir, for depth information cant we do what we do for stereo vision like place 1 lighthouse slightly below, 1 lighthouse above and the two lighthouses activate different

photodiodes. So, you will be able to similar to stereo vision, but except here you are (Refer Time: 20:38) distances be much larger.

um So, so I think that is I think that is a reasonable idea, I think you are getting information from I think even in your example you have had to use two different photo diodes though right, or would the same cause you cannot tell the same photo diodes is being hit simultaneously right. So, is there a difference between one light house sweep hitting two different photo diodes versus two different lighthouses hitting two different diodes.

I think it may more or less be the same thing you are just creating more visibility rays, each of which adds more constraints to the problem right, either way your idea or this idea there is a eliminating degrees of freedom from the problem right. So, I think it is to do that, but I do not think it is necessarily better than this and in an implementation of this they use multiple lighthouses. Anyway, but that is mainly for occlusion just to make sure that the you know that hopefully not too many of the photodiodes are blocked, if I am turning my head completely away from the base station that is producing the laser stripes, then at least I will be facing another one that is producing the laser stripe.

So, I think there is more motivation for multiple lighthouses for that reason, any other questions. These are good questions and it shows that people are I think paying attention and understanding what we were doing here.