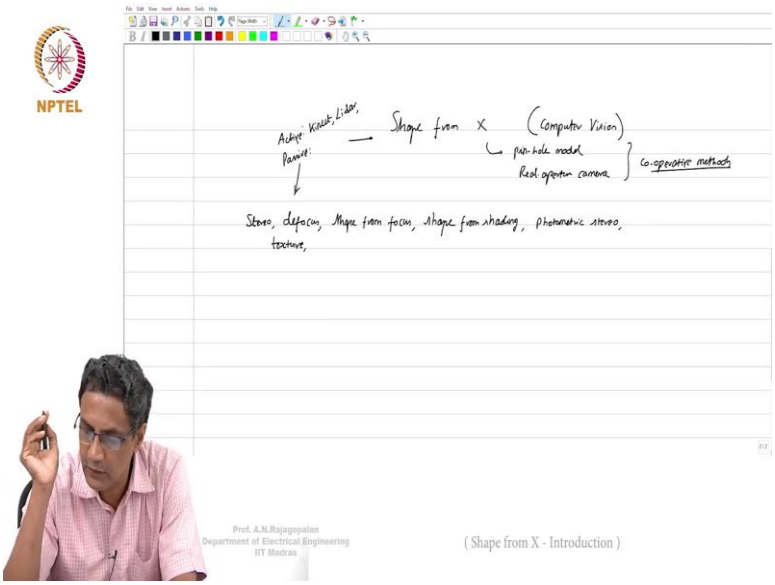
Image Signal Processing
Professor A.N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology Madras
Lecture 26
Shape from X – Introduction

(Refer Slide Time: 00:17)



What that really mean that X could be anything and this could come out of pinhole model. Whatever is this information that you are seeking from X, so the X could be some information that is coming out because you have a pinhole camera or it could be simply that you have a real aperture camera or you know there are actually methods that do cooperative kind of algorithm which means that they will exploit the fact that there are some things that are valid with respect to the pinhole model too even though you may be using a real aperture camera like we also did.

Certain things that we knew over valid, equally valid even though you are using a real aperture camera but there are certain principles from pinhole model which are still applicable and therefore we use them. So, there are actually methods that could simply assume everything to be completely in focus which means like you know pinhole model. So, you cannot get any kind of blur there.

Real aperture camera where you might say that well you know it depends upon the scene and all that and 3D scenes will generate blur and so on and which means they may simply exploit something related to image formation in lens and there are these other methods that might actually try to combine (())(01:39) coming from both.

They may say that the central ray actually still follows the stereo law and therefore I mean you can use stereo algorithms that are typically used on focus. You use them even on a real aperture camera and so on. So, there are these cooperative methods what are called cooperative method and there is a whole lot of things that happened here and this is really a vision sort of thing. These things come and what is called computer vision and the boundary between computer vision and (())(2:15) is a little blurred.

Somewhere you do not know where you are but typically computer vision is more about telling more about the scene. It is less about the image per se but using the image to tell more about the scene. Okay whereas image processing is like you have an image and you want to process it and so on. So, normally image processing will fill into computer vision.

Now here right so what do you think could be like cues? I mean suppose I gave you an image, what kind of cues you think you will use in order to like for example, we ourselves what do you think we use? Right when I see you, I know that there is a scene that I am so careful and I write I know what is where. What are all the things that possibly the uses? Let us get ourselves write them down one by one. First thing is stereo right?

Okay we all know that there is a pair of (())(03:07) therefore there is a reason for it. Stereo, let me just write it properly. Stereo, anyways stereo also has something called as 2 view stereo then multi-view stereo and so on. Stereo does not always mean just an AUX stage with 2 images or something. That is like 2 views trivia.

So, most common thing you can have multiple views, that is called multi-views stereo. Alright, defocus blur, there is something called along with this you know, so there is something called shape from focus also and I will tell you what those things mean. Shape from focus, something called shape from defocus and the shape from focus and stereo, defocus, shape from focus then there is something called, how many of you have seen shape from shading? There is something called shape from shading. I will show you some examples so that you get an idea.

And then there is something called photometric stereo. So along with stereo I told you there is Multiview and all, so in all of that the camera is moved but there is something

else in which the camera is not moved. It is called photometric stereo. There is also something called shape from texture. This is not so common but it is there.

Texture, then shape from shading, shape from focus, okay by the way talk about shape from focus I forgot to tell you that there are kind of say 2 kind of classes there. One that is active. One that is actually passive. So, active is your say, when you talk about the Kinect or when you talk about Lidar, all these are active in the sense that they kind of throw something. So, they are active in the sense that it is not a passive skip.

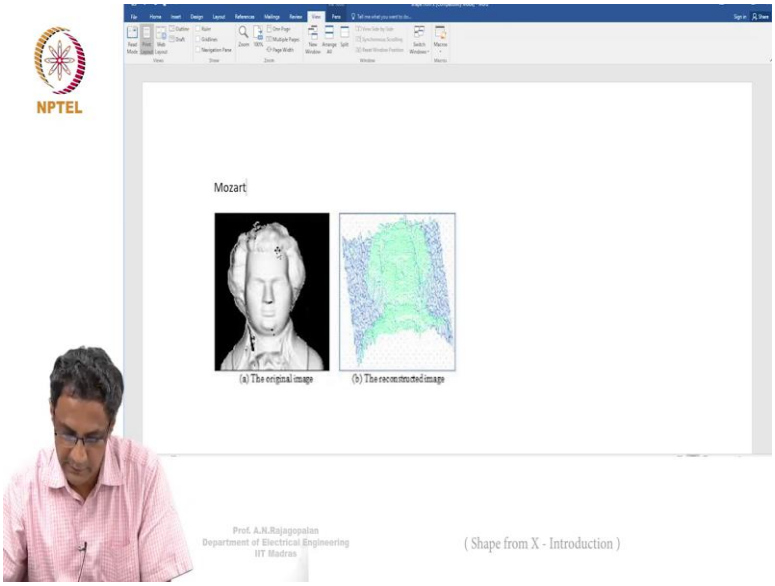
So, they would either throw a light pattern on the scene or they would actually send out a beam. They will actually try to compute time of flight just like you do in echo. You send something and you wait for the echo return. So, you get a sense of how far away the objects are. So, all these are like active.

And active things are you know are typically either they are very expensive or in terms of their usage where you can only use them at a certain constraint conditions and so on. The passive ones there is full of lot of interest in passive. Passive is someone which we normally for us passive basically read views images. So passive is where kind of we are now. So all that we are saying is the cam right we are simply capturing images.

Okay so the scene is kind of active for us. The scene is giving out intensity and we are capturing them and now once you have these images we want to be able to tell what is where. And shape typically means the 3D shape of the object. You can also call it dep map or you can call it shape. Because wherever dep map then you know say relatively where every point is irrespective of the other that is shape.

Now, photometric stereo texture and all these. So, our focus is here in passive and that too we are not going to do all of this and so on okay. There is no time for that. So, I am going to talk about just one method in details.

(Refer Slide Time: 06:21)



So, the first one right is really a photometric no this is shape from shading. So here you have this one image. Shape from shading typically means that if you see an object, this give you single, this is one of the hardest by the way. This says that I just actually we can do it rather well. We as humans. Because when somebody shows us a photograph there is nothing like a 3D in the photograph right.

It is a planar thing and on a planar surface something is there and then when we just look at it even if it give you let us say an image of the sphere. When I just see this sphere I know that this is the sphere. That I know that it is probably curving and all that. So, you bought one of the things the we use is of course our brain uses multiple things but one of the things that we use is what is called shading. So, what shading means is that depending upon where the light source is and depending upon where the camera is and depending upon how the surface normals are on the object surface.

So, the surface normals wait a minute if you think of a sphere, so the surface normal suppose think of it like this and let us say you have a camera looking at it from somewhere and then assume that you have a light source shining from there. So, what it will mean is, so you have a surface normal that is going right here you know like this and then it will slowly turn and then by the time you come to that top of the sphere it is upwards.

So, what will happen is, so the angle that the surface normal makes with respect to the light source is what will tell you and there is a law for that and that is upon that law

that will tell you it will give you the intensity that you likely to get in the camera for that point and this is also depends upon what is called an albedo of the surface and so on because you can have a material character stick that could change through the surface.

If a sphere made out of one material then probably it does not matter but then you could have multiple materials and you know being exposed to a light source. So, the shading is this change in the intensity as a function of the surface normal and weather light sources.

So, it seems like now when you see an image that we are able to make out all of that. We may not be able to exactly tell what the at what point what depth and so on but we have a rough idea just by looking at one single image. So, shape from shading is something that banks on that law and it is a kind of well understood law. Except that (())(8:49) problem because we eventually have one image to look at and then you have this albedo which is unknown and then you have the surface normal which is unknown. Therefore, there is only one known which is an intensity and based upon one known you have to (())(9:03) multiple unknowns.

So, it is kind of an ill posed problem. So, this is not something that we do in this course but I just wanted to show you an example. So, can you see this guy's face sort of there is a nose, there are the eyes, I do not know. Guys at the back maybe I should have picked up a say a better reconstruction. This is I picked up from the net. Somebody had done this. I can show you an example that a way you can probably appreciate it better. But I hope you are able to see the eyes there, the nose there and then the hair there. You can make out no? Then the next one.

Student: (())(09:46)

Professor: We are just using this one single image. Okay the one image on the left which is using this image. Okay so this is like one single image of the statue that you have. So, saying if I had say let's Mozart statue in front of me and they had some light source which is shining on it and then I viewed it through a camera then whatever is that image that I get the idea is that the image function are the surface normals of this object and also relative to these normals where is my light source kind of see direction.

Student: (())(10:22)

Professor: The right is actually a dep map.

Student: (())(10:27)

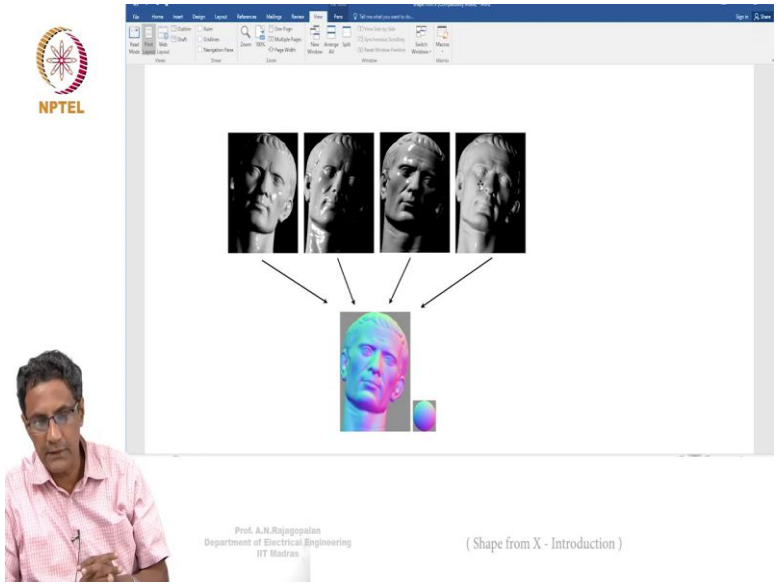
Professor: Well I mean if you talk about something that is totally uncalibrated in that case you do not know where the light source is. You do not know anything. But if it is calibrated situation you can assume that you know where the light source is and then only the normals are unknown. That is okay. But there is something called a base relief kind of is ambiguity.

The base relief, ambiguity depends upon the what all you assume, what all you do not assume and so on. The most general case you may not even get, so for example if you are asking whether I get absolute dep map, no. You may get the dep map up to a similarity, up to a Euclidean depending upon what you assumed to be known an what you assumed to be unknown.

So, that is a whole thing, that is a whole set that is kind of uncalibrated which is the most complex thing where you do not assume anything at all and then a calibrated thing where you can say tie things down at least you know that some things are known. And the more uncalibrated you go in the sense that you just take a camera walk around, you capture pictures, do not assume anything at all.

So, the more uncalibrated you then the dep map that all you estimate is only up to a factor. That factor could be simply scale. I mean if you know a lot then it can go up to a similarity. It can go to an affine depending up on how much you do not know. But that is okay. You know people are happy with that. Because you know instead of not having anything, you at least have something now. So, this is shape from shading. I am not able to write there. This becomes too big.

(Refer Slide Time: 11:59)



The other one, this is called photometric stereo. What do you see here? Again, this is supposed to be a dep map of this reconstructed face. What do you find? There are 4 images here. Exactly, so the light source is changed. So, from where you are shining the light is changed but the object is remaining exactly there. So, it is like the camera is not being shifted. So, every point, so for example so the ear here if I mark a point that the ear is exactly the same location. Nothing has changed.

So even from where the camera is taking the picture, that is not a big change. But then it is called a photometric stereo but actually it does not use in stereo you would actually translate the camera. Here you do not translate the camera. The camera is sitting bang there. Your object is sitting right there. You do not move these two. Only thing that you move is the light source. You take one from here. You know you shine a light from there. You shine of course not from the back. Then you shine something from here. Then you can get 4 or more or whatever. And there are certain conditions that you have to meet.