


Image Signal Processing
Professor A. N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture 32
Shape from Focus - Examples

(Refer Slide Time 0:17)





Shape from focus

Focus operators: Look purpose: SML

$$SML(g(x,y)) = \sum_{i=2}^{x+N} \sum_{j=2}^{y+N} ML(g(i,j))$$

$N=8, 2$





Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

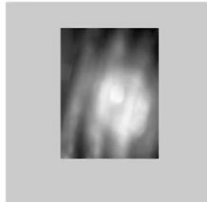
(Shape from Focus - Examples)

So I thought I will show you some of the outputs that you will likely get if you use SFF.

(Refer Slide Time 0:21)

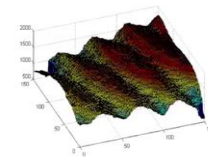


SFF using Tensor Voting




Stack generated using a microscope


Click on the image for video



Reconstructed 3-D shape



Focused image



Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

So you saw this already, right? So this example we have already seen, okay, where we have a bunch of wires and then it is kept under a microscope. And then you move, capture a bunch of frames, this is what is being shown here. And then you run the shape of focus algorithm.

Now, on this side, you also see an image that looks like it is all focused. If you see this image, it does not look like any of these images here. Does not look in the sense that every one of them looks like they have had some problem in terms of blurring, somewhere or the other there is blur, whereas this one looks uniformly in focus.

So this image is not something that the microscope gave. So this image is not something that this microscope can ever give because of the fact that there is this object which has a 3D variation, and your microscope will always end up giving images which will always be blurred somewhere or the other. So which means that this is an image that we reconstructed, how do you think we arrived at this focused image? We do not use any software at all, we have all that we need to kind of build this focused image up. So how would we do it?

Student: () (1:40)

Professor: No, it is much more simpler than that.


Student: () (1:49)

Professor: Exactly. So yeah, you wanted to say something?

Student: () (1:54)

Professor: Correct. Correct. So it simply boils down to, now let me also write that a little bit. So one of the other things that you would like to do is actually is this.

(Refer Slide Time 2:10)



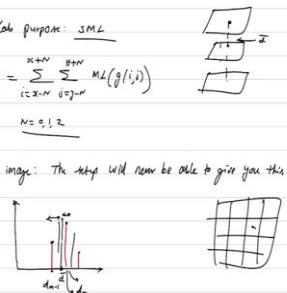
Shape from focus


Focus operators: Lab purpose: SML

$$SML(g(z,y)) = \sum_{(z,x=N)}^{(z,x=M)} \sum_{(y=0)}^{(y=M)} M_L(g(i,j))$$

$M = 0, 1, 2$

Reconstruction of the focused image: The steps will now be able to give you this!!





Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

So the other thing that is of interest is actually a construction or reconstruction, typically it is called reconstructing or recovering, reconstruction of the focused image. And the important thing to note is that the image setup that you have used will never be able to give you this, will never be able to give you a focused image give you this.

And therefore it is something that we have a stack so we need to walk through the stack to find out where exactly are those focused intensities. So we can go back to the plot that we had for the focus measure. So we saw that after we plot the focus measure, so we said that we could have these values right there.

And then we said that we will do a Gaussian interpolation and let us say that we get our d bar somewhere there, I mean, that is why let us say that we can also get our optimal value and whatever it is right maybe that is the FP for that okay. That is the max value that we get when we do the Gaussian fitting.

Now, what you could do is, you could simply look at the frame which is actually nearest to d bar. So in this case, the frame rate nearest d bar is the frame that you would have captured here, if this was the other way around, then no, we have to see that whether it is closer to this frame, this frame or that frame, whichever it happens to be closest to so we can just go there.

And since we know that the all the frames are aligned, so all that we need to do is for that pixel, find out right from basically which frame should we actually pick up the intensity. So it is like saying that in this whole stack, suppose I find out that for this pixel, so this pixel let us say comes into focus somewhere right above this.

So, your d bar is somewhere there, so we have not hit the d bar exactly, but then we are just slightly away. So, like this plot is showing here so we are somewhere there. So, simply what we could do is we could simply pick up the intensity value that is there that currently exists in the frame that is actually closest to d bar, and simply pick that and actually paste that in your output grid. So, in your output image where you want to reconstruct the focus image, you can simply copy that pixel here.

Another thing that you could also do is a linear interpolation, where maybe you want to make use of this as well as this value if the peak lies here. If the peak lies there, if it is on this side, then maybe you might want to use this intensity, you know your d bar so you know that this is some $d_m - 1$, you know that right this guy is a d_m . So you can look at a ΔI I mean you can look at a difference right between these two, these are separated by d bar minus d_m

minus 1, these two are separated by whatever d_m minus d_{bar} and therefore, we can do a linear interpolation just like we did earlier. But this is in 1D, what we did earlier for image interpolation was on a 2D grid, this would be simply a 1D plot is a simple 1D interpolation, so we can make use of both the intensities, and do a weighted averaging, and simply copy that pixel here and repeat this everywhere. I repeat this for every d_{bar} , so you are kind of traversing the grid.

So you are kind of going from $1 \times$ to x, y to another x, y and everywhere you are trying to look for d_{bar} , and on the one hand, you can keep as you are constructing your depth plot, and simultaneously on the other hand, you can have an output grid that will actually produce a focused image. Right? Yeah.

Student: (())(5:48)

Professor: The intensity? No.

Student: The Focus (())(5:56)

Professor: Well, you could do that, but well yeah. So I said you have two alternatives; one is you can still write interpolate or you can simply take the one right, which is closer. So you are saying that why not I pick the one that has the maximum value or something.

Student: Instead of finding the d_{bar}

Professor: Yes, instead of finding the d_{bar} .

Student: Instead of finding the d_{bar} can we take the (())(6:26)?

Professor: Ah, well, I mean, yeah, in this case, it turns out that that it is going to happen that way. Yeah, in a way, what he is saying is as good as saying that it, wherever my kind of say f_m peaks that is where I am going to see pick up. But if you wanted to do some kind of a linear interpolation or something, you would still want to know how do you do the interpolation, but typically it is better to bank on multiple values than to bank on one value, because the speak if it is slightly noisy or something we do not know that peak value in that frame could be that it is affected by noise.

So, instead of simply copying that value if you have D_{bar} , then you would at least know where you are and how much of a weight should we have for that, if your d_{bar} reinforces the fact that actually the focus is occurring very close to that, then you can be more sure that you

can add, you can actually put more weight for that intensity, but let us say if it so happens, that your that your d bar is equally away from these two peaks, it makes more sense to take both intensities. So that is it, you can still use d bar. I mean, I would say use it because of the fact that you will at least know what you are doing is right or not because instead of blindly picking whatever is a maximum you may still want to know whether I am right and kind of doing so.

If you are closer to this you can simply pick that in which case you can assume that probably because, but interpolation why not because you have two values better bank on both because some averaging will happen and averaging always counters noise to some extent. We do not know how much noise and all is there in these images, but if there is noise and to counter noise, it is always good to average.

And because these are all aligned, see things are not aligned, then it is a problem, right? averaging and all does not make sense. But because that pixel is sitting right there in the other frame, it is right, right up, right below, right center, it is always good to average because then you will actually counter noise to some extent. Yeah.

Student: (())(8:25)

Professor: What image? Rough?

Student: (())(8:30)

Professor: No, why would it be rough? Because you are only kind of say picking...

Student: (())(8:40) different pixels from different images.

Professor: No, no. All that it means is, you have the stack, and all that you are saying is right, I mean for example, one location is coming in focus here. The other location is probably coming in focus there. Since you are only interested in the focus intensities, it is okay. I mean, you will have to do that, I mean, you cannot you cannot pick one frame to get all the focus intensities because you cannot do that anyway. So you will have to walk through the stack to find out from where I should pick and that will not give you any sort of an unpleasant image or something.

Student: (())(9:25) maximum intensity.

Professor: Maximum, not the maximum intensity you can interpolate. Okay, yeah, go ahead.

Student: Is it only the maximum intensity in which focused images lag in the sense it has to be 0. Due to some blurring, if it is converted to grey scale some 50 or like that then it cannot be focused further.

Professor: No, no, which is a reason why no we are not looking at the intensity value. So when I am doing this FM plot, those are not intensities What I am plotting as FM values, I am saying that it is sum operator at sum modified Laplacian, or at any of those operators that we mentioned.

So these operators are not just looking at the grayscale value and all, they are looking at their entire neighborhood, they are trying to make some sense out of how this whole region is around that pixel. And based upon that we are saying that there is a maximum, maximum does not mean you have a maximum intensity, even though we typically believe that is something is in focus, it is supposed to be the maximum intensity, but then what will also happen is, you may have a higher intensity in a frame above some frame from where you are actually drawing the intensity, eventually.

What it amounts to saying is that you could have noise in that particular frame, and because of noise maybe you are looking at a higher intensity, but then it does not mean that the SML will go and pick that pixel, it will work around the region and try to find out where it supposedly is coming into focus, so that need not be the frame where this pixel has the highest intensity.

So in fact you can actually probably verify that, but then of course, you like to add noise and so on, because the images that we give you are not so noisy, but if you try to add noise and then blur and so on, you may be able to show that it is not always true that a pixel where it has a maximum intensity is what will eventually that is what you will pick, there is no guarantee.

You might pick something that is above the frame or below the frame, depending upon what this operator says. So, those peaks are not corresponding to intensities, those peaks correspond to what we think about the sharpness of that region. Any other questions?

Student: (())(11:34)

Professor: For Delta d? Yeah. Okay, so, so I thought, I will show the results, then I will come back to what are all the weaknesses and strengths and so on, then we will talk about delta.

Typically, Δt is chosen such that it should at least be equal to DOF, which is depth of field. So for example, if I have a microscope that has a depth of field of 25 microns, so having a Δd less than 25 microns, it does not make sense because if two points are separated by let us say more than 24 microns that is when there will be a change in blur. So, the sort of a ballpark figure is Δd should be at least greater than equal to your depth of field of the lens.

You can go higher than that if you feel that if I just take DOF then I have to capture too many frames and that increases my computation. But choosing anything less than DOF may not help you, you will end up capturing too many frames and then across range you may not see a difference that you want to actually see.

So anyway, I will kind of come to that in a minute, after we go through it. Any other doubts there? All these are valid doubts. They are very, very, very valid. Anything else? Okay, so let me show you some of these output results. Okay, so we were here, right?

(Refer Slide Time 13:01)

The slide is titled "SFF using Tensor Voting". It features the NPTEL logo in the top left. The main content is divided into three parts:

- Stack generated using a microscope:** A square image showing a stack of blurred grayscale images of a textured surface. Below it is the text "Stack generated using a microscope" and a small link "Click on the image for video".
- Reconstructed 3-D shape:** A 3D surface plot showing the topography of the surface, with a color gradient from blue (low) to red (high). The axes are labeled with numerical values.
- Focused image:** A single grayscale image showing the same textured surface as the stack, but now in sharp focus.

 At the bottom left, there is a small inset video of Prof. A.N. Rajagopalan speaking. At the bottom center, his name and affiliation are listed: "Prof. A.N. Rajagopalan, Department of Electrical Engineering, IIT Madras". At the bottom right, the text "(Shape from Focus - Examples)" is displayed.


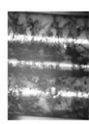
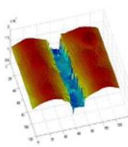
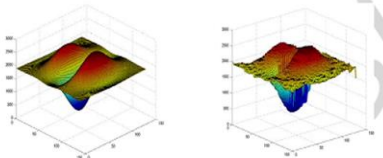

So we are showing you this. So this focus image is actually constructed like that. Of course, it is hard to say what might have been the original image because we do not have the original with us. I mean there is no way to validate this, there is no way to validate, because we do not have the true focused image with us. Going by intensity is not really a smart thing to do. Because just because of noise, you may have higher intensity there. But that does not mean that that is how the original intensity was like.

You should rather go with this focus operator, which gives you a sense of how that region is evolving, and where that region comes in because see what happens when something comes into focus, typically, it sets out a whole set of set of these pixels in and around because objects are like that. Objects are not like noisy like image, you can think of images, I mean, you can have random noise in images, but you will never see random noise pattern in a kind of 3D world, it is unthinkable.

So typically, there is always something called a generic prior that they say everybody uses right across noise filtering for all not just images everywhere What it means is that locally things look similar, there is a fundamental fact. Fundamental fact of nature is that locally things look similar.

So, this is a generic prior that you can exploit anytime anywhere. Seldom will you go wrong unless like there is a discontinuity or something where you can break down. Okay, so this is that plot right and again, how do you really verify that this is true and all, there is a bunch of other things that you have to do to verify. But right now, I am just showing you how those how those output results look like.

(Refer Slide Time 14:44)

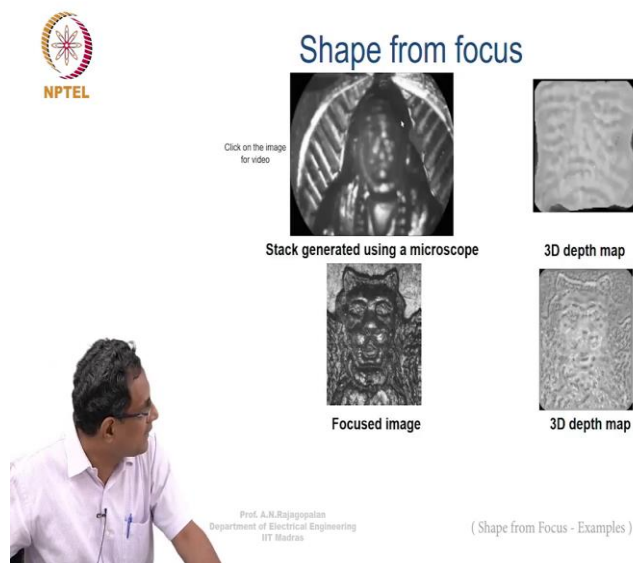


Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

And then here is another example, there is a trench which you want to say pick up and here is a focus image corresponding to that, this is a simulated situation. The first one was real.

(Refer Slide Time 14:53)



Can you see the face out like this of face alone cropped. More than that, if you look at this, so this have you seen somewhere? Where have you seen this? In the coin, in the coin you have this or what is that lion? Ashoka this one. So if you see this, so this is the focused image by the way, but now can you kind of see a depth map here, I mean, you can see these two things and then you can see the faces of lion and so on.

So all these things that would be next to impossible to do using stereo and all, because these are shiny objects getting feature correspondence is next to impossible, having a baseline, of course you have these microscopes that are stereo microscopes, people do have those, but again to image objects like these, which are shiny and all it is very hard and to do those feature correspondences.

Whereas you know, doing something like this will still give you that is somewhat reasonable number. Okay, then, before I go further, I wanted to show you one more thing, right? What if you have to ignore the parallax effect and what if you did not have a tele centric setup and you still attempted something like this then what would happen?

So you would expect things to not work out that well because that is a phenomenon called, we call it pixel migration. What it means is that if you take a camera and move forward, see just as when you translate a camera we said that is normally called stereo, when you translate in plain, that is like stereo, that is a regular stereo.

There is something called axial stereo, axial stereo means you take a camera and walk along the optical axis, you are translating, but you are not translating like this. This is in plain

translation, you translate along the optical axis that is called axial stereo, axial stereo is seldom used because the parallax that you get, you have to move a lot in order to see a parallax that you can actually identify, reasonably well. But as far as shape from focus is concerned, what happens is, if you try to walk with the camera, when you walk forward and you do not have this kind of a telecentricity in your camera most camera, most of our regular cameras do not have that kind of a telecentric setup.

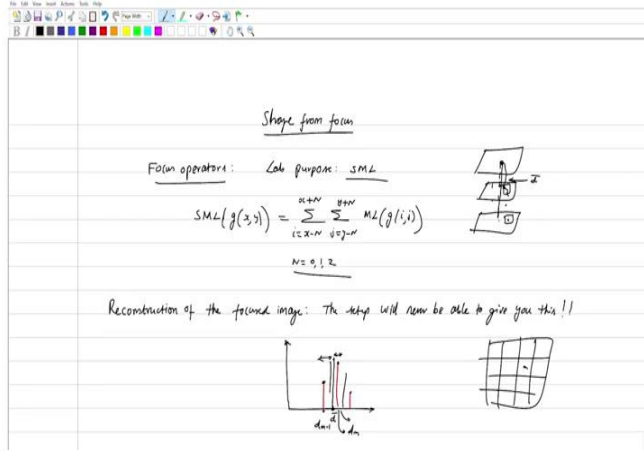

Therefore, when you move right what will happen is each pixel will now start to move across the frames as a function of (z) in depth. It is not just scaling, if you think that these are all scaled versions of each other, it is not correct, because when you move, there is an axial stereo going on.

Only thing is that pixel motion is much less than what happens when you translate, when you translate, if you see the parallax effect that will be much higher. It is also the reason why you do not find stereo where they have cannot to translate this way. They will typically have a stereo rig where they keep the baseline separated like that. And then the opposite axis this way, not that way.

So it is not like I turn on camera on here, and then I turn the next camera on there or maybe whatever, you do not have a setup like that. So typically, it is like there is a baseline, they are separated in play. But if you walk like this, that also is stereo that is called axial stereo. But what happens is right in this case, because it is not a stereo because you are getting these blurred frames.

So you get an effect that is both, there is a combination of axial stereo plus the blur. But axial stereo itself is not really such as sort of a great cue for depth. So what happens is, but then if you tend to ignore that, and if you say that I will still run my operator through the stack like that, what could have happened is, what could have possibly happened is.

(Refer Slide Time: 18:26)



Shape from focus

Focus operators: Lab purpose: SML

$$SML(g(z, y)) = \sum_{(x, z) \in \Omega} \sum_{(i, j) \in \Omega} M_L(g(i, j))$$

$M = 0, 1, 2$

Reconstruction of the focused image: The setup will now be able to give you this!!

Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

So here, we assume that all these points are aligned, right that need not be the case anymore. So this point could have moved slightly away here. And then in the other frame it could have moved somewhat a little more, and so on. So unless you know where it is going, I mean if you know where it is going, you could apply the operator there, but then to know where it is going, you should know its depth, which is exactly what we are trying to find out. So we do not know where it is going. So unless you account for that motion and if you simply blindly say that I will go ahead and apply my focus operator here, you will end up going wrong. Do you accept that? So if there is a parallax delay, let it be axial, but still there is going to be some parallax.

So is this called pixel migration, where these pixels start to migrate depending upon where they are in the scene, each pixel will migrate in its own way.

Student: (())(19:15) delta d.

Professor: That is okay, no delta t is known but then for that delta d each scene point will move in its...

Student: (())(19:23)

Professor: No no, you will not know that, to solve for the depth you need to do this focus measure plot, do the focus measure plot here assuming that the same pixel exists at the same location across the frames, but that is no longer true because they are moving and you do not know where they are going. So for example, you do not know that next time you should be

putting the operator here and you end for this frame, you should be putting the operator there you do not know that because the moment you translate rate, there is a parallax, parallax is not scaling.

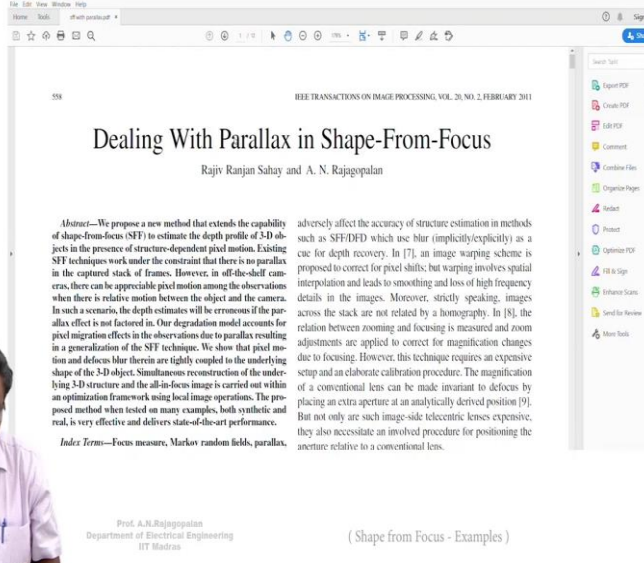

See, for example, when you translate like this, when you translate like this, we do not say that there is a simple translation in the scene. When you cannot have a homography that will simply translate one to kind of map onto the other, because of the fact that different points are existing at their own depth, and therefore each one moves independently of the other.

Same thing with this, see, the whole idea is that the only case when you can have a homography relating multiple views is when the camera does not move. Whenever you translate, whether you translate in plane or whether you translate out of plane, there is always a parallax, except that the parallax that is visible more visible is the one that when you do in play, that is one that gives you a better cue for depth if you are using stereo.

In this case, because (())(20:41) from focus has constructed this way that you actually move along the optical axis. So, what you encounter is actual stereo. But, but then because of the fact that we do not know the original depth, we do not know where these guys are going. So if you simply blindly apply a focus operator, then you are actually mapping wrong, you are not looking at the same pixel anymore because something else has migrated there and you are looking at its measure, then you are looking at something else that has migrated into the next frame and you are looking at its measure and therefore you are not likely to do well.

So right so we had one of my students this is again bold work, but I just wanted to share that work with us right, So where we had handled this one SFF with parallax. So I will show you those example.

(Refer Slide Time 21:28)



IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 2, FEBRUARY 2011

Dealing With Parallax in Shape-From-Focus

Rajiv Ranjan Sahay and A. N. Rajagopalan

Abstract—We propose a new method that extends the capability of shape-from-focus (SFF) to estimate the depth profile of 3-D objects in the presence of structure-dependent pixel motion. Existing SFF techniques work under the constraint that there is no parallax in the captured stack of frames. However, in all-the-shelf cameras, there can be appreciable pixel motion among the observations when there is relative motion between the object and the camera. In such a scenario, the depth estimates will be erroneous if the parallax effect is not factored in. Our degradation model accounts for pixel migration effects in the observations due to parallax resulting in a generalization of the SFF technique. We show that pixel motion and defocus blur therein are tightly coupled to the underlying shape of the 3-D object. Simultaneous reconstruction of the underlying 3-D structure and the all-in-focus image is carried out within an optimization framework using local image operations. The proposed method when tested on many examples, both synthetic and real, is very effective and delivers state-of-the-art performance.

Index Terms—Focus measure, Markov random fields, parallax.

Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

It's just interesting to see what happens, is this full view? I do not know how to get the full view. So let me just go down I will just show you how those how those images look like.

(Refer Slide Time 21:40)

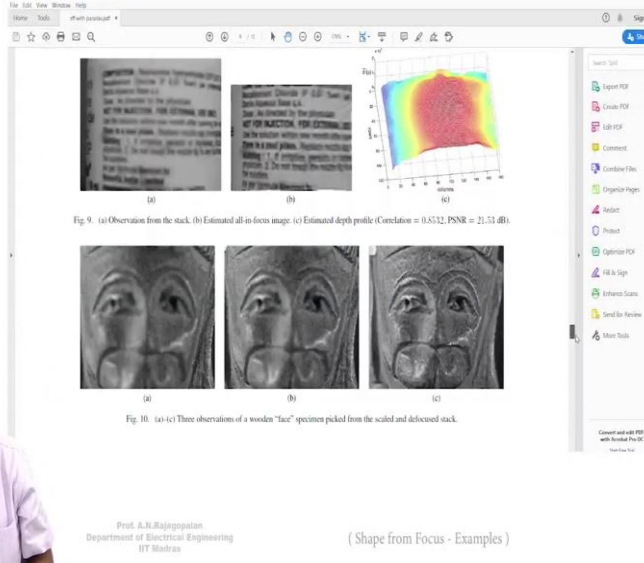



Fig. 9. (a) Observation from the stack. (b) Estimated all-in-focus image. (c) Estimated depth profile (Correlation = 0.532; PSNR = 21.53 dB).


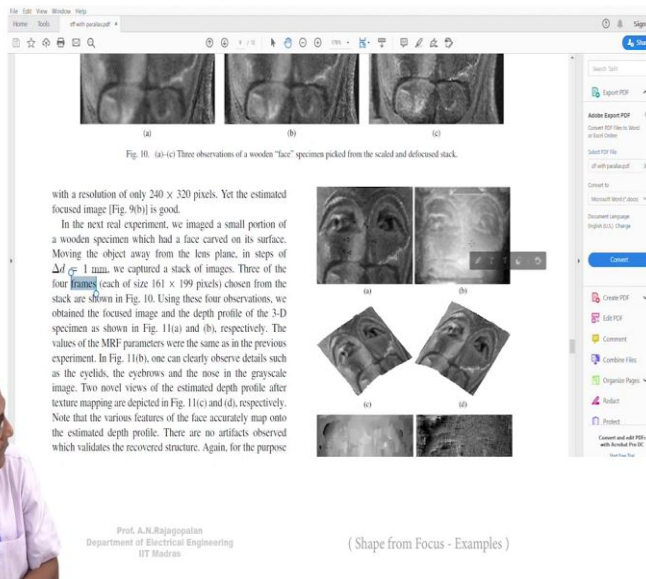
Fig. 10. (a)-(c) Three observations of a wooden "face" specimen picked from the scaled and defocused stack.

Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

See this, okay. Now this is actually a statue, this is like a handheld wooden statue. And he had his camera and he was walking, this is not a microscope anymore because we wanted to see how this works. If you will actually take a camera and walk.

(Refer Slide Time 22:02)

with a resolution of only 240×320 pixels. Yet the estimated focused image [Fig. 9(b)] is good.

In the next real experiment, we imaged a small portion of a wooden specimen which had a face carved on its surface. Moving the object away from the lens plane, in steps of $\Delta d \approx 1 \text{ mm}$, we captured a stack of images. Three of the four frames (each of size 161×199 pixels) chosen from the stack are shown in Fig. 10. Using these four observations, we obtained the focused image and the depth profile of the 3-D specimen as shown in Fig. 11(a) and (b), respectively. The values of the MRF parameters were the same as in the previous experiment. In Fig. 11(b), one can clearly observe details such as the eyelids, the eyebrows and the nose in the grayscale image. Two novel views of the estimated depth profile after texture mapping are depicted in Fig. 11(c) and (d), respectively. Note that the various features of the face accurately map onto the estimated depth profile. There are no artifacts observed which validates the recovered structure. Again, for the purpose

Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

So if you see, if you get a look at look at this depth map, this is how it looks like and it looks okay, despite the fact that the camera actually moved and then here is a focused image corresponding to that. If we did not take parallax into account then see what you get, see this nose that has gotten squished, see what this actually means is that the nose should have come straight, and because of the fact that these pixels are migrating you are actually planting at the planting wrong d bar now.

This is what will happen if you do not get this, if you do not account for this parallax effect. I just wanted to show this image so that you get an understanding that this face looks like a contorted face, it does not look like the original face anymore.

(Refer Slide Time 23:02)


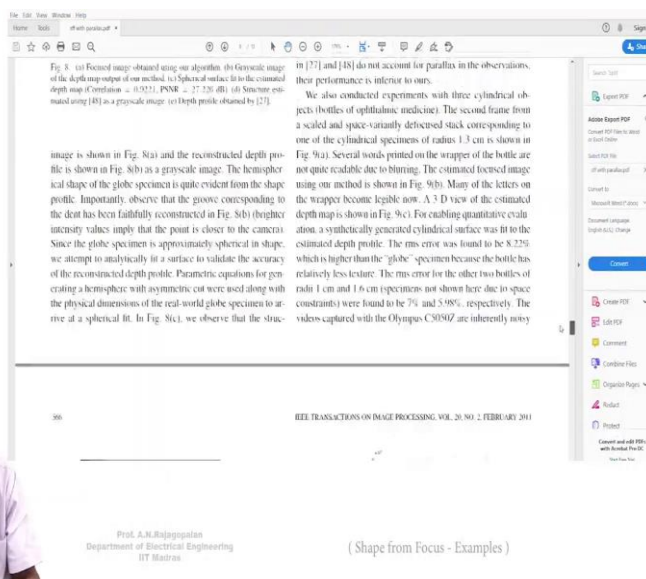



image is shown in Fig. 8(a) and the reconstructed depth profile is shown in Fig. 8(b) as a grayscale image. The hemispherical shape of the globe specimen is quite evident from the shape profile. Importantly, observe that the groove corresponding to the dent has been faithfully reconstructed in Fig. 8(b) (brighter intensity values imply that the point is closer to the camera). Since the globe specimen is approximately spherical in shape, we attempt to analytically fit a surface to validate the accuracy of the reconstructed depth profile. Parametric equations for generating a hemisphere with asymmetric cut were used along with the physical dimensions of the real-world globe specimen to arrive at a spherical fit. In Fig. 8(c), we observe that the structure in [27] and [18] do not account for parallax in the observations, their performance is inferior to ours.

We also conducted experiments with three cylindrical objects (bottles of ophthalmic medicine). The second frame from a scaled and space-variantly defocused stack corresponding to one of the cylindrical specimens of radius 1.3 cm is shown in Fig. 9(a). Several words printed on the wrapper of the bottle are not quite readable due to blurring. The estimated focused image using our method is shown in Fig. 9(b). Many of the letters on the wrapper become legible now. A 3-D view of the estimated depth map is shown in Fig. 9(c). For enabling quantitative evaluation, a synthetically generated cylindrical surface was fit to the estimated depth profile. The rms error was found to be 8.22%, which is higher than the "globe" specimen because the bottle has relatively less texture. The rms error for the other two bottles of radii 1 cm and 1.6 cm (specimens not shown here due to space constraints) were found to be 7% and 5.98%, respectively. The videos captured with the Olympus CS3040 are inherently noisy

Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

I think I have a few more examples like that. Yeah, but that one is the most. These are all ones that work. And the one that fails, right I think this is was one that we had shown. And then I also wanted to show you about this. See this, I mean here is a globe, this is just your globe, you have kind of multiple images and you see a dent here on this globe, there is some dent there.

(Refer Slide Time 23:13)

SARAY AND RAJAGOPALAN: DEALING WITH PARALLAX IN SHAPE FROM FOCUS 563

Fig. 7. Images of a globe specimen affected by space-variant blurring and magnification. (a)-(c) Three frames from the stack.

ture estimated with our method closely matches the spherically fit surface. The actual radius of the globe specimen is approximately 1 cm. While capturing the stack of frames, the background region of the specimen which was brought into focus in the first frame was approximately 2 mm above the circular edge of the hemispherical globe. The maximum height of the esti-

Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

Now, if you reconstruct it correctly, then what you will see is this dent here, which actually shows up pretty correctly, and you have a smooth surface, and then there is a small little dent, I think you can make out dent here.

(Refer Slide Time 23:38)

ture estimated with our method closely matches the spherically fit surface. The actual radius of the globe specimen is approximately 1 cm. While capturing the stack of frames, the background region of the specimen which was brought into focus in the first frame was approximately 2 mm above the circular edge of the hemispherical globe. The maximum height of the estimated depth profile is, therefore, 8 mm which is in accordance with the physically measured dimension of the globe. The rms error in d is only 4.3%. We again performed comparisons with recent methods. The depth map using the focus measure operator proposed in [48] is shown as a grayscale image in Fig. 8(d). We observe that the shape estimated is coarse and error-prone and the groove near the center is not at all visible. The output of [27] is given in Fig. 8(e). There are several holes in the estimated surface and the overall quality is not satisfactory. The correlation score (which is computed with respect to the fit surface) for [48] and [27] is about 0.6 while the PSNR value is about 18 dB. The work in [27] aims to recover the shape of the 3-D object by reconstructing the focused image surface (FIS). This is done by searching over the entire 3-D volume spanned by the stack of observations using the technique of dynamic programming. In the scenario considered in our paper, depth-dependent pixel motion (parallax) will result in a distorted and warped FIS. Also, more than one point on the 3-D specimen can map to a point on the sensor plane. Most importantly, pixels are not tracked in existing methods which precludes the possibility of unwarping

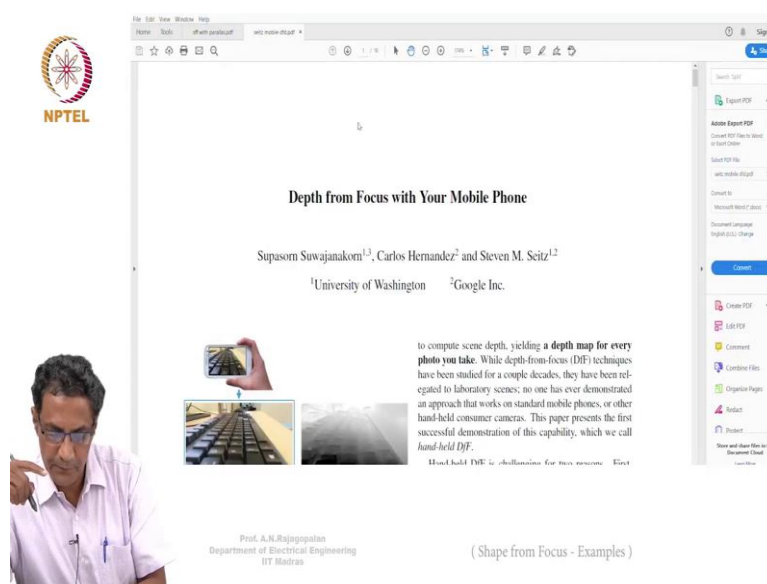
Prof. A.N. Rajagopalan
Department of Electrical Engineering
IIT Madras

(Shape from Focus - Examples)

But then if you do not take parallax into account, then what you end up with something like this image, you end up with something like that. So you can go and see very wrong, which is also the reason while I said hat time that people were still not talking about taking shape from focus out of the lab, because they knew that you had to solve all this. If you do this, you could still achieve. I mean if you wanted to escalate the challenge and if you wanted to take it out and do it, you still can do it, but then provided you take into account all these factors.

In fact, recently right I know recently I saw a paper in 2015, this is still pretty old, I mean this is which year was? This was 2011, right?


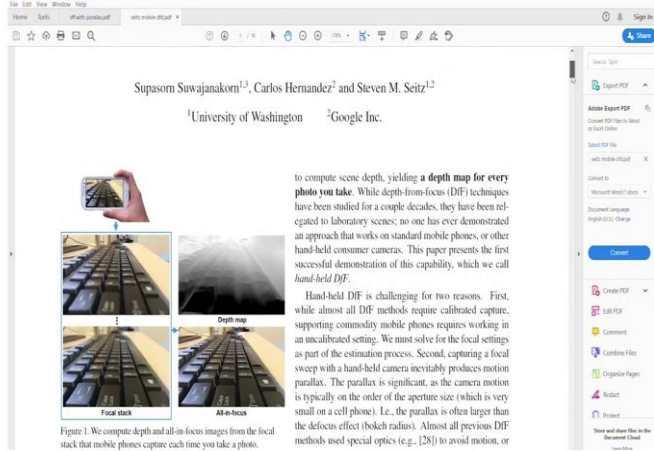
(Refer Slide Time 24:17)



The image shows a presentation slide with a background image of a PDF document. The PDF title is "Depth from Focus with Your Mobile Phone" by Supasorn Suwajanakorn^{1,3}, Carlos Hernandez² and Steven M. Seitz^{1,2}. The authors' affiliations are ¹University of Washington and ²Google Inc. The PDF content includes a small image of a hand holding a mobile phone over a keyboard, with a depth map overlay. The text in the PDF discusses hand-held depth from focus (DF) techniques. The NPTEL logo is visible in the top left corner of the slide. Below the PDF image, there is a small inset image of a man in a white shirt, likely the presenter, and the text "Prof. A.N.Rajagopalan, Department of Electrical Engineering, IIT Madras" and "(Shape from Focus - Examples)".

Now, look at the paper that came in 2015. So, he talks about depth from focus so shape from focus what he calls it depth from focus with your cell phone, a mobile phone. Again the same issue okay that one also for example, he shows that on a keyboard, you can actually do it okay.

(Refer Slide Time 24:36)

Supasorn Suwajanakorn^{1,3}, Carlos Hernandez² and Steven M. Seitz^{1,2}
¹University of Washington ²Google Inc.

to compute scene depth, yielding a **depth map for every photo you take**. While depth-from-focus (DfF) techniques have been studied for a couple decades, they have been relegated to laboratory scenes; no one has ever demonstrated an approach that works on standard mobile phones, or other hand-held consumer cameras. This paper presents the first successful demonstration of this capability, which we call *hand-held DfF*.

Hand-held DfF is challenging for two reasons. First, while almost all DfF methods require calibrated capture, supporting commodity mobile phones requires working in an uncalibrated setting. We must solve for the focal settings as part of the estimation process. Second, capturing a focal sweep with a hand-held camera inevitably produces motion parallax. The parallax is significant, as the camera motion is typically on the order of the aperture size (which is very small on a cell phone). I.e., the parallax is often larger than the defocus effect (bokeh radius). Almost all previous DfF methods used special optics (e.g., [28]) to avoid motion, or

Figure 1. We compute depth and all in-focus images from the focal stack that mobile phones capture each time you take a photo.

Prof. A.N.Rajagopalan
 Department of Electrical Engineering
 IIT Madras

(Shape from Focus - Examples)

So he appears a keyboard for which he gets a depth map right there. And then he shows some examples I mean pretty impressive, and these are people that will actually focus stack alignments. So if you see this work I mean, previous work corrected for magnification changes to scaling and translating or a similarity transform.

However, these global translations are inadequate for doing for local parallax and that is what I said. So you cannot account for that, to account for that you should know the depth, and which is like the chicken and egg problem, that if you knew the depth and when you know where to go but that is the whole point, we do not know what is the depth and then he shows some examples here.

(Refer Slide Time 25:15)


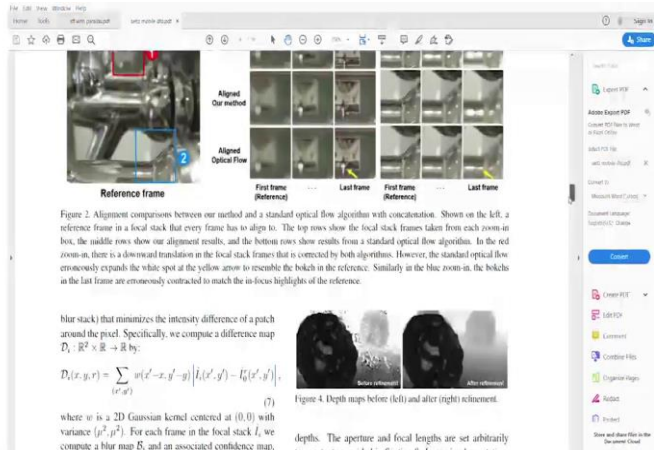



Figure 2. Alignment comparisons between our method and a standard optical flow algorithm with concatenation. Shown on the left, a reference frame in a focal stack that every frame has to align to. The top rows show the focal stack frames taken from each zoom-in box, the middle rows show our alignment results, and the bottom rows show results from a standard optical flow algorithm. In the red zoom-in, there is a downward translation in the focal stack frames that is corrected by both algorithms. However, the standard optical flow erroneously expands the white spot at the yellow arrow to resemble the bokeh in the reference. Similarly in the blue zoom-in, the bokeh in the last frame are erroneously contracted to match the in-focus highlights of the reference.

blurred stack) that minimizes the intensity difference of a patch around the pixel. Specifically, we compute a difference map $D: \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ by:

$$D_i(x, y, r) = \sum_{(x', y')} w(x' - x, y' - y) |I_i(x', y') - I_0(x', y')|$$

where w is a 2D Gaussian kernel centered at $(0, 0)$ with variance (μ^2, μ^2) . For each frame in the focal stack I_i we compute a blur map B_i and an associated confidence map.

Figure 4. Depth maps before (left) and after (right) refinement.

depths. The aperture and focal lengths are set arbitrarily to constants provided in Section 8. In our implementation

Prof. A.N.Rajagopalan
 Department of Electrical Engineering
 IIT Madras

(Shape from Focus - Examples)

Of course, I see these are not easy anymore okay, in the sense that the moment you start walking, then there is all this parallax and what did they actually do is what is called an optical flow. So, optical flow is sort of a generalization of your homography notion, homography helps you with one law, it helps you map the entire image right, this apply with that one law, homograph is like applying one linear transformation that you can apply on the entire image.

An optical flow is not like that, on optical flow you can think of trying to tell where each pixel is going independently of the other. So it is like saying that if you translate it so it is like something like if you computed a stereo, it will tell you where exactly each point went, optical flow will do something like that.

It will tell you about where each pixel has gone, in fact, it is used to compute stereo depth maps and so on. So it assumes that your intensity would not change across the frames much, there is the result sort of an illumination constancy constraint and all they put and they try to solve it some optimization problem.

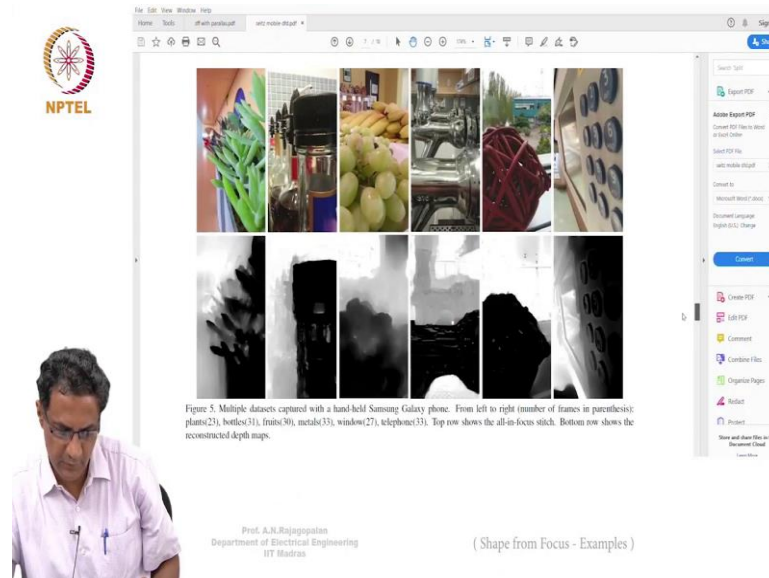
The thing is here, the optical stereo see, they could always ask when there is already a stereo cue why use blur then, if there is already stereo cue. The point is they are only using the stereo cue to the extent that they will not align the image because they are still banking on blur as a cue, because like I said if you want a stereo, you would translate this way and not this way.

This gives you very little of this parallax, but if the little amount is not taken care off then you will go wrong when you try to do this focus measure curve so anyway, it has to be accounted for. So finally this depth that he is computing is still from blur as a cue that is why he writes it as the depth from focus. It does not say depth from stereo. Although there is parallax inherently there is stereo running right across images, there is a cue but that cue is not really what he is using, he is using that to simply align it, he is not using that to, you could theoretically use both, but in this paper that is not what they do.

What is called really a cooperative defocus stereo that is called where let us say people use actually both cues to come up with a depth map and so on that is called a cooperative method. So here is using the optical flow to kind of do a compensation for this pixel motion. And because the pixel motion is compensated for, now we can run the autofocus operator and then

you are going to happy doing it and then okay, now, here are some examples. Let me just go down and show you some examples.

(Refer Slide Time 27:46)



So right so here is what is it so he says multiple, okay, data sets, captured with a handheld Samsung Galaxy. Little right number of frames, so plants, so he has got about 23 frames of that. So he has also given the number of frames, bottles, some of these mini frames fruits, metals, 33, window is I think metals is this, this is some window, I think, and then there is a phone a telephone, and for each one of them he constructs a depth map and that is what they show.

So, I just wanted to tell you that it will take some time for these things eventually come on the phone and so on, but as academicians we are always interested in knowing what all does an image reveal to you. No, the optical flow is what will actually, the way it works is optical flow will tell you that these two objects are at the same time because they are moving by the same amount, it is the other way around.

See, optical flow is computed across frames, you have a bunch of frames you compute the optical flow means what should be the motion, x, y motion that I should apply on each pixel that is on the first frame so that it aligns with the second. Now if it turns out that a pixel here and some other this one right a pixel there, if they both show that they are both moving by the same x, y amount that it means that they are both at the same depth. So, optical flow is something that will I mean use optical flow to kind of reconstruct, a depth map. And from

that they are you can actually make a domain, either you look at the optical flow, interpret it or you use the optical flow to construct a depth map and then you interpret it.

But the idea is still the same, that if I let a bunch of these pixels move by the same amount, as let us say some other bunch, then it means that right that these two objects, they could be the same object or they could be say different objects in the scene, but they are both at the same depth. But here, he is not using all of that, he is just doing some optical flow alignment, a crude alignment, so that he can run the focus measure of it. His idea is to extend different focus to a real world, where you can take a mobile phone and do it. So the idea is not really stereo is not the idea, and I do not want us to focus wrongly.

Stereo is not the idea here, stereo is only to do that (())(29:57) composition that you need in order to make these frames aligned, because then only you can run, so his depth cue is all due to blur and why would he capture that many frames; 30 frames, 35 frames for stereo you need two frames.

He is running he is capturing a whole stack now, capturing a stack but then he is aligning the stack because he understands that there is a parallax which needs to be accounted for. And once the frames are aligned then he runs a depth from focus algorithm. I have not read the paper fully to know what focus operators are but some focus operating must have been used then, okay?