# Implementing High Resolution Structured Light by Exploiting Projector Blur

Camillo J. Taylor
GRASP Laboratory
University of Pennsylvania
Philadelphia, PA USA
cjtaylor@cis.upenn.edu

## Abstract

*This paper presents a novel variant of structured lighting which exploits the inherent blur in the projector system to overcome the discrepancy in resolution between typical Digital SLR cameras and typical projector systems. More specifically, the scheme estimates the coordinates of the projection of each illuminated scene point in the projector frame with subpixel precision and this additional level of accuracy helps to improve the quality of the resulting 3D reconstructions.*

## 1. Introduction

Structured Light has long been a popular method for recovering the 3D structure of an object or scene. In its simplest form, structured light schemes make use of a projector system that illuminates the scene with a time varying pattern and a camera system which captures a series of images under this varying illumination. This image sequence is then decoded in order to establish correspondences between points in the image and points in the projector frame.

Structured Light owes its popularity to a number of felicitous characteristics: it is relatively simple to implement requiring only a projection system and a camera, one can implement self-calibrating versions of the scheme which simplifies deployment, and the method provides dense depth estimates even on untextured areas of the scene where stereo methods typically have difficulties.

One basic issue with structured light reconstruction that has become increasingly annoying over the past decade is the growing disparity between the resolution of the image sensors and the resolution of the projector systems.

The resolution of image sensors, driven by the market for digital cameras and Hendy's Law, a corollary to Moore's law, has been increasing exponentially. Typical consumer grade Digital Single Lens Reflex (SLR) cameras now offer tens of megapixels in a single image. Twenty megapixel cameras are not uncommon and higher resolutions are available.

Projector technology, which is governed by a different set of physical limitations and which is not being driven by the same mass market forces, has failed to keep pace. Standard Digital Light Projector (DLP) based projector systems range in resolution up to 1920x1080, the resolution required to support High Definition Television. The highest resolution DLP based projector system available at this point in time offers a resolution of approximately 4 Megapixels.

With standard structured light decoding schemes one is limited by the resolution of the projector. That is, while one can decode a corresponding projector pixel coordinate for every pixel in the image frame, the quantization of the projector ultimately limits the accuracy of the reconstruction.

This paper proposes a novel approach to overcoming this limitation. The insight is that we can exploit the blur induced by the optics of the projector to achieve subpixel resolution of the recovered projector coordinates. This allows us to more precisely localize scene points in the projector frame and, hence, improve the accuracy of the resulting 3D reconstruction.

### 1.1. Related Work

Projector blur has been exploited previously by Zhang and Nayar [10] who describe a scheme for measuring projector blur using a confocal projector camera system. They explain how one can, with a properly calibrated system, deduce the estimated projector defocus at every point in the scene from a series of images.

The method proposed in this paper differs substantially from this previous work since it does not involve estimating depth from the observed variation in projector defocus. The proposed scheme estimates depth via triangulation by considering the disparity between the location of points in the camera image and the projector image. The defocus blur here is being used to establish subpixel correspondences in the projector frame.

A number of interesting approaches have been developed which seek to combine range information from structured light with per pixel estimates of surface normals derived from photometric stereo. One of the earliest exemplars of this work is the paper by Nehab *et al*. [8] who describe an efficient technique for merging per-pixel normal and range data to improve the quality of the reconstruction. Subsequent work by Aliaga and Yu [1] described a system that made use of multiple cameras and projectors in a self calibrating network. Their system also provided a photogeometric reconstruction by fusing photometric stereo results with structured light. More recently Lu *et al*. [7] described an impressive system for fusing photometric stereo results obtained using an ultra-high resolution imaging system with low resolution structured light depth measurements to obtain range results that rival those obtained from a high accuracy laser scanning system.

All of the previous approaches rely on a photometric stereo step that makes fairly strong assumptions about the reflectance properties of the underlying surfaces. Typically, one assumes that the surface is Lambertian so that one can recover estimates for the surface normal at each pixel from a relatively small number of intensity measurements. Some of the photometric stereo methods require carefully calibrated lighting systems in order to produce accurate results. These methods also involve solving a large, but sparse, linear system in order to fuse the range and normal measurements into a final surface estimate.

Classical structured light techniques project a plane of light into the scene and then search for the center of the resulting scan line in the image [4]. These approaches often model the Gaussian blur in the illuminant intensity and seek to recover the peak of the beam in the image with subpixel accuracy. This approach is quite different from the one advocated in this paper since it seeks subpixel coordinates in the image plane rater than the projector frame. The number of lines in the scan is still limited by the resolution of the projector. Further the approaches to recovering the subpixel coordinates in the image typically assume that the underlying surfaces have constant albedo or smooth geometry so that variation in image intensity can be related to the beam position. These assumptions are frequently problematic in practice.

The proposed approach is more similar to the spatiotemporal analysis approach described by Curless and Levoy [2] who also note that scanning a plane of light over a surface results in a Gaussian profile in the spatiotemporal volume of images. In their work they consider a scanning system where a camera and laser plane are fixed relative to each other and translated relative to the scene. Their analysis of this geometry lead them to consider slanted lines in the spatiotemporal volume for better beam localization. The analysis in the present paper differs since it uses a fixed projector and camera system and the resulting analysis takes place *entirely in the temporal domain* which allows us to give an accurate depth estimate for each pixel in the frame through a relatively simple analysis of a 1D signal.

Another popular approach to recovering subpixel correspondence values is the method of sinusoids which involves projecting a series of shifted sinusoidal patterns on the scene and analyzing the resulting intensity variation at each pixel [9, 11]. Unfortunately, this method assumes that one can produce an accurate sinusoidal pattern with the projector which is surprisingly difficult. The pixelation of the projector also imposes fundamental limitations which are more difficult to overcome. A more detailed comparison is presented in Section 3.2.

The remainder of this paper is organized as follows: Section 2 describes the technical details of the implementation of the high resolution structured light system. Section 3 describes a quantitative and qualitative evaluation of the results obtained with the proposed scheme and the final section presents some of the conclusions drawn from this work.

## 2. Technical Approach

The proposed scheme builds upon the tried and tested projector based structured light paradigm depicted in Figure 1 . This figure shows a scene being illuminated by a projector and observed by a digital camera. In our experiments we utilized a BenQ PB2640 DLP based projector with a resolution of 1024x768 pixels and a Canon 50D 15 megapixel Digital SLR camera with a resolution of 4770x3177. The camera was outfitted with a zoom lens which allowed for focal lengths between 28 and 135 millimeters.

Figure 1. In a structured light system the scene is illuminated by a projector and the images are captured with a camera.

The first phase of data collection uses a standard binary stripe sequence to recover the integral row and column indices of every pixel in the scene. A standard gray code stripe pattern is employed to cut down on decoding errors. Since we are interested in recovering both the row and column coordinates of every point, two stripe sequences are used, 10 images of horizontal stripes and 10 images of vertical stripes.

In the second phase of data collection a different set of horizontal and vertical stripe patterns is employed. In this set each stripe is one pixel wide and the stripes are spaced 8 pixels apart. An example of this stripe pattern is shown in Figure 3. Shifted versions of the stripe pattern are projected onto the scene so that the stripes appear to march across the surface one pixel at a time as the patterns advance. In all 8 sets of horizontal stripes and 8 sets of vertical stripes are projected. The same stripe pattern was employed by Guhring [4] in his work on 3D surface acquisition although he used the resulting image measurements in a completely different manner.

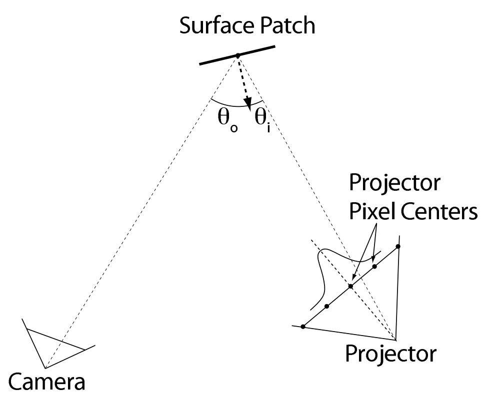



Figure 2. The effective irradiance that a pixel in the projector contributes to a point in the scene is related to the displacement between the projection of that scene point on the projector frame and the center of the pixel. This falloff is modeled as a Gaussian.

If we consider the simplified 2D depiction shown in Figure 2 we can relate the observed scene radiance measured at a pixel in the camera, $I$, to the irradiance supplied to the corresponding scene point by the projector, $E$, as follows:

$$I = f(\theta_o, \theta_i) E \cos\theta_i \tag{1}$$

where $f(\theta_o, \theta_i)$ represents the BRDF at the scene point. The net effect of the projector optics is modeled with a Gaussian blur kernel so the irradiance supplied by the projector to the scene point is given by the following expression:

$$E(k) = E_0 \exp\left(\frac{-(k-\delta)^2}{\sigma}\right) \tag{2}$$

where $k$ denotes the stripe index, $\sigma$ models the width of the blur kernel at that point in the scene and $\delta$ is the projection of the scene point in the projector frame which is the parameter that we ultimately recover with subpixel precision. In this model the irradiance that a point in the scene receives from a pixel in the projector is related to the disparity between the center of that projector pixel and the projection of the scene point on the projector grid.

Combining Equations 1 and 2 yields the following expression which models how the observed intensity of the pixel, $I(k)$, varies as the stripe is marched across the scene.

$$I(k) = I_1 \exp\left(\frac{-(k-\delta)^2}{\sigma}\right) + I_0 \tag{3}$$

In this equation $I_1 = f(\theta_o, \theta_i) E_0 \cos\theta_i$ while $I_0$ models the scene irradiance due to ambient illumination, this quantity can be estimated by considering the lowest intensity value recorded over the 8 frame sequence. Once this has been subtracted off, one can easily recover the remaining parameters by taking logs of the residual intensity values and fitting a quadratic function to the resulting values in the vicinity of the maximum observed intensity. Equation 4 shows the relationship between the parameters of interest and the coefficients of the local quadratic fit.

$$\log(I(k) - I_0) = \left(\log I_1 - \left(\frac{\delta^2}{\sigma}\right)\right) + \left(\frac{2\delta}{\sigma}\right) k - \left(\frac{1}{\sigma}\right) k^2 \tag{4}$$

This procedure recovers a floating point offset between -0.5 and 7.5 at each scene point which effectively corresponds to the lower bits of the projection of the scene point in the projector frame. This result is then spliced onto the integer code recovered in the previous phase to obtain the final x and y projector coordinates. Note that in this scheme the effective BRDF of the scene point and the orientation of the surface patch *do not* affect the sub pixel localization scheme since they only enter equation 3 as scale factors. Notice also that this method recovers the size of the blur kernel, $\sigma$, at each pixel independently, this is important since the blur will in fact change based on the depth of field.

The model assumes that the intensity readings from the camera vary linearly with scene irradiance so all of the imagery is captured in raw format and all of the processing is done on the raw measurements prior to the indignities of the debayering process. At the end of the decoding process for every illuminated pixel in the camera image one recovers two floating point values, $(x_s, y_s)$, which represent estimates for the projection of the scene point onto the projectors frame with subpixel precision.

At this point one can exploit the projector camera duality to auto-calibrate the system and recover the intrinsic parameters of both the camera and the projector [5, 3]. This stage proceeds exactly as one would proceed when provided with a set of correspondences between two images, the only difference being that you literally have millions of correspondences from which to choose. The fact that the system is self calibrating significantly simplifies the deployment process. The proposed scheme chooses a small subset of the correspondences at random for computational convenience and then recovers an estimate for the fundamental matrix relating the camera and the projector from these measurements in the usual manner. This matrix is then decomposed to yield estimates for the focal lengths of the camera and the projector and the rotation and translation relating the two frames. This initial estimate is then refined using the publicly available Sparse Bundle Adjustment package [6]. This bundle adjustment procedure also models the radial distortion of both the camera and the projector. In our experiments, it was convenient to include both quadratic and quartic radial distortion terms.

Once the intrinsic and extrinsic parameters are recovered we perform non-linear resectioning to compute the depth of every point in the scene that has been decoded. This resectioning procedure is simplified by choosing the camera as the base frame of reference. For each of the pixels in the image that have been decoded we perform a non-linear optimization to recover the depth that is in best agreement with the decoded projector coordinates. Given the Z depth at a pixel and the intrinsic parameters of the camera one can recover the associated X and Y coordinates in the scene. Note that the procedure recovers the depth at each pixel *independently*. This means that the there are no significant issues in the vicinity of depth discontinuities. Further the independent depth estimates provide a firm foundation for subsequent filtering operations.

## 3. Experimental Results

In order to characterize the blur kernel associated with our projector system an experiment was carried out wherein a flat matte surface was set up perpendicular to the optical axis of the projector. Under these imaging conditions, the intensity variation observed on the target should mirror the intensity variation produced by the projector without foreshortening artifacts.

The planar target was approximately 50 centimeters on side and was constructed by mounting a matte surface on a half inch thick sheet of plexiglass to help ensure planarity. This surface was then illuminated with the aforementioned stripe pattern. Figure 3a shows one of the images where the scene is being illuminated with one of the single pixel thick stripe patterns. Figure 3b shows a closeup view of a small section of the image showing the details of that stripe pattern.

Figure 4a plots how the intensity varies across the columns of one such region and finally Figure 4b shows how the log of the intensity varies in and around one of the local maxima. The purple curve shows the quadratic function that was fit to this data. This quadratic fit seems to be in good agreement with the observed variation which helps to justify the simple Gaussian blur model.

a. b.

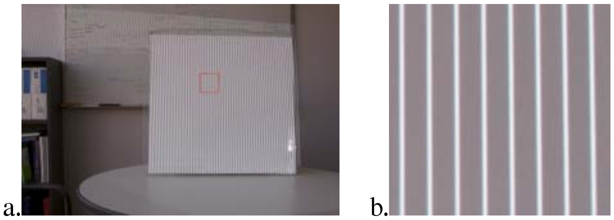
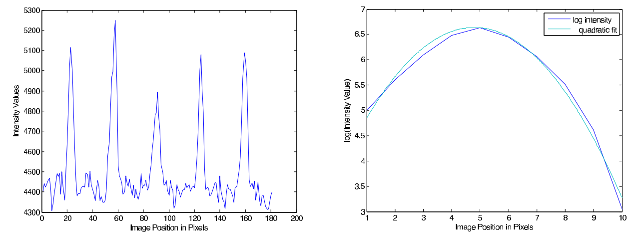


Figure 3. a. Planar target illuminated with a single pixel stripe pattern b. Closeup of image on a small region of the target.




a b

Figure 4. a. Profile of intensity variation along a row of the image shown in Figure 3. b. Closeup of a local maxima of the profile showing a quadratic fit to the log of the intensity values.

| Distance to target | baseline | subpix. | mean error | median error | max error |
|---|---|---|---|---|---|
| 2664.7 | 1110.2 | N | 1.397 | 1.270 | 6.123 |
| 2664.7 | 1110.2 | Y | 0.488 | 0.409 | 3.140 |
| 1559.9 | 1027.9 | N | 0.606 | 0.584 | 1.963 |
| 1559.9 | 1027.9 | Y | 0.153 | 0.132 | 0.732 |

Table 1. Table summarizing the residual error associated with the planar target. All values are in millimeters.

### 3.1. Comparison to Standard Structured Light

Experiments were also carried out to help quantify the accuracy of the proposed scheme and to compare the results to those that would be obtained without subpixel interpolation. The entire reconstruction procedure was carried out to recover the position of every illuminated point in the scene with respect to the camera. A least squares procedure was invoked to fit a planar surface to most of the pixels on the planar target and the residual to this fit was recorded and analyzed.

The reconstruction experiment was repeated using the projector coordinates recovered using only the first decoding stage which produces integral projector coordinates. The planar target was then moved to a different position with respect to the camera and the entire procedure was repeated.

The results of the experiment are summarized in Table 1 which shows how the mean, median and maximum residual values over the area of regard differed between the subpixel and non-subpixel variants and how the errors changed as the depth to the target was varied. To further characterize the variation in error, the plots in Figure 5 show histograms of the residual error for all four test cases.

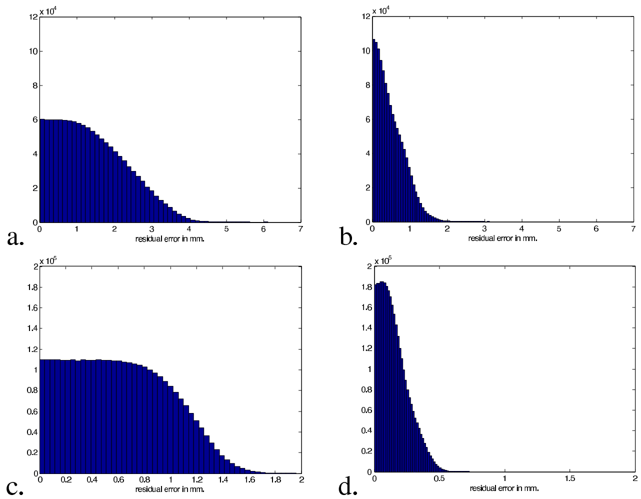



Figure 5. These histograms compare the distribution of residual errors to the planar fit before and after subpixel refinement. The first row shows the error distribution when the target was placed 2.7 meters away and the second row shows the error when the target was 1.5 meters away. The first column corresponds to the errors recorded without subpixel refinement while the second shows the errors observed with subpixel refinement. Note that the subpixel residual errors are smaller.

These results demonstrate that the subpixel variant offers a clear advantage over the standard structured light procedure and reduces the residual error by approximately a factor of 3. In order to investigate how the proposed procedure would impact the perceived quality of the 3D reconstruction results experiments were carried out using the scene shown in Figure 6.

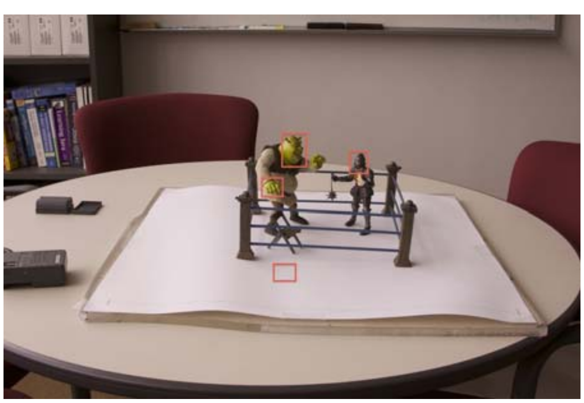


Figure 6. Sample scene used in reconstruction experiments. The highlighted areas are expanded in Figure 7.

Figure 7 shows side by side comparisons of the depth reconstruction results obtained using subpixel and non-subpixel correspondences in the regions marked by boxes in Figure 6. Note that these 3D reconstruction results have not been smoothed or filtered in any way so one can observe some outliers.

Figure 7h shows the 3D results returned by the standard non-subpixel structured light method on a planar region in the foreground of the scene. Here one can clearly observe the classic sawtooth error pattern induced by the quantization error on the pixel coordinates. This error is largely eliminated in the subpixel version.

### 3.2. Comparison to Method of Sinusoids

Another technique that is commonly used to produce high accuracy correspondences in structured light systems is the method of sinusoids [9, 11]. The approach involves projecting a sequence of shifted sinusoids onto the scene and then analyzing the resulting intensity variation of the individual image pixels. The theoretical model predicts that the intensity of a pixel in the frame should vary according to Equation 5:

$$I(k) = I_1 \sin(\alpha k + \phi) + I_0 \tag{5}$$

Where $\alpha$ denotes the phase difference between successive intensity patterns. In our experiments we have chosen to project a sequence of 8 patterns separated in phase by $\frac{\pi}{4}$ radians.

Given a sequence of intensity values taken over time from a single pixel one can readily recover the unknown parameters, $I_1$, $I_0$ and $\phi$ from a least squares procedure. The phase associated with each pixel, $\phi$, encodes its position in the projector frame. In principle, this value can be recovered with subpixel precision. In practice, this sinusoidal model makes a number of assumptions which are difficult to satisfy. Firstly it, assumes that one can precisely control the illumination of each pixel in the projector so as to produce a reasonable approximation of a sinusoid. In fact, for most projector systems this is a very questionable proposition. Various non-linearities in the intensity response are introduced by gamma corrections in the display card and by other intended and unintended intensity distortions in the projector.

This problem is illustrated in Figure 8 which plots the series of 8 intensity values collected by a single camera pixel as the sinusoidal illumination patterns are applied. The continuous curve shows the best fit sinusoidal model in a least squares sense for those samples. The divergence between the samples and the curve is quite alarming.

These modeling problems are reflected in the subpixel results returned by the method. Figure 9 shows a direct comparison of the subpixel coordinates returned by the method of sinusoids and those returned by the proposed

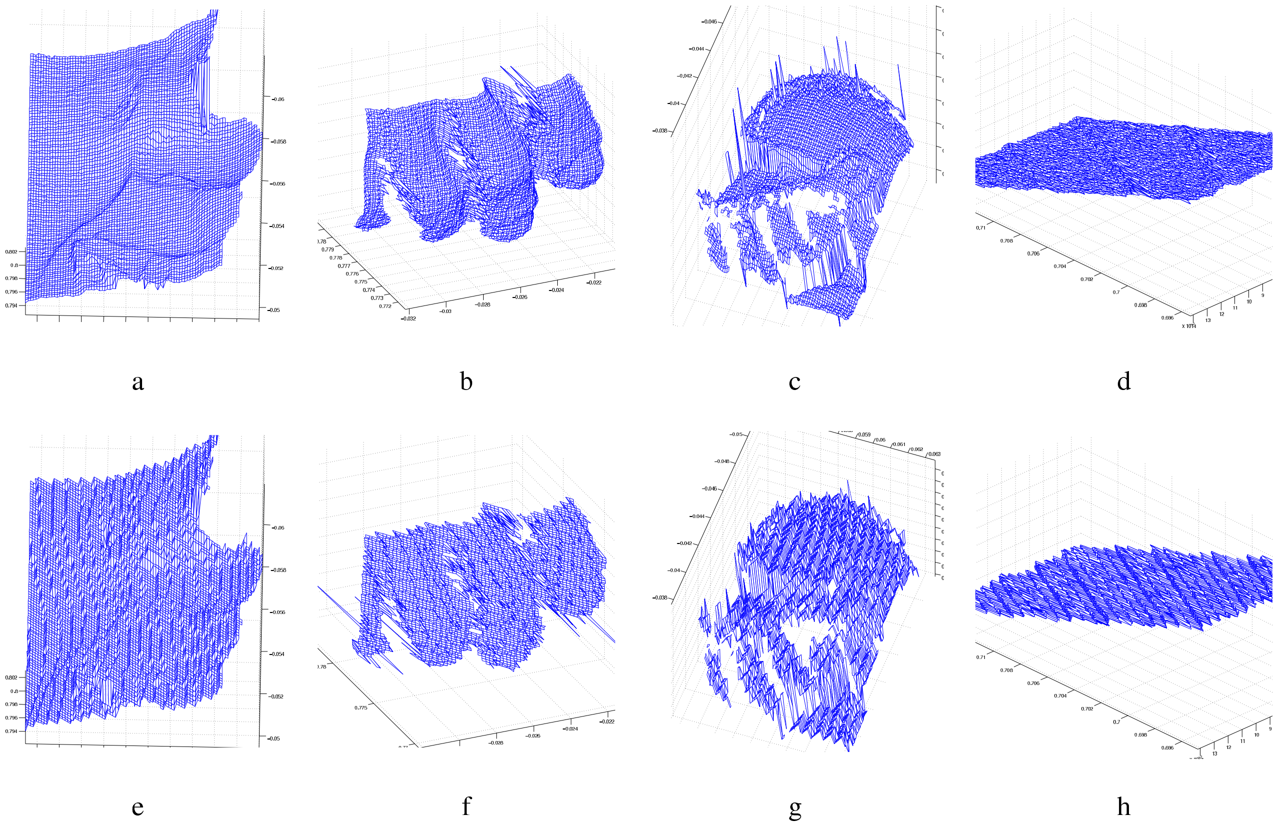



Figure 7. This figure compares the reconstruction results obtained on the scene in Figure 6. The first row corresponds to the results obtained with subpixel refinement and the second row shows the results without subpixel refinement.

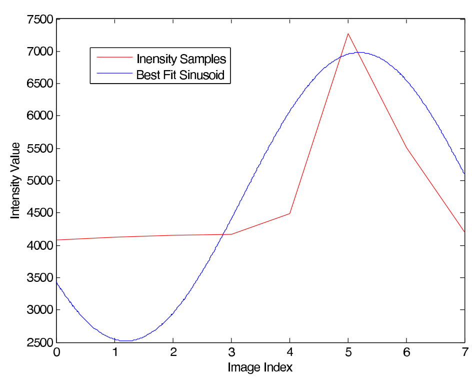



Figure 8. This figure shows how the intensity of a single pixel in the image changes over time as it is illuminated by a sequence of shifted sinusoids. The continuous curve is the best fit sinusoid for the intensity samples.

projector blur method over the span of a horizontal line segment in the camera image. Since this was an image of our flat target we would expect the subpixel coordinates to vary linearly with pixel position. The bottom plot shows the results from the method of sinusoids while the top plot shows the results from the proposed method. The two plots have been deliberately offset for ease of comparison. In this plot one can clearly see that the proposed method produces results that are much more linear. Quantitatively, for this set of 200 pixels the root mean square error associated with the sinusoidal error was 0.6029 pixels while the RMS error of the projector blur method was 0.0470 pixels.

In order to obtain accurate results with the method of sinusoids on standard resolution imagery one is typically obliged to perform a fairly complex photometric calibration step to recover the curve relating programmed image intensity to actual scene irradiance as described in [11]. As such we believe that the proposed approach offers an advantage since it involves fewer assumptions and is simpler to implement on typical projector systems.

In the context of high resolution imagery, a more fundamental limitation is imposed by the pixelation of the projector. With a projector system the illumination pattern that is produced is not a true sinusoid but a discrete approximation thereof. If the projector is properly focused on the scene, this discretization will be apparent in the high reso-

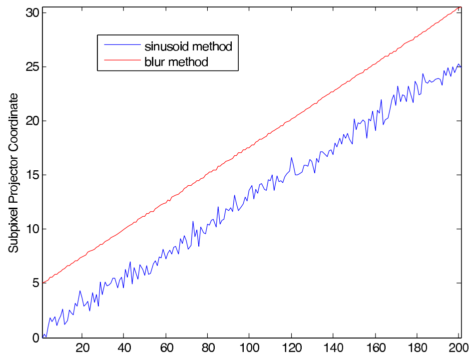



Figure 9. This figure compares the subpixel values returned by the method of sinusoids (the bottom plot) to the subpixel results returned by the proposed projector blur method (the top plot) over a span of pixels in the image. Both plots should be straight lines.

lution imagery and the method of sinusoids would offer no subpixel advantage since all of the points in the scene that are illuminated by a given pixel in the projector would vary in exactly the same way and would, therefore, decode to the same phase value.

With low resolution imagery the discretization of the projector is matched by the discretization of the camera so the camera pixels effectively low pass filter the intensity variation and help to justify the sinusoidal model. With a high resolution camera the discrete nature of the projector is much harder to hide since each pixel in the projector now covers several pixels in the camera image.

### 3.3. Scene Reconstruction

One compelling application of the proposed reconstruction technique is model acquisition for graphics. The fact that the procedure can be used to capture high resolution, high accuracy range scans that are naturally registered with the color information from the camera allows us to effectively render the scene from novel vantage points as shown in Figure 10. These renderings were produced using a point splatting procedure. This approach is particularly well suited to the data sets produced by the proposed reconstruction procedure which consist of individual scene points with both position and color information. Prior to rendering, the depth images were smoothed with a simple 3x3 median filter to identify and remove obvious outliers.

We observe that the resolution of the Digital SLR is typically an order of magnitude higher than the resolution of the novel views one wants to generate for fly-throughs or other purposes which means that several pixels in the scene typically contribute to every pixel in the rerendered view. The resolution and accuracy of the scans allows the user to effectively zoom in on areas of interest in the scene.

## 4. Conclusions

In this work we choose to make a virtue out of a necessity. Instead of lamenting the inevitability of projector blur and the uncertainties that it can raise in the decoding process, we instead exploit the blur to improve the accuracy of our procedure.

The method has been evaluated both quantitatively and qualitatively and has been shown to provide a substantial improvement over traditional binary structured light methods and over the commonly used method of sinusoids.

The goal of this effort has been to build upon the strengths of structured light methods namely, simplicity, flexibility and accuracy. The resulting method is, in practice, identical to standard structured light the only difference being that it employs an additional round of stripe patterns which are analyzed to produce the subpixel estimates. The scheme is entirely self calibrating which means that the camera and projector can be repositioned and refocused as needed without issue.

The calculations involved in implementing the scheme are relatively straightforward. The subpixel coordinates at each pixel are recovered independently as are the depth estimates which eliminates the need for large scale relaxation computations that fuse information from neighboring pixels and simplifies the handling of scene discontinuities.

Modeling the net effect of the optics as a Gaussian blur is a simple but effective choice which appears to work reasonably well in this context. It may be interesting to explore the advantages of more complex blur models in future work.

Finally, while the scheme has been described as a two frame method involving a single camera and projector it is a relatively straightforward matter to extend the method to deal with multiple cameras and projectors. For example one may choose to use a single camera with multiple placements of the projector so as to ensure that every pixel in the camera image is illuminated by one or more of the projector scans. The self calibration framework can easily be extended to handle such a situation [5, 3]. Similarly one could use the proposed method to scan a scene from multiple different vantage points and fuse the results into a single representation for rendering or modeling purposes.

## References


[1] D. Aliaga and Y. Xu. A self-calibrating method for photogeometric acquisition of 3d objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(4):747 –754, 2010. 10

[2] B. Curless and M. Levoy. Better optical triangulation through spacetime analysis. *IEEE International Conference on Computer Vision*, 0:987, 1995. 10

[3] R. Furukawa and H. Kawasaki. Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning. In *3-D Digital Imaging and Mod-*

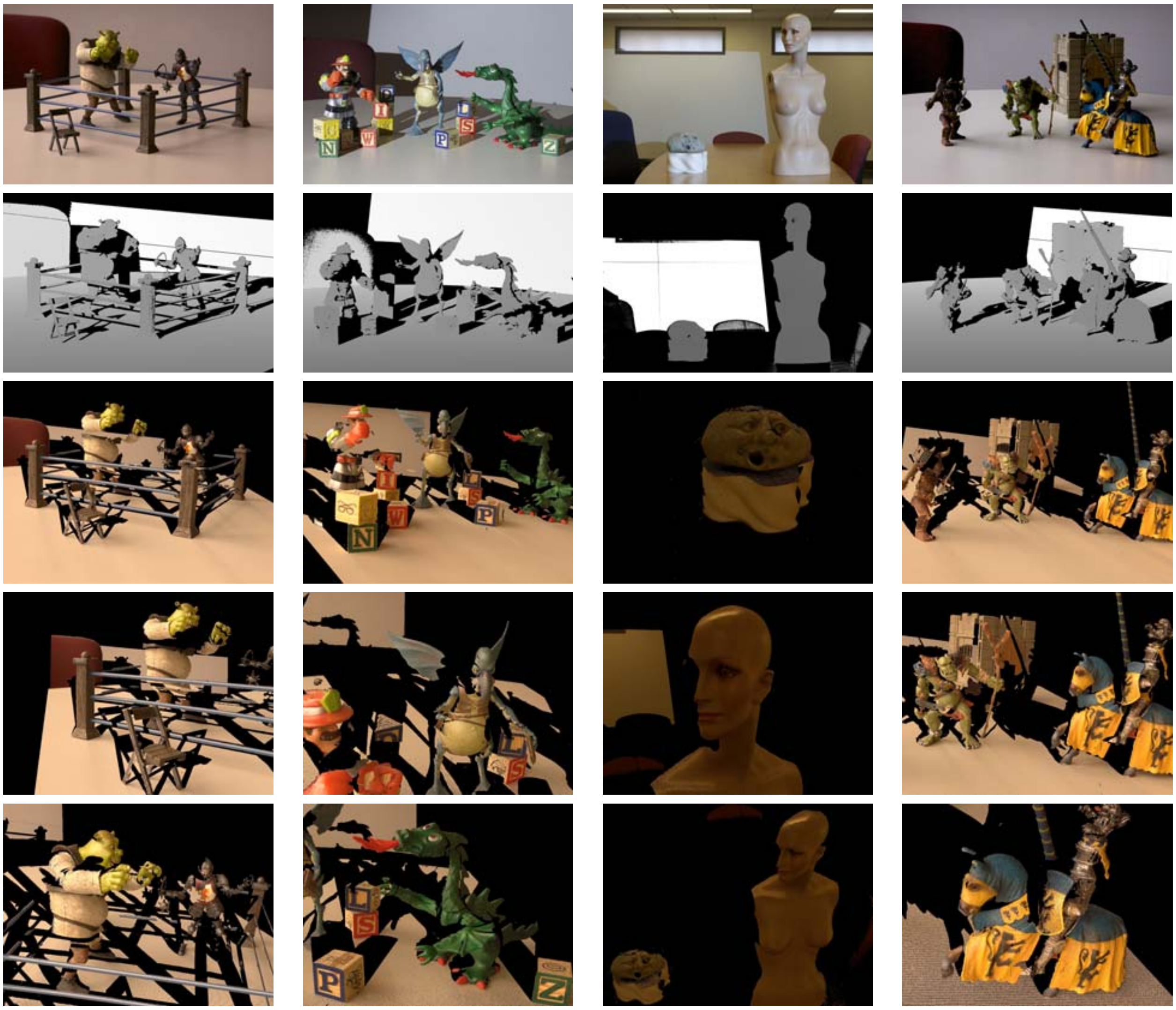


Figure 10. This figure shows results obtained by applying the proposed reconstruction scheme to a few different scenes. The first row corresponds to the original input image. The second row depicts the recovered depth map and the remaining 3 rows show novel views of the scene produced using a splatting renderer.


*eling, 2005. 3DIM 2005. Fifth International Conference on*, pages 302 – 309, 2005. 11, 15

[4] J. Guhring. Dense 3-d surface acquisition by structured light using off-the-shelf components. In *Proc. Videometrics and Optical Methods for 3D Shape Measurement*, pages 220–231, 2001. 10, 11

[5] R. E. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. 11, 15

[6] M. A. Lourakis and A. Argyros. Sba: A software package for generic sparse bundle adjustment. *ACM Trans. Math. Software*, 36(1):1–30, 2009. 12

[7] Z. Lu, Y.-W. Tai, M. Ben-Ezra, and M. Brown. A framework for ultra high resolution 3d imaging. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1205 –1212, 2010. 10

[8] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph.*, 24:536–543, July 2005. 10

[9] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using strucured light. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 195–202, 2003. 10, 13

[10] L. Zhang and S. K. Nayar. Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Jul 2006. 9

[11] S. Zhang and S.-T. Yau. Generic nonsinusoidal phase error correction for three-dimensional shape measurement using a digital video projector. *Appl. Opt.*, 46(1):36–43, 2007. 10, 13, 14