

# Multi Focus Structured Light for Recovering Scene Shape and Global Illumination

Supreeth Achar and Srinivasa G. Narasimhan

Robotics Institute, Carnegie Mellon University

**Abstract.** Illumination defocus and global illumination effects are major challenges for active illumination scene recovery algorithms. Illumination defocus limits the working volume of projector-camera systems and global illumination can induce large errors in shape estimates. In this paper, we develop an algorithm for scene recovery in the presence of both defocus and global light transport effects such as interreflections and sub-surface scattering. Our method extends the working volume by using structured light patterns at multiple projector focus settings. A careful characterization of projector blur allows us to decode even partially out-of-focus patterns. This enables our algorithm to recover scene shape and the direct and global illumination components over a large depth of field while still using a relatively small number of images (typically 25-30). We demonstrate the effectiveness of our approach by recovering high quality depth maps of scenes containing objects made of optically challenging materials such as wax, marble, soap, colored glass and translucent plastic.

**Keywords:** Structured Light, Depth from Focus/Defocus, Global Light Transport

## 1 Introduction

Active illumination techniques that use projectors as programmable light sources have been applied to many problems in computer vision including depth recovery [15], surface normal estimation [9], BRDF estimation [4], separating direct and global components of illumination [12] and probing light transport [14, 13]. Because projectors have large apertures, most active illumination algorithms are limited to a shallow working volume in which the projector is in focus. This limits their applicability to scenarios where the scene relief is small and laboratory or industrial settings where the relative geometry between the scene and the projector-camera system can be carefully controlled. Additionally, global light transport effects like inter-reflections and sub-surface scattering are often ignored, but they can induce large, systematic errors in active shape recovery techniques like structured light and photometric stereo. Since global illumination effects are present in virtually all scenes to some extent, it is important to be able to account for their effects during shape recovery.

Pattern coding strategies like gray codes [10] degrade gracefully when illumination is defocused. In [6] patterns are designed such that they are all attenuated

to roughly the same extent by projector blur and [8] uses a sliding projector as the light source. These methods have some robustness to illumination blur but they do not explicitly model illumination blur and use a single projector focus setting. When depth variation in a scene is very large, the structured light patterns in some areas will be blurred too severely for shape recovery to be possible.

Global illumination can be handled in structured light depth recovery by using suitably designed illumination patterns. When the spatial frequency of the pattern is high compared to the frequency of the scene’s global illumination, the contribution of the global illumination to the observed radiance at each scene point becomes almost independent of the pattern [12]. Thus, using high frequency patterns can ameliorate problems caused by global light transport during shape recovery but it makes designing and decoding patterns more difficult as projector-pixel correspondences become ambiguous. Previous solutions to this ambiguity include using a very large number of patterns like in [3] or using techniques like phase unwrapping as was done in [6].

In this paper, we present a structured light algorithm that extends the working volume of the projector-camera system and is capable of producing high resolution depth maps over large working volumes. Our algorithm models both illumination defocus and global illumination effects like scattering and inter-reflection. In addition to a depth map of the scene, our algorithm recovers the direct and global components of illumination. It can be used to scan optically challenging materials like wax, marble and translucent plastic.

A naïve approach to expanding the depth of field would be to project a complete set of structured light patterns at each focus setting and then combine the resulting depth maps, but such an approach would require an inordinately large number of images. Our algorithm uses multiple focus settings but projects only a small number of patterns at each setting, keeping the overall number of images required small. The key insight of our method is that even an illumination pattern that is not in focus at a scene point can aid in pattern decoding, provided the projector blur kernel has been carefully characterized. We do this characterization by calibrating the projector to find the blur kernel as a function of scene point depth for each focus setting.

Previous work in structured light associates a fixed, depth independent code word with each projector pixel. In contrast, in our approach a projector pixel’s code has a defocus induced dependency on the depth of the point it is illuminating. To test a candidate projector-camera pixel correspondence hypothesis, we first compute the scene point depth implied by the hypothesis. This depth value can be used to predict the defocused illumination received by the scene point from the projector. If the candidate correspondence is correct, this projector output prediction should match well with the intensity values observed at the camera pixel. By using a range of focus settings, we ensure that at least some segment of a projector code is always in sharp focus at a point in the scene. Our algorithm seamlessly combines two complementary depth cues - triangulation based cues which provide high depth resolution but require sharp illumination focus (and thus suffer from narrow working ranges) and defocus based cues which

work over a large range of depths but provide coarse depth estimates. Our shape recovery algorithm is purely temporal and does not use spatial windows for decoding projector patterns which allows it to recover high quality depth maps with few artefacts at scene discontinuities. Once the shape has been predicted, we automatically have an estimate of the illumination received by each point of the scene in each image. We use this information to recover the direct and global components of illumination.

## 1.1 Related Work

The idea of exploiting projector defocus as a cue to recover shape was proposed in [19]. The approach involved estimating a measure of the projector pattern blur occurring at each illuminated scene point and mapping this measure to a depth value using a calibration function. They could recover accurate depth maps, but the fixed blur-to-depth mapping could not handle global light transport effects like sub-surface scattering. Gupta et al. [7] proposed a method to simultaneously model both projector defocus and global illumination. Their technique allows for depth recovery in the presence of global illumination and is based on the observation that unlike defocus blur, the blur induced by global light transport effects is almost independent of projector focus. Both [19] and [7] use colocated projector-camera systems and recover depth solely from focus/defocus cues.

In contrast, our approach does not use a colocated configuration but performs stereo triangulation between the camera and projector to measure depth. It has been shown that in principle, depth from defocus is similar to stereo triangulation [16] but focus/defocus cues have a baseline equal to the size of the aperture. Since triangulation cues are computed over the wider projector-camera baseline, our method is capable of producing more fine grained depth estimates. Although we do not use defocus cues explicitly (by using an illumination sharpness measure for instance), they are used implicitly as our projector codes are modeled as being depth dependent due to defocus. Previous work that combines camera defocus and stereo cues includes [18] and [17].

In structured light literature, some methods have been proposed to prevent errors due to global light transport. In [3] a large number of high frequency random bandpass illumination patterns were used to mitigate pattern decoding errors caused by inter reflections. In [5], global illumination effects are handled by designing a set of light pattern codes that work well with long range effects like inter reflections and a second set of patterns that work well with short range effects like sub-surface scattering. For scenes with both types of effects, ensembles of codes are generated and a voting scheme is used to estimate depth. Unlike [5], we do not seek to assign a binary code to each pixel and instead attempt to fit a model to the observed projector and camera values at a pixel, so we can use a single set of patterns to handle both types of global illumination effects.

Modulated phase shifting [2] modulates the sinusoids used in phase shifting by high frequency signals so both shape and global illumination of a scene can be recovered, but it does not consider the effects of illumination defocus.

Micro phase shifting [6] is a phase shifting variant that uses a narrow band set of high frequency sinusoids as the projected patterns. All the patterns are high frequency so the effects of global illumination are avoided. Because the patterns all have similar frequency they are attenuated similarly by projector defocus which lends some robustness to projector blurring. However, it should be noted that while this has some robustness to blur, it does not model defocus or use multiple focus settings so it can not handle large variations in scene depth.

In [11] illumination defocus is exploited towards a different end. Sinusoidal patterns are generated by projecting binary patterns with a defocused projector. DLP projectors can project binary patterns at very high frame rates which allows the phase shift algorithm to run in real time and recover dynamic scenes.

## 2 Modeling Image Formation and Illumination

Let  $S^t(x)$  be the value of the projected structured light pattern at time  $t$  at a scene point imaged by camera pixel  $x$ . The brightness  $I^t(x)$  observed by a camera pixel is a weighted sum of the direct illumination  $I_d(x)$  and the global illumination  $I_g(x)$  of the scene point. When the pattern  $S^t(x)$  has a high spatial frequency and a 50% duty cycle, it can be shown that the contribution of the global illumination to the observed brightness is approximately pattern independent and equal to  $\frac{1}{2}I_g(x)$  [12]. The pattern modulates the direct component so its contribution to the observed brightness is  $S^t(x)I_d(x)$ . Thus we have

$$I^t(x) = \frac{1}{2}I_g(x) + S^t(x)I_d(x) \quad (1)$$

We use  $\pi$  to denote the correspondence between projector pixels and camera pixels that illuminate/image the same scene point,  $p = \pi(x)$ . The projector value seen at time  $t$  at a scene point at depth  $z$  illuminated by projector pixel  $p$ , is a defocused version of the projector pattern value at that pixel  $L^t(p)$ . It has been shown that unlike camera defocus blur, the defocus blur kernel for a projector is scene independent in the sense that the kernel at a scene point depends only on the depth of the point, not on the geometry of the neighborhood surrounding the point [19]. Thus, without resorting to assumptions like local fronto-planarity, the effects of projector defocus blur can be modeled by convolving the projector pattern  $L^t(p)$  with a spatially varying blur kernel  $G(p, z, f)$ .

$$S^t(x) = \tilde{L}^t(\pi(x)) = (L^t * G(\pi(x), z, f))(\pi(x)) \quad (2)$$

The blur kernel  $G$  depends on the scene point depth  $z$  and the projector focus setting  $f$ . Additionally, we allow the function  $G$  to vary spatially with projector pixel coordinate as this helps better model the projector's optical aberrations.

Although the original high frequency illumination pattern  $L^t(p)$  is blurred due to defocus, Equation 1 still holds. The defocus blur reduces the amplitude of the high frequency components of the pattern but does introduce any low frequency content into the signal. We use a small aperture on the camera ( $f/10$  in our experiments) and model it as a pinhole camera that does not introduce any additional blurring due to camera defocus.



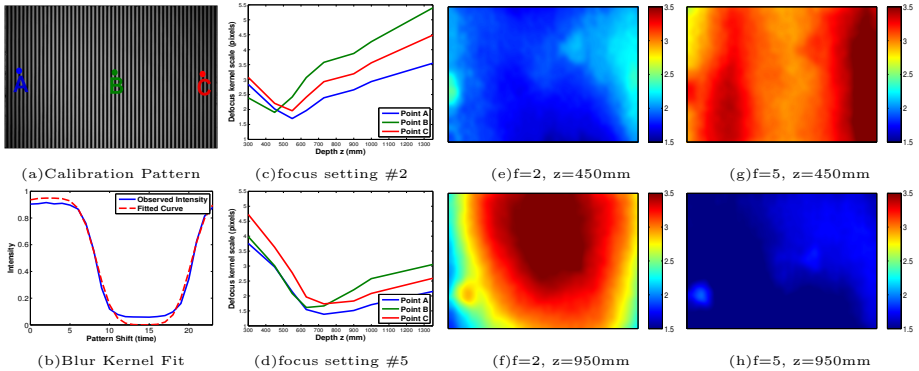


Fig. 1: Characterizing Projector Defocus: (a) - image of one of the square wave patterns for estimating projector blur. (b) is the temporal intensity profile at point B and the Gaussian smoothed square wave fit. (c) and (d) - the blur kernel scale  $\sigma$  for the projector pixels A,B and C for two different focus settings as the scene depth is varied. (e) to (h) - maps of blur scale  $\sigma$  across the projector image for different combinations of focus setting and scene point depth. The value of  $\sigma$  clearly varies across the image, especially when the projector is out of focus.

**Characterizing Projector Defocus** We model the projector blur using a spatially varying, isotropic Gaussian kernel. The scale of the blur kernel  $\sigma(p, z, f)$  is a function of projector pixel location  $p$ , the depth  $z$  of the scene point being illuminated and the current focus setting of the projector  $f$ . A more general class of kernels may allow for a more accurate characterization and allow more complex types of aberrations to be modeled, but we found that isotropic Gaussians were sufficient for our purpose.

For a given focus setting  $f$  and target depth  $z$  we estimate the defocus blur by projecting a sequence of patterns onto a planar target at depth  $z$ . The patterns are horizontal square waves with a period of 24 pixels (fig. 1a). We capture 24 images as the pattern translates one pixel at a time. The temporal profile of intensity values observed at a pixel is modeled as a square wave convolved by the blur kernel (fig. 1b). A similar scheme was used in [19] to estimate a mapping between illumination defocus and scene point depth. We find the blur kernel scale  $\sigma(p, z, f)$  that best fits the observed temporal profile for each projector pixel. This characterizes the defocus blur at one depth and focus setting (example  $\sigma$  maps are figs. 1e-1h). We repeat the process at a set of depths for each focus setting ( $f = 1, 2, \dots |F|$ ). We sample  $G(p, z, f)$  at every projector pixel  $p$  and focus setting  $f$ , but only sparsely in depth  $z$ . When queried for the blur kernel at a given focus setting and depth, we return the kernel at that focus for the nearest calibrated depth. Projector characterization is a one time, off line process.

### 3 Illumination Control and Image Acquisition

We recover shape and perform direct-global separation with a set of structured light patterns captured at different projector focus settings. The focus settings are chosen so that the projector’s plane of focus spans the entire working volume of the scene and that every part of the scene has at least one setting where the illumination is in reasonably good focus. For each of the  $F$  focus settings we capture a small number ( $N$ ) of structured light patterns. Although we have chosen to capture an equal number of patterns at each setting, this is not a requirement for the algorithm, the number of patterns used could be varied adaptively depending on the scene.

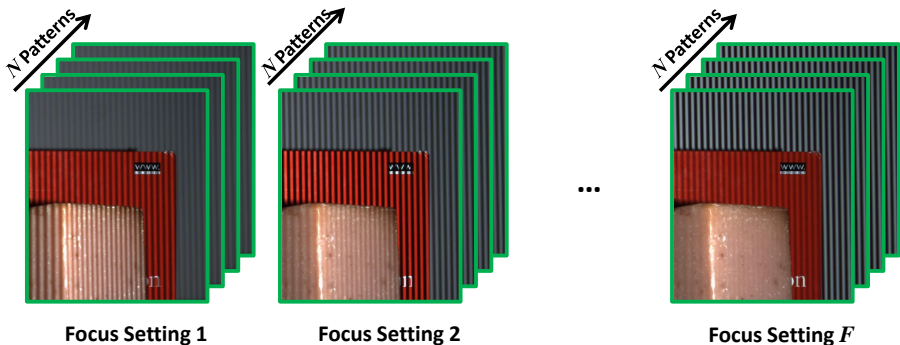


Fig. 2: Input to our Algorithm: We use binary stripe patterns of varying width. Unlike most other structured light algorithms that use a fixed focus setting on the projector, we change the focus setting to move the plane of focus backwards during the image capture process (the camera focus however remains fixed). We capture a total of  $T = F \times N$  images. In our experiments,  $T$  typically ranged between 20 and 30. As the figure illustrates, near by objects receive focused illumination in the earlier parts of the sequence and distant objects come into focus later on.

The structured light patterns we use are vertical binary stripes with randomly varying widths. Higher frequencies are less susceptible to global illumination errors, but very high frequency patterns are not displayed well by projectors. We let the period of the stripes in a pattern fluctuate between 10 and 14 pixels. This frequency range is high enough to prevent global illumination errors in most situations while still being in the band where contemporary projectors works effectively. We select patterns that do not correlate with each other to ensure that there is little redundancy between patterns.

## 4 Recovering Shape With Defocused Light Patterns

Temporal structured light algorithms project a series of patterns onto the scene, the time sequence of values emitted by a projector pixel form a code for that pixel. Camera-projector correspondence is established by finding the projector code that best matches the time sequence of intensity values observed at each camera pixel. The code can be binary (eg. gray codes) or continuous (eg. phase shifting), but it is assumed that the code for each projector pixel is independent of the scene geometry.

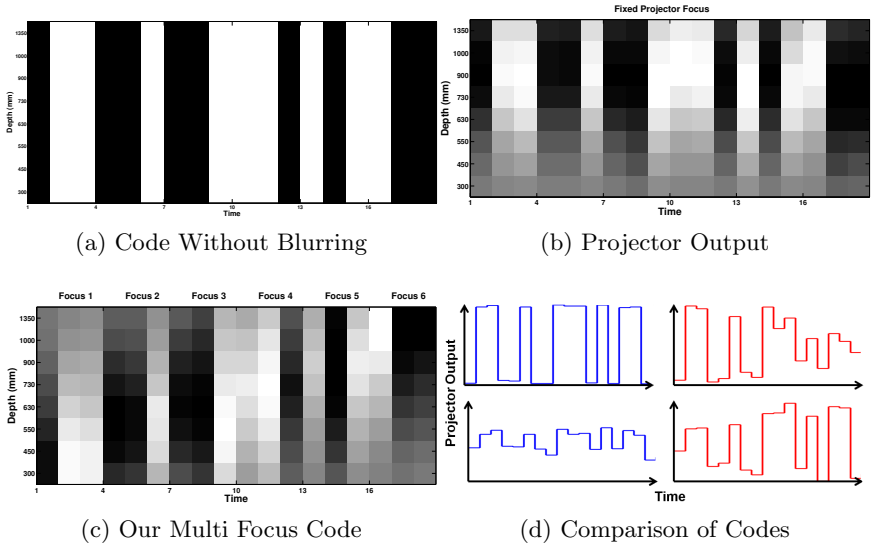


Fig. 3: Effect of Defocus on Codes: When illumination defocus is not modeled, the temporal code associated with a projector pixel (a horizontal cross section of (a)) is independent of depth. However, as (b) shows, outside a narrow working range, the actual appearance of the code is depth dependent. In (c), we use 6 focus settings and 3 patterns per focus. Using multiple focus settings allows us to expand the systems working volume. Also, we model illumination defocus so blurred codes do not cause errors. When regular codes are in focus, they work well (upper blue graph in (d)), however for scene points that are out of focus, contrast is very poor (lower blue graph in (d)). On the other hand, our multiple focus codes always have parts that are well focused and thus have high contrast (red graphs in (d)).

In contrast, our multi-focal structured light algorithm explicitly models illumination defocus effects, so a projector pixel’s code becomes a function of the depth of the scene point it is illuminating. This idea is illustrated in figure 3. It

is clear, that when the depth variation in a scene is large, defocus can strongly affect how a projector code manifests in a scene. As seen in figure 3b, when a pattern is out of focus, different values become difficult to distinguish. Decoding such a blurred pattern reliably with a defocus-blind algorithm would necessitate very high illumination power and high dynamic range on the imaging sensor. As figure 3c shows, even in large working volumes, some part of our code is always in sharp focus. This allows our method to work at lower illumination power levels over extended depths.

If we hypothesize that projector pixel  $p$  corresponds to camera pixel  $x$ , we can perform triangulation to find the scene point depth  $\tau_z(x, p)$  implied by the hypothesis. Using our defocus model (equation 2), we can then simulate the projector value  $\tilde{S}^t(x, p)$  that would be observed at this scene point by convolving the projector illumination pattern  $L^t$  with the defocus kernel,

$$\tilde{S}^t(x, p) = (L^t * G(p, \tau_z(x, p), f))(p) \quad (3)$$

Stacking together these values for all the patterns  $t = 1, \dots, T$  gives us the projector code for the pixel.

$$\tilde{S}(x, p) = [\tilde{S}^1(x, p), \tilde{S}^2(x, p), \dots, \tilde{S}^T(x, p)] \quad (4)$$

This projector code needs to be matched against the sequence of observed intensity at camera pixel  $I(x)$

$$I(x) = [I^1(x), I^2(x), \dots, I^T(x)] \quad (5)$$

If the hypothesis that pixel  $x$  and pixel  $p$  correspond to each other is correct, then by our illumination model (equation 1), there should be a linear relationship between the observed intensity values at the camera and the simulated projector values. We quantify the quality of a camera-projector correspondence hypothesis by computing the correlation coefficient between  $I(x)$  and  $\tilde{S}(x, p)$ . We can then find the projector pixel  $p = \pi(x)$  corresponding to camera pixel  $x$  by maximizing this correlation.

$$\pi(x) = \underset{p}{\operatorname{argmax}} \rho(I(x), \tilde{S}(x, p)) \quad (6)$$

We use a calibrated projector-camera system so with the epipolar constraint we limit the search in equation 6 to a 1D search along the epipolar line. We compute  $\rho(I(x), \tilde{S}(x, p))$  for every  $p$  along the epipolar line corresponding to a positive depth (Figure 4). To compute disparity to sub-pixel accuracy, we interpolate  $\rho$  scores between projector pixels when searching for maximae.

## 5 Recovering Direct and Global Illumination Components

Once the camera-projector correspondence map  $\pi$  has been estimated, we can compute  $S^t(x)$ , the projector pattern value at each camera pixel taking defocus

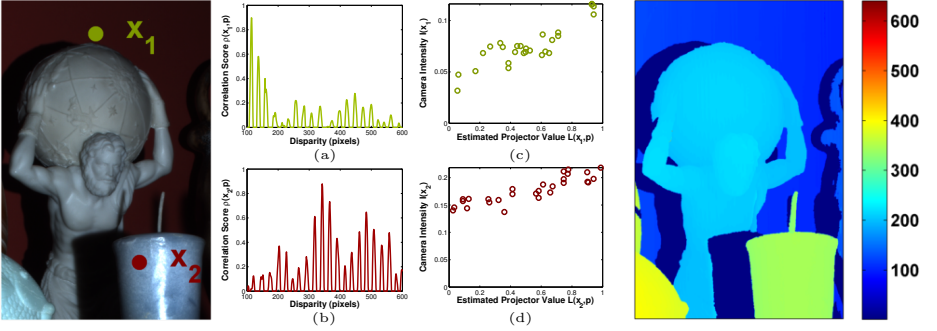


Fig. 4: Part of a scene (left) and the computed disparity map (right). Graph (a) shows the correlation score for point  $x_1$  as a function of disparity to the projector. The disparity that leads to the best match is 115. There are many peaks in the correlation score graph, but modeling of illumination blur causes the peaks to decay as we move away from the correct disparity value. Graph (c) shows the intensity observed by the camera against the simulated projector illumination value for the best disparity. Graph (b) and (d) are the same trends for point  $x_2$ . Because of strong sub-surface scattering at  $x_2$ , the global illumination component is large and the direct component is relatively small. This can be seen in (d).

blur into account using equation 2. Under the image formation model (equation 1), there is a linear relationship between the projected pattern value at a point  $S^t(x)$  and the brightness observed by the camera  $I^t(x)$ . Fitting a line to this model at each pixel yields estimates of the global and direct illumination. However, it is possible that even over the entire sequence of projected light patterns, some camera pixels would have seen only a small range of projector intensity values. There will be significant ambiguity while fitting a line to data at these pixels and hence there will be numerous plausible solutions to the direct-global separation. We resolve these ambiguities using a smoothness prior as was done in [1] by finding the direct image  $I_d$  and global image  $I_g$  that solve

$$\operatorname{argmin}_{I_d, I_g} \sum_{t \in T} \|I^t - \frac{1}{2}I_g - S \circ^t I_d\|_2^2 + \lambda_d TV(I_d) + \lambda_g TV(I_g) \quad (7)$$

$\lambda_d$  and  $\lambda_g$  are scalar parameters that weight the smoothness terms for the direct and global components respectively.  $A \circ B$  is the Hadamard (element-wise) product between  $A$  and  $B$ .  $TV(F)$  is the isotropic total variation of the function  $F(x, y)$

$$TV(F) = \sum_{\text{Domain}(F)} \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2} \quad (8)$$

Parts of the scene far away from the projector receive less light than regions close to the projector. As a result, there is a pronounced falloff in the recovered

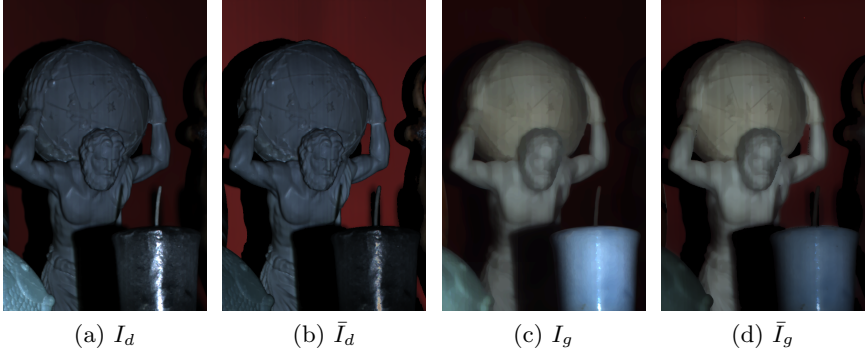


Fig. 5: The direct (a) and global (c) components of illumination estimated by our algorithm. Since we have recovered a depth map of the scene, we can also correct for projector fall off. This is particularly useful in scenes with large depth variations where objects in the background appear much darker than those in the foreground because they are further away from the light source. After accounting for the fall off, we get corrected estimates for the direct and global component illumination ( images (b) and (d) respectively).

direct and global illumination images. Because we have recovered scene geometry, we can roughly correct for this falloff by assuming it follows an inverse square relationship with depth. We can compute depth dependent correction factor  $K(x)$  at each pixel

$$K(x) = \frac{\alpha}{\tau_z^2(x, \pi(x))} \quad (9)$$

where  $\alpha$  is an (arbitrary) positive scale factor. We can then solve for the corrected direct and global illumination components ( $\bar{I}_d$  and  $\bar{I}_g$ ) by modifying equation 7:

$$\underset{I_d, \bar{I}_g}{\operatorname{argmin}} \sum_{t \in T} \|I^t - \frac{1}{2}K \circ \bar{I}_g - K \circ S \circ^t \bar{I}_d\|_2^2 + \lambda_d TV(\bar{I}_d) + \lambda_g TV(\bar{I}_g) \quad (10)$$

## 6 Results

**Experimental Setup** Our experimental setup consists of a projector and a camera mounted in a stereo configuration. We use a 500 lumen DLP projector with a resolution of  $1280 \times 800$  (InFocus IN1144). The camera is a  $2448 \times 2048$  color CCD (Point Gray Research GRAS-50S5-C). The camera is calibrated geometrically and radiometrically. The projector is calibrated geometrically at each focus setting and its blur kernel has been characterized (as described in Section 2). Our method uses only binary stripe patterns but we calibrated the projector radiometrically so that we could compare our method to micro phase

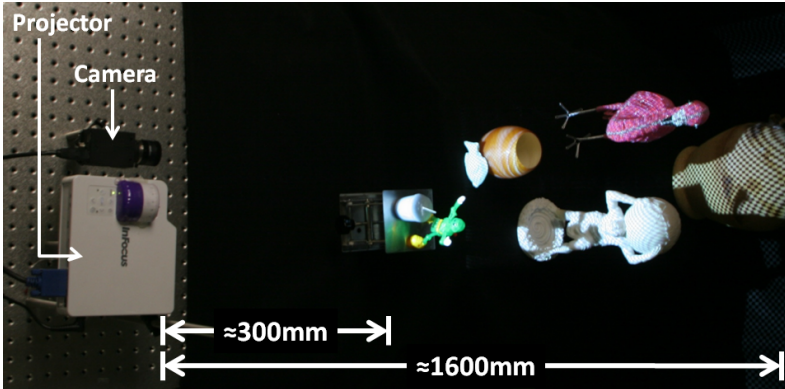


Fig. 6: Experimental Setup: With a small, 500 lumen DLP projector and a small number of images, we are able to scan scenes over a large working volume to recover accurate depth maps and perform illumination separation.

shifting [6]. The projector-camera baseline is fixed and known. Since the projector intrinsics change with focus setting, we correct for this by warping images before projecting them so that geometrically they all appear to be projected by a projector with fixed intrinsic parameters.

In our experiments the focus ring position was changed by hand and we used 4 positions ( $F = 4$ ). Between the shortest and longest focus settings, the working range of the system covers depths from  $350\text{mm}$  to  $1600\text{mm}$ . For all experiments the camera lens aperture was set to  $f/10$ , the exposure time was  $133\text{ms}$  and the camera was configured to capture 8 bit images.

## 6.1 Depth Recovery

We present results from our depth map recovery algorithm on two challenging scenes (top row in fig 7). Depth maps from our algorithm (second row in fig 7) were generated using 7 structured light patterns at each of 4 focal lengths, a total of 28 images. Our algorithm is able to recover accurate depth maps of both scenes with very few errors. We compare against a simple depth from illumination focus algorithm (bottom row) and micro phase shifting (third row).

The illumination defocus algorithm we compared against projects a shifted sequence of square waves (14 images) at each of 8 projector focus settings and then finds the focus setting at which each camera pixel’s illumination contrast was maximized. Each focus setting can be mapped to the depth of its corresponding plane of focus to find the depth map. Since the baseline for this method is limited to the aperture of the projector, the resulting depth estimates are coarse along  $z$  and tend to be inaccurate at large distances.

For the micro phase shifting experiments, we chose the high frequency (16 pixels per cycle) pattern set with 15 frequencies [6]. Micro phase shifting uses a fixed projector focus so we set the projector to be in focus in the middle of

the scenes. The total number of patterns used is 17. Using more patterns at this frequency is difficult because micro phase shifting requires all projected pattern frequencies to be in a narrow band.

Micro phase shift has some robustness to illumination blur but since it does not actually model defocus, it breaks down when the depth variation in a scene is too large. This is evident in scene 1 where the shape of green plastic robot in the foreground is not recovered by micro phase shifting. In comparison, our method is able to recover the robot. Our algorithm also works better on low albedo or poorly lit regions like the red funnel in scene 2. Since we change focus settings, there are always some images where the contrast of our projected illumination is high, so low signal to noise ratios are less of a problem for our algorithm.

The candle in scene 1 is very difficult to reconstruct as from some directions, it reflects almost no light directly back to the camera, almost all the observed radiance is due to sub-surface scattering. As a result, all the methods are unable to recover depth at some points on the candle surface.

## 6.2 Recovering Direct and Global Illumination

To obtain ground truth direct and global illumination images for our scenes, we projected 14 shifted stripe patterns at 8 projector focus settings and used the multiple focus separation technique proposed in [7]. The results presented for our method are computed using the same 28 images that were used to estimate the depth maps. Although our technique uses fewer images and involves a smoothing term, it generates output that is similar to the ground truth. Additionally, we can correct for the effects of projector fall off as demonstrated in Figure 8.

## 7 Discussion

We presented an algorithm that can reconstruct shape and recover direct and global illumination in a large working volume with a small number of images.

Our algorithm’s robustness to global illumination effects relies on the assumption used in [12]- the global illumination must vary slowly compared to the spatial frequency of the projected patterns. If this assumption does not hold, for example when specular interreflections occur, our method fails.

We currently use randomly chosen stripe patterns. Optimal pattern sets for structured light are usually derived by trying to maximize the distance between codes to minimize the chance of a decoding error. In our setting, we would have to consider the fact that defocus causes our codes to vary with depth. Also, for the direct-global component separation to work well, each pixel’s code word must contain a large range of projector intensity values. Carefully designed patterns may allow our algorithm to work well with fewer images.

**Acknowledgements:** This research was supported in parts by NSF grants IIS-1317749 and IIS-0964562 and ONR grant N00014-11-1-0295.



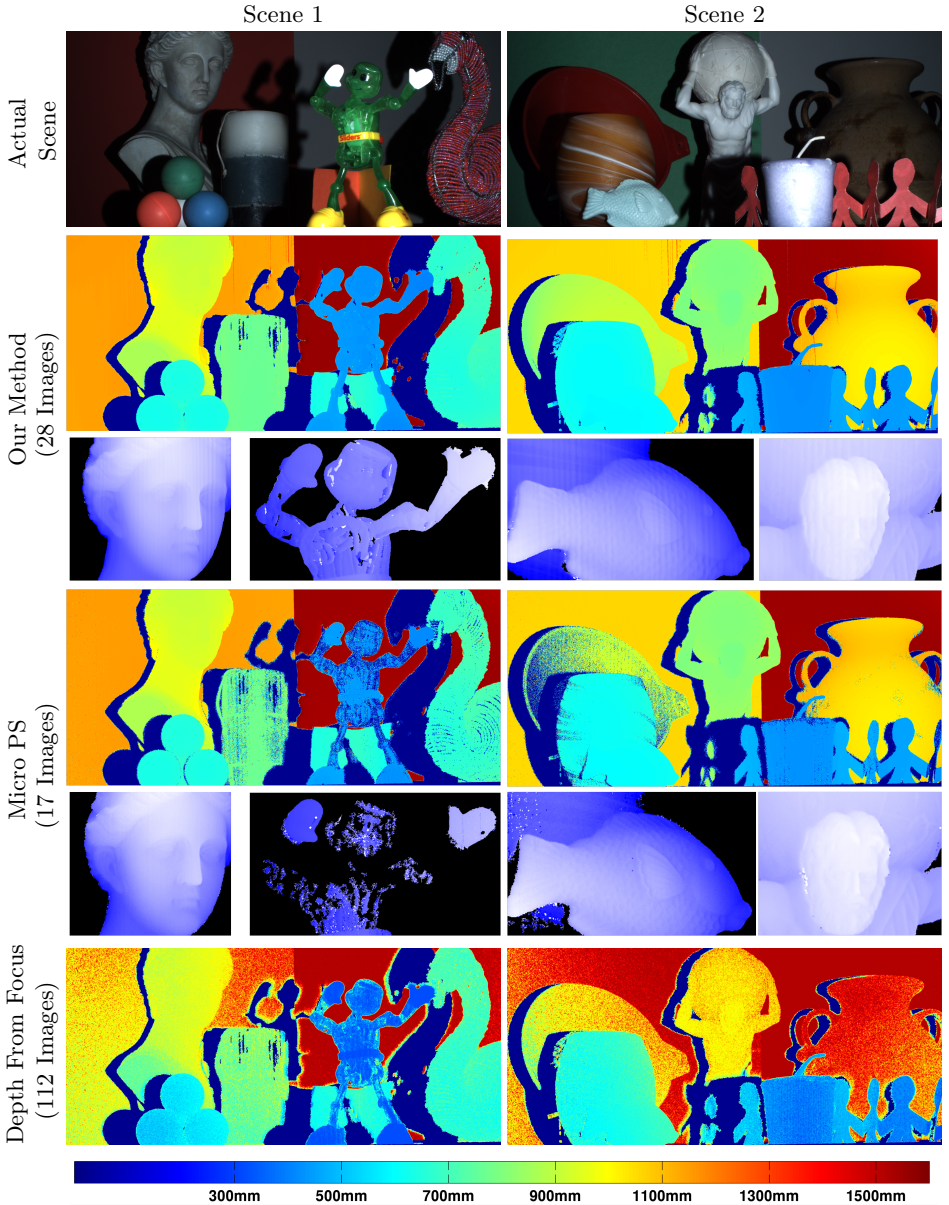


Fig. 7: Recovering Depth: Our structured light algorithm is able to recover depth maps for scenes containing challenging objects over an extended working volume with relatively few images. The insets for our method and micro phase shifting show (rescaled) depth maps for small parts of the scene. Many fine details on objects like the scales on the soap fish are successfully resolved.



Fig. 8: Direct-Global Separation. Groundtruth was computed using 112 images and our method used 28. In scene 1, the white hands on the robot toy appear much brighter in the global image computed by our method than the ground truth. This is because our algorithm tried to fit a linear trend to saturated (completely white) camera pixels. In scene 2, the shading on the soap fish and the white statue becomes is very clear in the direct illumination image

## References

1. Achar, S., Nuske, S.T., Narasimhan, S.G.: Compensating for Motion During Direct-Global Separation. International Conference on Computer Vision (2013)
2. Chen, T., Seidel, H.P., Lensch, H.P.: Modulated phase-shifting for 3D scanning. IEEE Conference on Computer Vision and Pattern Recognition pp. 1–8 (Jun 2008)
3. Couture, V., Martin, N., Roy, S.: Unstructured light scanning to overcome inter-reflections. International Conference on Computer Vision (2011)
4. Goldman, D.B., Curless, B., Hertzmann, A., Seitz, S.M.: Shape and spatially-varying BRDFs from photometric stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(6), 1060–71 (Jun 2010)
5. Gupta, M., Agrawal, A., Veeraraghavan, A., Narasimhan, S.G.: Structured light 3D scanning in the presence of global illumination. IEEE Conference on Computer Vision and Pattern Recognition pp. 713–720 (Jun 2011)
6. Gupta, M., Nayar, S.K.: Micro Phase Shifting. IEEE Conference on Computer Vision and Pattern Recognition (2012)
7. Gupta, M., Tian, Y., Narasimhan, S.G., Zhang, L.: A Combined Theory of Defocused Illumination and Global Light Transport. International Journal of Computer Vision 98(2), 146–167 (Oct 2011)
8. Hermans, C., Francken, Y., Cuypers, T., Bekaert, P.: Depth from sliding projections. IEEE Conference on Computer Vision and Pattern Recognition (2009)
9. Hern, C., Gabriel, V., Bjorn, J.B., Roberto, S.: Non-rigid Photometric Stereo with Colored Lights. International Conference on Computer Vision (October 2007)
10. Inokuchi, S., Sato K, Matsuda F.: Range imaging system for 3-d object recognition. In: International Conference on Pattern Recognition (1984)
11. Lei, S., Zhang, S.: Digital sinusoidal fringe pattern generation: Defocusing binary patterns VS focusing sinusoidal patterns. Optics and Lasers in Engineering 48(5), 561–569 (May 2010)
12. Nayar, S.K., Krishnan, G., Grossberg, M.D., Raskar, R.: Fast separation of direct and global components of a scene using high frequency illumination. ACM Transactions on Graphics 25(3), 935 (Jul 2006)
13. O’Toole, M., Raskar, R., Kutulakos, K.N.: Primal-dual coding to probe light transport. ACM Transactions on Graphics 31(4), 1–11 (Jul 2012)
14. Reddy, D., Ramamoorthi, R., Curless, B.: Frequency-Space Decomposition and Acquisition of Light Transport under Spatially Varying Illumination. European Conference on Computer Vision (2012)
15. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. IEEE Conference on Computer Vision and Pattern Recognition (2003)
16. Schechner, Y., Kiryati, N.: Depth from defocus vs. stereo: How different really are they? International Journal of Computer Vision 39(2), 141–162 (2000)
17. Tao, M., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from Combining Defocus and Correspondence Using Light-Field Cameras. International Conference on Computer Vision (2013)
18. Yuan, T., Subbarao, M.: Integration of Multiple-Baseline Color Stereo Vision with Focus and Defocus Analysis for 3-D. In: Proceedings of SPIE. pp. 44–51. No. i (Nov 1998)
19. Zhang, L., Nayar, S.: Projection defocus analysis for scene capture and image display. ACM Transactions on Graphics (2006)