MITOCW | mitmas_531f09_lec03_1.mp4

ANKIT MOHAN: Hello, everyone. I'm Ankit Mohan. I'm a post-doc with Ramesh. And I'm sorry he's not here today. I'm going to be talking briefly about what's called epsilon photography, or film-like photography. It's actually how can we improve film like photography. And I'm not sure what part-- some of this might be very straightforward and obvious to many of you. So if it seems like I'm going too slow, please let me know. Or if you want more detail on any of these topics, again, stop me and ask the question.

So before we can try to improve film-like photography we should understand what I mean by film-like photography. And this is basically what's been the camera obscura model, where you have a pinhole or a center of projection and you have rays of light that goes through that point and form an image on the sensor or a film plane.

So what you see over here is-- on the left, this is traditionally how you draw an optics ray diagram. You have the object of the scene on the left. And rays always travel from left to right. And there are people who do hardcore optics who can get really annoyed if you don't follow this model. So it's always a good thing to go from left to right.

So you have the scene on the left. You have a center of projection, which is a pinhole, in this case. What I haven't shown here is-- basically, you can imagine a box that's surrounding the central projection and the sensor. And only a single point allows light to go through. What this gives us is a single ray from every point in the scene is allowed to go through the camera and forms an image on the sensor.

Now, most objects around us are actually diffuse. What that means is-- technically, it's called Lambertian. What that means is the rays-- when you have an internal light coming on an object, it reflects light in all directions. And most objects are diffuse, in that all the rays that come out of a point on the object have roughly the same intensity, whereas the other case would be a specular object, which reflects light in primarily one direction and not in all directions. So because most objects are diffuse, when you have a pinhole camera taking a photograph, it looks very similar to what it appears to the eye. So it captures most of the information coming from the scene.

Now, what a lens does is slightly different, is that it actually integrates over an angular exchange. So in this case over here, you have rays coming from a point in the scene. But not just one single ray gets imaged on the sensor. But you have a whole cone of rays that get imaged on the sensor. So in this case, all the rays coming from a point in the scene that go through the lens aperture get focused onto the sensor plane. And this is basically how a lens works and how a camera, modern camera works.

Now, again, this is very straightforward stuff. But a lens obeys certain properties, in that the ratio of the distances has to obey certain properties. And what this basically tells us-- and I'm going to skip over some of this stuff-- is-- this is, I think, the most important thing. If you have a lens, then only one plane in the scene gets imaged onto the sensor exactly. And there's a one-to-one correspondence between which scene is going to get imaged based on the focal length and the distance between the lens and the sensor.

This was not the case if it were a pinhole. In the case of a pinhole, everything appears in focus. And you have what's called an infinite depth of field. So unlike a pinhole camera, a camera with a finite aperture lens actually has a finite depth of field.

Now, depth of field has an interesting definition. If you look it up in Wikipedia, in the case of film-based photography, it was defined that as when you take a picture of a scene and you print the picture at a certain resolution-- at a certain size paper, and then a standard human observer stands some finite distance from it, can he or she has a difference between two points, whether it's one is in focus or not in focus?

So based on that and certain perceptual tests that they did, they came up with this definition of how far you can get from the plane that's going to be in perfect focus and still give the appearance to a viewer that the plane is in focus. So in the case of the digital camera, what it roughly translates to is that when you go away from the plane of focus, you are going to get-- if you look over here, you're not-- rays are not going to focus to a point, but they're going to create a test-like blur. And if the size of this just like blur is smaller than the size of a pixel, usually you cannot tell the difference between whether it's in focus or out of focus.

So there is this finite region around a plane of focus that's called the depth of focus. And it's actually-- it's not symmetric. It's usually greater behind the plane of focus and smaller in front. And there's a corresponding depth of field on the sensor side. So are there any questions about that? Is this obvious stuff?

What's interesting is that the size of this depth of field depends on the side of the aperture. So in the case when we had a pinhole where our aperture size was infinitely small, the size of the depth of field is infinitely large. So everything is in focus. But as we increase our aperture size-- like you can see from here, we went up here. The corresponding size-- because these cone forms are much larger cone angle, the region in which the size of the blur would still be smaller than a pixel, it becomes smaller. And you have a much shallower depth of field.

So there's something that photographers often use when they take pictures, like portraits or macro-photography, is they try to open the aperture or keep the aperture as wide as possible. And that results in a very shallow depth of field. So only the plane of interest is in focus. Anything behind and in front of it appears like a blur. It has a nice blurry appearance.

On the other hand, if you're doing something like landscape photography, you want the tree that's 10 meters from you to be in sharp focus and also a mountain that's five kilometers away to be in sharp focus. So usually, people use a smaller aperture size in order to get everything in focus.

So we'll come back to this a little later when I talk about how you can computationally modify the depth of field and things like that. But in general, it depends on the application. The application dictates what kind of depth of field you need. And most cameras give the photographer an opportunity to set the aperture size, which sets the depth of field.

So because there is only a single plane in the scene which is actually in sharp focus, if you use a camera that does not have a pinhole aperture, which is most cameras, you need to be able to select which plane you want to focus on. And that's usually done these days using autofocus.

The cameras use different techniques for autofocus. The most common one these days in SLR cameras is the space-based autofocus, which is a really interesting technique that I think was first proposed by Minolta way back in the late '70s. And what they essentially do is they form two separate images from the different-- so this is I think likely from the pattern. This is the main aperture of the main lens. And what they essentially do is create a rangefinder, where the baseline is equal to the diameter of the lens aperture.

What that means is essentially doing stereo or creating one image from one corner of the lens and another image from the other corner of the lens. And looking at those two images, if the scene is in focus, those two images are going to be exactly the same. If it's not in focus, there's going to be a phase mismatch. And by observing the phase mismatch, you can determine which direction the lens needs to move in and by how much. So it's a single-shot focusing technique where, by just getting this one reading, you can move the lens in the right direction and get an in-focus scene.

And this is usually very fast because you don't have to constantly keep searching. And the downside is that you need the special hardware in your camera. And most SLR cameras have this kind of hardware in them.

Another technique that most point and shoot compact cameras use is contrast-based autofocus, where-- since you have a live view coming from the sensor directly, you can look at one frame. And you can try to maximize the contrast and move the lens back and forth until you get the maximum contrast. And since you don't get an estimation of the phase, like the previous case, you cannot-- it's not a single-shot operation.

You have to usually search through the span of the focus settings and find out which one has a maximum contrast and stop over there. That's where it's in focus. And it's usually slower than the previous case, but you don't need dedicated hardware in order to do this. And most compact cameras use this.

Another technique that some of the older film-based compact cameras used was using ultrasound or infrared-based estimation of how far a scene is. And it's something that's not very prevalent anymore. And it's also not very accurate.

Another technique that I don't mention here is what's called a rangefinder camera. And usually, that's a separate unit from the main camera. The difference here is that this lens-- the autofocus occurs through the lens. So what gets on the image plane is what's used to determine whether it's in focus or not in both these cases. In the case of a rangefinder camera, there is a separate unit, which basically does this shifting and trying to find when it's aligned. And it's usually done manually rather than automatically.

I think the important point over here is there's lots of work that's gone on even before computational photography came into its being in the area of trying to find very quickly, and effectively, and repeatedly set the focus automatically on a camera. And there's lots of engineering that's gone into that.

So focus was the first thing that a camera needs to worry about when it tries to take a picture. The second thing is what's called exposure. And what I'm trying to show here is that the brightness of something that's daylight versus something that's dark is widely different. And you have just 0 to 255 8-bits or, at most, 12 bits or 14 bits to work with in order to compress all of that information in there.

And usually, this span or the dynamic range does not-- it cannot go through more than two, or three, or four decades at most. And so what needs to be done is need to decide what exposure to use on a camera. And so this is a scene that goes underexposed versus overexposed. Overexposure means you let too much light into the camera versus underexposure, if there wasn't enough light and the scene was-- the image was dark.

So exposure itself is comprised of these three things. One is aperture size-- the larger size, more of the light coming in, and the brighter the image is going to be. The shutter speed, how long you keep the shutter open for-- if you keep it open for longer, you get more light in and the image is brighter. And the film sensitivity, the light coming in, how many-- in the case of film, how much chemical can it translate, can it change or chemically modify? In the case of your digital sensor, it's the digital-- it's the analog to ADC converter gain is what is set by the ISO.

So these three things together determine what the exposure should be on your camera. So if you set a certain shutter speed, you need to determine what the corresponding aperture size and the sensitivity should be in order-- before you can take a picture. And once again, older cameras require you to do this manually.

Usually, you would have film, which was of a certain sensitivity. And you set an ISO 100 on it. You would select the aperture size. And then you would have to-- based on some of rule of thumb or using an exposure meter, you would determine what the shutter speed should be.

So this was drastically changed by Nikon in, I think, mid or late '80s, where they proposed this Nikon Matrix Metering Scheme. And the idea over here is-- so this is what the SLR camera looks like. You have the main lens. You have a mirror. The film plane is back here. The light coming in gets reflected up here into the pentaprism. And inside the pentaprism, it bends, points, and it goes through the viewfinder into the viewer's eyes.

But what happens here is another small mirror reflects it up to the top, where there are these-- there's these five-- I think you have five different zones. And they had a light meter at each of these zones, which was basically capturing how many photons are coming in at that zone.

So even before the picture is taken, the camera knows how bright the scene is. And based on that and based on some heuristics that they came up with, they determine what the correct exposure should be for the given photo. And this was supposed to be a very revolutionary technique back then. It did away with all the various rules of thumb that people had come up with before this in order to estimate a good exposure.

And this is what led to the change where you can just Auto mode on a camera. You can just press the shutter release. And you don't have to worry about either the focus or the exposure. And once again, I come back to these things in the realm of computational photography and computational cameras in a bit.

The one last thing I want to touch on is the concept of color in digital cameras. And most digital cameras have what's called a Bayer filter. And it looks kind of like this. So adjacent pixels have different colored filters placed on top of them. And usually, there are two green filters for every red and blue filter.

And what this gives them-- so the image you get is this interspersed blue channel, red channel, and the green channel on the same sensor. And then they use demosaicing or interpolation techniques in order to recover a high resolution image in color. There are other sensors, such as the Foveon sensor, which does this spinning in depth rather than spatially. So for each pixel, they get a red, green, and blue color value.

One more thing over here I wanted to say is that the electromagnetic spectrum that ranges from radio waves to gamma waves, it's only a very small portion that we are interested in for photography. It's usually from 400 to 700 nanometers. And this region gets actually split up into these three channels, the three color channels that you have-- red, green, and blue.

But the only reason you have these three channels is because of the human eye, which also has a similar three channels. And cameras try to mimic the functioning of the human eye in that sense. But if you look at multispectral cameras-- and I think we'll come back to that in some other class-- you can have a whole number of channels between the 300-- 400 and 709 range.

So this is the CIE Chromaticity Diagram. This is how the human eye visually interprets color. So what you have is-- it's on an xy scale. And what you have on the edges over here are the pure colors, or the color-- primary colors that correspond to pure wavelengths going from 380 or 400 to 700 nanometers. And so anything that lies outside here is a pure color, just a single wavelength. That's what a laser or some LEDs would give you.

Anything within this is a mixture of various colors. And the interesting property of this color space is that if you have any two points on this color space and you mix those two colors in various proportions, you're going to get a color which lies on the line that connects those two points perceptually. And so if you have-- if you have a triangle like this, which is a color space, the XRBG color space in this case, which is what most monitors and LCDs use, you would get-- if you have color primaries that are at the three vertices and you mix those color primaries in various proportions, you're going to get a color within that triangle. And by simply varying the [? wave ?] of the three primaries between 0 and 1, you can go from completely white, which is in the center, to one of the three colors.

And that's what the color response-- the curve for just the red, green, and blue color primary looks like for film and for a typical digital sensor. What's interesting to note is that even though we've advanced quite a bit from film to digital, the basic technique still remains the same. We still have the same three color primaries. They look almost identical. There's very little difference between them.

And that's one of the goals of computational photography, is to do away with the film with the baggage that we still have associated with film. And part of this lecture is actually going to go in the other direction and say, how can we improve on that? So the rest of the class is going to be more about how can we get away from film, whereas this class is more on how can we improve on film.

AUDIENCE: In the graph that you have shown, it looks like the film has the colors more orthogonal being sensed rather than the digital sensor. You see the blue is leaking into the green and the red is leaking? But there, it seems it's-- in some sense, it's very less leakage.

ANKIT MOHAN: Yeah.

AUDIENCE: Is it in general too?

ANKIT MOHAN: I'm not sure, in this case, why it's like that. And also, note that this is just one film which is optimized for certain kinds of photography. I think Velvia is supposed to be very good for landscape photography, and sunsets, and those kind of things. And that's something that you could do with film. You could have a film that's suited for a particular task and different-- has different primaries, whereas for cameras, it has to be-- for digital sensors, it has to be something that goes across the board for different kinds of scenes and things like that. So that could be the reason why it's like that.

AUDIENCE: In the previous slide, [INAUDIBLE] two green cubes, two green squares and only one for red and blue.

ANKIT MOHAN: Yeah.

AUDIENCE: Is that because the eye is more sensitive towards the green channel?

ANKIT MOHAN: Yes, it's because when-- I think that's roughly the proportion of the cones in our eye also. And green, if you look at the value, v, the luminance, chromaticity relationship between RGB and that, green is the one that has the most corresponding-- most weight. Yeah?

AUDIENCE: [INAUDIBLE] You mentioned the film can be more specific [INAUDIBLE]. Would it make sense-- would it be possible to actually have different kinds of sensors that would be specific for different kinds of photography in digital?

ANKIT MOHAN: It's hard for you to change sensors once you have a sensor and it's baked in.

AUDIENCE: Yeah, I mean, if we could change the sensor--

ANKIT MOHAN: Yes, yes. And some of the stuff that, I guess, at some point we'll talk about in this course is there has been work on how to make more flexible digital sensors, not just digital sensors, but making-- how do you make the whole camera more flexible so you can programmatically change those responses? And you could do something of that sort.

But it turns out that, for most photography, it doesn't matter that much. And just by doing things in Photoshop, if you have enough bit depth over there, it doesn't matter too much. But it does matter for things like remote sensing, where you need-- even between 400 and 700, there'll be 30 or 40-- they're divided into 30 or 40 different channels, which are almost completely orthogonal.

And so going back to what you were saying, if you look at the response curve of the human eye, even that has a huge overlap. So it's actually quite similar to this one. It's not [INAUDIBLE]. Any other questions?

So that was a very quick overview of what I thought would be useful for you to know about film photography in general. And what I'm going to talk about during this class is what's epsilon photography.

And this is a term that-- this is a term that Ramesh coined some time back. And the idea here is-- the goal of epsilon photography is to improve on film-based photography, not to try and do something new, but just how to do what we could already do with film, but do it better. And the way it's done in almost all cases is by taking multiple pictures or capturing more data.

So you capture multiple photos, each with slightly different camera parameters. And usually, the parameters that you vary are the exposure settings, the color settings, the spectrum settings, the focal settings, the camera position, and the direction in which it's looking, or even the scene illumination. So you change one of these settings. And you capture a whole number of images. And then you somehow use an algorithm to combine those images together. And you get one image that looks better than any one of those individual images. That's basically what epsilon photography is.

And there are a number of ways in which you can do this epsilon photography. And I'm going to go through each one of these one by one. You could do-- you could take multiple pictures over time. You could take one-- you can take one picture, save it, take a second picture, save it, take 10 different pictures, and then combine them together somehow.

Or you could do it over sensors. You could have 10 different cameras co-located at the same point and take one image, one picture with each camera at the same time. Or you could do epsilon over pixels. And that's-- I'll get back to that in a minute. Or you could do a combination of all of these.

So epsilon over time is something which is the most common. And it's what most photography manuals refer to as bracketing. And the idea of bracketing is a little different because, in the end, you end up using just one image. So when you're not sure of what the exposure should be or you're not sure of where you should focus, you take multiple images with slightly different exposures or slightly different focus settings or aperture settings.

And most cameras have inward features for doing this. So you just have to press the Shutter button once and it takes five images for you. And then when you go home, you can decide which one is the best and just use that. But epsilon in time is similar. You take multiple images. But then you use some algorithm or some-- something smart to combine those images together and get one resulting image.

So the case where it's the most commonly used is for high dynamic range photography. And I believe Ramesh talked about this. In last class, he mentioned it. So as I was saying earlier, that you need to have the correct exposure in order to get the image of a scene.

It turns out that, for many scenes, even if you have the correct exposure, you cannot capture everything that the scene contains. Your scene can have very bright parts, such as daylight, and very dark shadow regions. And the contrast ratio of these two can be as high as [INAUDIBLE]. And most cameras would not capture anything more than a ratio of about-- excuse me-- about 1,000.

So one way of going around this limitation is to capture a number of images, and then use an algorithm to combine all those images together and create what's called a high dynamic range image. And I'm sure we've all heard of this term. If you just go on Flickr and search for "high dynamic range images," you will get millions of pictures that people have captured using this technique, just capturing a bunch of images and putting them together. And there's been lots of research into how you should put these images together.

And it turns out that once you've done all of this, there's a related dual problem, which is how do you display that image. And I'll get back to that in a minute. But one way of displaying that is what's called tone mapping. And there is work on sophisticated algorithms on how do you compress a 12-bit or a 14-bit image back to an 8-bit image. And there's interesting work in that area, which we're not going to cover in this class.

Another example over epsilon over time is this example that I really like. This is-- I think-- I'm not sure-- but it's one of the first color images-- or it's from the set of one of the first color photographs produced. And it was by-- this guy's name I cannot pronounce. But he went around Russia in the early 20th century during-- I forget, but during the early 20th century. And he took a whole bunch of photos of people just living their lives-- going and farming, hunting, and just sitting, and things like that.

And then the way he took these pictures was he would take three images-- one with a red filter placed in front of the camera, then one with a green, and one with the blue filter. And then once he had processed his images and so on, he developed a projector which would project a red image, a green image, and a blue image on the same screen. So when you were viewing it, you would view a colored image.

And as recently as about 10 years ago, until about 10 years ago, there was just these films that were lying in the Library of Congress, which were then digitized and hand-aligned. And now, you can download all of these color images from their website. So this is the very simple case of epsilon in time. You just take three images with three different pictures in front of your camera.

Another example, which is actually used-- it's probably using this projector-- is that there's-- most DLP projectors have a color field, which stands in front of the DMD mirror. And part of this-- it's a little hard to see. But I think this is red, green, blue, and then it's green again. And I think this part is just white in order to increase the intensity of the-- maybe this is red and this is just transparent.

But when you have the red part of the wheel in front of the DMD, you project the red image. Then when you have the green part, you project the green image, and so on so that when you actually view the projected image-- and this happens really, really fast, that the eye integrates over time and actually gives you the full color image.

And one way to see this happening is if you take a camera and you just capture an image with a really fast shutter speed of about 1 over 1,000, you can actually sometimes get half the screen green, half the screen is blue. Or you can get really interesting and [INAUDIBLE] if you try to do that.

Now, this won't work if you have an LCD projector because an LCD projector actually uses a color LCD. And you get all the colors at the same time. It's actually spatially interpolated. It's a spatial sort of multiplexing instead of a temporal thing like this.

So this was doing epsilon over time. The next one is doing epsilon over sensors. And this usually means two things. You can either have multiple cameras or you can just have multiple sensors within the same camera. And multiple sensors within the same camera is what's popularly called a 3CCD imaging system. It's what's used in most high-end video cameras and camcorders.

And you have this trichroic lens with a prism, which actually-- depending on the index of refraction, the rays get-- they just pass through or they have total internal reflection. And so the red, green, and blue images are formed on three different sensors, which are exactly the same optical distance away from the scene. And so when you take all of those three images-- I think that image over there shows you have white light coming in from behind the prism. And then you have the green, blue, and red components getting separated as they go through.

And so at the same time, using three different sensors, you can capture the three color channels. So it's similar to putting the three filters in front of the camera, but it happens at the same time. So you can use this for moving objects and so on.

And yeah, so also the sensor itself is-- it usually has a very broad spectral response. So it actually responds to any incoming wavelength between 400 and 700 nanometers. It's only the prism that does the separation.

AUDIENCE: So why is this being used? Why is that not being used in digital SLRs?

ANKIT MOHAN: It's just big and clumsy. I mean, I think the question I would ask is the opposite. Why don't they use Bayer sensors in camcorders? And I'm not entirely sure why. I think this might be something that's just stuck around since the first camcorders were developed. And that was probably before Bayer pattern filters became popular.

AUDIENCE: [INAUDIBLE].

ANKIT MOHAN: Right. So probably those edges-- edge effects show up more in a video camera than they do in digital, just still photography. I'm not really sure why they're still in use. I mean, they're definitely better. They do give a higher resolution, as was pointed out.

So recently, Morgan McGuire and others at Mitsubishi Research, which is just across the street, they came up with this really-- they took this to the extreme. And they said, instead of just having three of them, why don't we have eight of them and make a whole tree of these kinds of multiple optically co-located cameras? And so they came up with a very interesting beam-splitter arrangement.

And each of these eight cameras are actually optically co-located at the same point. And so the image formed on each of them, if there was nothing else changed, would be exactly the same. But what this gives you the flexibility to do is now each of these eight can have different focus settings, for example. They can focus on a different plane. Or they can have a different color filter in front of it. And you can get eight different spectrum-- spectrum channels at the same time.

And what I think, on the right, it shows is just that he's shown-- he shines a laser through the cameras to align them to see that that ray of light actually goes into each one of those cameras. So he used this setup for-- or a simplified setup for doing matting or defocus matting, where he used, I think, two or three cameras to focus on the plane and one that's focused in the background in order to do a separation between what's in the front and what's in the background. But it's certainly something that can be used for various other things. It's basically the same concept as multiplexing over sensors.

Another way you can do this is simply by using camera arrays. And this was work, I believe, on one of the first camera arrays. And it was done at CSAIL. And it's epsilon-- the difference between camera arrays and a SAMP or 3CCD is that this imposes a certain epsilon on your setup, that there's always going to be epsilon in the camera position.

You can put other epsilon on top of it. You can have a different filter in front of these cameras or you can have a different focus setting on each of these cameras. But just by itself, it gives you an epsilon in the camera position. So each of the cameras in this camera array is actually-- they're not co-located. They're slightly translated from one another.

And that itself can be used to give interesting things, like I'll get back to that later in the talk. But this is another way of doing this kind of epsilon. And Stanford has their own version of camera arrays. And now, actually you can just buy a camera array, which is a 5 by 5 profusion camera. We have one upstairs, which-- it's one unit. And it actually has a very well-aligned and precisely calibrated camera-- array of cameras.

The last one is epsilon over pixels, where different pixels are actually capturing different information. And we already talked about this one, which is where the Bayer filter is essentially doing this. Each pixel has a different color filter in front of it. And [? Roarke ?] is going to talk about another technique a little later today, which is a very clever way of extending this and allowing you to do various other things without having to place the filters on the pixel itself or [INAUDIBLE] elsewhere, which is easier to do.

So going back to-- whose question was it? Someone asked, can you change the shape of the filters of the color response? I think you want to know? But [? Roarke ?] will shoe you a way of doing that by simply putting something in front of the lens.

And then you can have epsilon in multiple axes. So this is a very cute camera that also we have upstairs. It's got four lenses. And so it forms four images on the film. It's a film camera. And it also has four flashes.

The reason it has four flashes is not because of four different flashes, but because it can strobe them very quickly one after the other. So it opens one lens at a time. And when that lens is open, it strobes one of the flashes. So you get four images, which are from slightly different viewpoints and taken at a slightly different timing sensors. And there's a whole website full of creative stuff that people have done with these kind of cameras. But this is epsilon in time and also in the position of the camera, so in sensors. Yeah?

AUDIENCE: So is this on film?

ANKIT MOHAN: This is on film, yeah.

AUDIENCE: And so does it give you four pictures on the film? Or are they superimposed?

ANKIT MOHAN: They're not superimposed. They're four distinct pictures.

AUDIENCE: Right.

ANKIT MOHAN: So this is a work done by [? Srini ?] sometime back, which is-- it brings it all together in one nice package, where you can do all of these things together. And it's what they call generalized mosaicing. Are people aware of what mosaicing or how you capture a panorama? Basically, if I want to capture a panorama of this whole scene, I will take an image here, I would move, take an image here, move, take an image here, and take an image here, and then just stitch them together to create a mosaic, which has everything in here.

So what they came up with is instead of taking an image-- each image with the same setting, they put-- so that's the camera. They put a filter in front of the camera some distance between the scene and the camera itself. And this can be a filter which either has a ramp in neutral density gradient or different spectrums, different colors of polarization or even focus.

And you simply-- instead of taking one image at a time, you take a video as it's rotating. And from this, from the data that you capture, you can get either a high dynamic range of the whole scene, or a multispectral image of the whole scene, or an image that's focused at different points. So the way to think of this is that when you're taking an image here, different scene points-- something over there-- I'm going to get the blue channel of a pixel here. But when I rotate it, I'm going to get the green channel of the same pixel. I rotate it some more, I'm going to get the red channel of that pixel.

So you just do this complete panoramic motion. You have some missing data for the edges. But for anything in between, you'll be able to recover the complete information of this. That's why they call it generalized, because you could use it for any one of these things. And I think they also built a camera prototype like that which was more portable. And they just put this filter in front of the lens.

So we sort of already discussed this one, which is doing HDR capture by multiple images. You just take a whole bunch of images. And you can combine all that information together. And this is doing HDR over time.

So that was this one. So I just wanted to take the example of high dynamic range imaging and see how we can do this over time, over sensor, and over pixel. So this is the first one, which is doing HDR over time. This is the second one, which is the generalized mosaicing, which is sort of in between the three settings that you just put this filter in front of your camera, and you rotate the camera, and take a video.

This is using multiple detectors. This is similar to the SAM or the 3CCD setup that we saw. You have multiple cameras that are optically co-located that take multiple-- that take images at the same time with different filters in front of them. So they have different exposure settings on each one of them. And as you can see, each of these areas has had lots of work done in them.

So the last one is this-- using what's called assorted pixels. I think that that's a more generalized way of the other two. Yeah, so it's similar to the Bayer mosaic. But instead of having just an RGB Bayer mosaic, they actually had two or three different levels of neutral density filters also placed over each pixel.

So each pixel-- this blue is different from that blue in the amount of light that it captures. And so you can do an interpretation in the color. And you can also do interpolation in intensity in order to get a high resolution image.

And this is essentially what they call assorted pixel. But it's more like generalized Bayer pattern, Bayer filtering. And you can put polarization filters also on top of it. Or you could have other colors other than just RGB. And again, [? Roarke ?] is going to talk more about this later.

And this was actually work done at Columbia in collaboration with Sony. And Sony actually made a camera that did this. It was only a prototype. It was never sold. But they were able to get this picture from an assorted camera-- assorted pixels camera, which has a much higher dynamic range and captures all three color channels at the same time.

This is another example of doing this over time, this high dynamic range capture. And the way this was done is that you place an LCD in front of the sensor which is of a much lower resolution than the sensor itself. You capture the information on the sensor. And you see certain pixels go-- get saturated. They're too bright.

And in the next timestamp, you actually put a darker patch over those pixels so that you compensate for the brightness. And going through, iterating through this, you can get an image which is-- which has a lower dynamic range on the sensor. But once you combine that with the information that you pumped into the LCD, you can then recover a high dynamic range image.

The reason I wanted to just mention that is for this one, where this is actually a work done by Wolfgang Heidrich, who gave a talk in our group some time back at University of British Columbia in Canada. And this is stuff that's been brought over by Dolby. And they're actually using this stuff in Samsung LCDs now.

But this is a way of generating a high dynamic range display. And the setup is very similar-- it's very similar to the previous one. And it's actually very simple, in that instead of just using a projector and projecting on the screen, which is what a Bayer projector display does, they have a projector, and then they have an LCD in front of it. So they have two layers of control. And they get twice or the squared of what they had earlier as much control over the dynamic range of what they can display.

And so they can control the-- I think this is a very early prototype over here. So they have two layers of LCDs-- so one inside the projector itself and one placed over there. But you can also just do it with two layers of actual physical LCDs. And it turns out that the LCD at the back has to be-- can be of a much lower resolution than the LCD in the front. And just using that, you can get very high dynamic range.

And I think most HDTVs and so on, also they have this thing where they can dim the backlight. So they get this very high dynamic contrast, which is sort of confusing. But it's essentially not just modulating the LCD, but also modulating the backlight. That's essentially what this is doing, but not just the whole backlight. Backlight is modulated differently in different parts of the screen. Yeah?

AUDIENCE: Well, how come you can use a lower resolution for the front LCD?

ANKIT MOHAN: No, not for the front, for the back LCD. For the front, you still need full resolution because the back LCD is essentially acting like a backlight. I think they also had a diffuser here, which anyways reduces the resolution of the back LCD. Otherwise, you might get weird edges and so on.

So that was a little about high dynamic range. The next thing I wanted to talk about is what we discussed earlier, this concept of focus setting and how we can extend the depth of field. There are many applications where you want a very large depth of field, like I said, for example, landscape photography. You want the tree next to you and the mountain faraway to be in focus.

But as we discussed, in order to do that, you need to stop down the lens. You need to have a very small aperture, which means you are going to get very little light coming into the camera. And so your noise goes up. Or you might have things move while you're taking your exposure.

So a number of techniques have been proposed over the past 30 or 40 years, especially in the area of microscopy, in how can we extend the depth of field while still keeping the aperture size reasonably large. And there is recent work done in the area of light field cameras. And one I didn't write over here is tape recording, which I'm sure Ramesh will come back to later, which also allows you to extend the depth of field while still having a large aperture.

So the first technique that's the most interesting one here is what's called focal stacks. And the idea is very simple. It's basically epsilon over time again. And you take multiple images focused at different planes.

So for example, you have this ant sitting under a microscope or this-- when you focus in the foreground, you get things in the foreground are in focus, but its hind legs and the rest of the body is out of focus. If you focus on the background, you get focus-- the foreground is not in focus.

So what they instead did was they took a whole series of images. And I'm going to just flip through them. They're focused at different planes. So I think that's about 10 or 12 images that-- you can capture all of those. And you can do this because the object or the scene, in this case, is static just over time. Or you could use the SAMP kind of setup, where you capture all these images at the same time, but each camera is focused at a different plane.

And then you combine all of this information together in order to create one image that's completely all in focus. And so a similar [INAUDIBLE] from University of Washington proposes a very interesting and clever technique of how you can combine them together of finding out regions in-- so each image has certain parts that are in focus. So you do a contrast-based estimation of what parts are in focus. And that's what's shown on the right.

And then you do a gradient domain merging of various parts together. So the end result doesn't have any weird discontinuities. And it looks nice and smooth. And everything is in focus. And I think Ramesh is, at some other point, going to talk about this technique itself in more detail.

But what I wanted to mention is more of the focal stack. You can just take a whole bunch of images focused at different planes. And then you can put them all together in order to get one all in focus image.

AUDIENCE: Was the analysis done in a computer vision? Or did it use [INAUDIBLE] from the camera or the actual picture was taken?

ANKIT MOHAN: You mean this data?

AUDIENCE: No, the way to combine the images.

ANKIT MOHAN: So the combining the images was actually a different technique. That's this gradient-based merging technique, where you have stuff from one image and you have stuff from another image that you want to put together. But if you just cut that image and put it here, you're going to get weird discontinuities. And colors are going to be different.

But it turns out if you do that in the gradient domain and then do a [? percent ?] solver to integrate the image back, you are going to get a nice smooth image. And all the error is going to get distributed as noise throughout the image. So that was what the technique initially proposed. And they just used that technique for this focal stack in order to get this.

I think the example-- I'm sure Ramesh is going to talk about this at some point. The example they had in the paper was that you have a scene like this. And if you take a picture from here, you might get someone not looking at the camera, or someone caught yawning, or someone is-- just has a bad face. If you take 10 such images, each one of those images is going to have some people who are OK and some people who don't look OK. But there's not going to be one single image that has everyone looking at the camera.

So they developed this technique in order to combine all of those images together to get one image that has everyone looking at the camera the way you want it to be and still look like a picture that came from a camera. And I should have put that in somewhere. But it's called digital photo mosaic, I think. And it's [INAUDIBLE]. Yeah, a photo montage, digital photo montage. So you can do a similar thing with a light field camera. I don't know if Ramesh has introduced the concept of light fields yet. Has he?

AUDIENCE: Yes.

ANKIT MOHAN: Yes? So a light field basically captures all the information coming into a camera. And the way it's traditionally usually done is by putting a microlens array in front of the camera sensor. If you don't understand that, that's fine. I'm sure he'll go into more detail.

But what you can get from the light field is you can extract the focal stack out of the light field, and then do basically what we did in the previous case and extend the depth of field if you want. So a light field-- essentially, what's, important to remember is the light field can be used to extract the focal stack if needed. So that's another way of extending the depth of field.

There was another paper recently, which again, I did not put over here, which is interesting because it was from Sam Hasinoff at CSAIL, where he-- instead of taking one image with one aperture setting, he claims that if you take multiple images with two or three different aperture settings, and then you combine them, you are going to get much-- your total exposure time is going to be much shorter, and you're going to get less noise, and so on.

So that's yet another way of combining. It's similar to focal stacks. It's just [INAUDIBLE] focal stacks more smartly because focal stacks focuses at each plane. And what he said is that you can find an optimal set of planes that you need to focus in order to get the best results.

So that was extending the depth of field. The opposite problem is how do you make the depth of field shallower. And this is something that comes naturally when you use an SLR camera with a large aperture lens. You have a very shallow depth of field if you open the aperture all the way out. And so your main object is in sharp focus, but the background is nice and blurred. But if you use a small point and shoot camera, it's very hard to get that kind of an effect since your aperture size of the camera is usually very small.

And so the question is, how can you still use a small aperture camera and get results like the one at the top? So there's been a couple of-- three or four papers in this area that try to attempt to solve this problem. The first one is again from CSAIl by Fredo Durand and his student.

And what they did was-- you start with an input image. And then from just one single image, they estimate the depth of each point. So sorry, before I go into that, the reason why this is a hard problem is because firstly just from this image, it's hard to estimate what's in focus and what's not just by looking at the contrast. And even if you can get that, even if you know that the foreground is in focus and the background is out of focus, if you have multiple layers in the background, each one of those layers are going to be out of focus-- more or less out of focus, depending on the depth or the distance from the plane of focus.

So it's hard to estimate the 3D shape or the 3D structure of this from just a single image. It's much easier to do it from two images. But what this paper, "Defocus Magnification," did was they tried to estimate the 3D structure of the scene from just one image, and then use that one-- that 3D structure and the image that was captured in order to increase the defocus by simply applying a spatially very blurred filter on the image. So you can see the background is more out of focus than the image that they took over here.

And this is not really epsilon photography. I just put this here because it's important to the overall structure. But this is more of an image processing technique than anything else.

Another way of doing this is what's called synthetic aperture photography. And this is something that was proposed by Marc Levoy's group at Stanford. But it's something that's more general. And it's been used in radars and so on for a long time.

The idea is actually really simple, that what you want to do is you want to simulate a large aperture lens, such as the one shown here. However, you don't have the physical resources to create a large aperture lens. But what you can do is create a number of small aperture lenses, and then somehow take the information coming from each of those lenses and computationally combine them in order to simulate a large aperture lens.

And it's essentially what-- one way of thinking of this is from a light field camera kind of point of view or just a camera array. But if you just think of it simply, you can combine each of these rays coming together into the lens if you can find out what those rays are and get what you would have gotten from just this one large lens. And now, if you look at a different point in space, you combine a different set of rays, and you get the intensity corresponding to that point. So it's essentially doing this-- what gets done with a large aperture lens in optics. You're doing that computationally by combining all these rays.

So this is what their setup looked like. This is one of the setups. I think they had five or six different optical configurations. But this one allows them to see-- so they used this in order to focus on something that was behind a bush. So let me see. Yeah, so that's what you would get with just one image.

And what's on the right, if the video plays, is-- OK, that's weird. But anyways, so you can simply, by computationally combining the rays-- and you'll actually learn how to do this kind of computation. I think it's in one of your assignments also. It's probably the next assignment, where you combine information from multiple cameras. And by simply shifting and adding, you can focus on a different plane.

OK, so maybe I don't have the video for this. It's just you can focus behind on a different plane. And anything in front goes out of focus. And since your depth of field is so shallow because you have this large synthetic aperture, everything over here actually goes out of focus and just blurs out. And you can see behind the foliage in this case.

AUDIENCE: Do you know if the bushes are moving in time? Does that--

ANKIT MOHAN: I don't think it matters because it takes just one image from each camera. It's a camera array. So it's not over time. It's over sensors.

But that's a good point. You could actually do this by taking a camera, and moving it around, and taking multiple images, which is exactly what the next paper does. This is actually work by Professor [INAUDIBLE] who was a visiting professor in our group last year and his students in Japan.

And they generalized this thing to instead of having a fixed, rigid camera array, you can just take a camera, and move it around, and take multiple images. And then using computer vision techniques to line up various synced components, they're able to get results kind of like that. So this is just with, I think, three or four images that he took just by moving a camera and just taking random images without any structure or any sort of calibration or anything. And from that, he's able to focus now-- in this case, focusing on the foreground and the background is defocused. And in that case focusing on the background. And the clock in the front is in focus.

This is also-- it's similar to the camera array thing, except you don't need a camera array. You can just take one camera. And over time, you can move it. So in this case, if the scene was changing, you would have problems reconstructing it.

Finally, the last technique I want to talk about is something that I worked on last year and really quickly try to go through this. I think I might have put too many slides in for this. But the idea over here is to do it more optically rather than computationally.

And the basic idea is that instead of keeping the camera and lens static while you're taking an image, you intentionally move the camera lens and sensor. And that's what we call image destabilization. You move the lens and the sensor synchronously with one another during the exposure.

And so to give you an intuition for how or why this works-- so once again, going back to the image we had in the beginning, if you have a plane in focus that's plane A, all the rays coming from a point on plane-- from the point A get focused on A prime. And the rays coming from B get focused on a point, which is a little in front of the sensor. So you get this defocused blur on the sensor. And the size of this blur on the size of the lens aperture. If you reduce the size of the aperture, it goes down. If you have a pinhole, then you just get one ray going through, which is what we saw in the beginning.

So now if you take the pinhole and you translate the pinhole over time, then you're going to get a blur over here which corresponds to the motion of this pinhole coming from the point B and similarly point from point A. Now, what's interesting is if you compare these two, they're not the same size. They're actually-- they have a different size. And the size ratio actually depends on the distance between A and B, and the distance between the points, and the pinhole and the sensor.

So what we do instead is while moving the pinhole, we also move the sensor, but we move the sensor such that the ratio-- such that one of the points actually remains fixed on the sensor. And the other point produces a blur. So now, we've taken-- I haven't showed the actual motion. But through this sequence of images, point A was always focused on the same point on the sensor. Point B was focused on a different point. And so point B results in this blur which point A does not.

And you can use this sort of a setup, where you have the lens and the camera in two different planes. And the camera is moving at a different velocity than the lens. And by the ratio of the velocities, you can focus at a different plane in front of that camera.

So this is an image you will get with just a small aperture lens, I think F22 or something, where everything is in focus. And this is the image you get using our technique, where just the thing in between-- this is still the same lens. You get something which is in the focus in the middle. And everything in front and behind goes more and more out of focus by simply translating the camera lens and the sensor over the exposure time.

And so the advantage of this is that you can simulate a large aperture lens or an SLR lens using a small camera lens. So you don't have to spend-- many of those lenses cost more than $1,000. So you can use a [? $30 ?] lens in order to create a similar effect as that produced by a $1,000 lens. But the disadvantage is that you need this motion. And so you need some space around the lens where it can translate, which is not all that hard because if you look at the space around the lens, most of it is wasted and not really used.

OK, at this point, maybe it might be a good idea to just switch to [? Roarke ?] and see what his technique lets us do. I have a couple of other things here. But I think we can do that after his stuff. But I don't know, do you want to break for 5-10 minutes first?

AUDIENCE: Yeah.

ANKIT MOHAN: Sure, so we can break for 5 minutes. And at 2:45, yeah--

AUDIENCE: [INAUDIBLE]. There is something called crosstalk. And there's optical crosstalk [INAUDIBLE] crosstalk, meaning you end up having those [INAUDIBLE] side by side. And you can have some photons that actually pass [INAUDIBLE].

ANKIT MOHAN: Do you want to draw that image?

AUDIENCE: I mean, if you go back to the [INAUDIBLE] then you can actually--

ANKIT MOHAN: I think what you're saying is that you have [INAUDIBLE].

AUDIENCE: Red and green in there. The photons can actually cross say-- yeah, exactly this. And it's going to [INAUDIBLE]. And then also, there is the fact that-- I mean, photons or different wavelengths have different energies. So it might be that one blue cross is Jupiter. But we'll set up a wall a little bit and actually [INAUDIBLE].

ANKIT MOHAN: Right.

AUDIENCE: So this is another kind of [INAUDIBLE]. And if you like-- I mean, yet another reason for that-- I mean, the refraction varies with wavelength. So in fact, if you have a single plane here, you're going to have photons with different wavelengths of focus [INAUDIBLE]. So this is what he calls chromatic variation. So you have to add extra objects in order to focus it.

ANKIT MOHAN: Right.

AUDIENCE: But people know how [INAUDIBLE] one of the major reasons why-- for high-end applications like manifold applications [INAUDIBLE].

ANKIT MOHAN: Right, of course, yeah, and also for the Foveon sensor has similar advantages in some of the cases. And I mean, of course, using a 3CCD sensor or multiple distinct sensors is always going to give you better results than the Bayer filter. It's going to be more expensive, perhaps.

But I think the question that I think you had was, why is it that we use 3CCD for video but not for still? Why does it matter more for video? And I'm not entirely sure.

AUDIENCE: You just get more light. You need more light for video in general. Otherwise-- the Bayer filter functions by blocking the light.

ANKIT MOHAN: No, but then even--

AUDIENCE: 3CCD actually splits up the light. So you use all the light.

ANKIT MOHAN: That's true, yeah. Yeah, so you've got three times as much light.