

Embedded Phase Shifting: Robust Phase Shifting with Embedded Signals

Daniel Moreno Kilho Son Gabriel Taubin
Brown University, Providence, RI, USA
{daniel.moreno, kilho.son, gabriel.taubin}@brown.edu

Abstract

We introduce *Embedded PS*, a new robust and accurate phase shifting algorithm for 3D scanning. The method projects only high frequency sinusoidal patterns in order to reduce errors due to global illumination effects, such as subsurface scattering and interreflections. The frequency set for the projected patterns is specially designed so that our algorithm can extract a set of embedded low frequency sinusoids with simple math. All the signals, patterns high and embedded low frequencies, are used with temporal phase unwrapping to compute absolute phase values in closed-form, without quantization or approximation via LUT, resulting in fast computation. The absolute phases provide correspondences from projector to camera pixels which enable to recover 3D points using optical triangulation. The algorithm estimates multiple absolute phase values per pixel which are combined to reduce measurement noise while preserving fine details. We prove that embedded periodic signals can be recovered from any periodic signal, not just sinusoidal signals, which may result in further improvements for other 3D imaging methods. Several experiments are presented showing that our algorithm produces more robust and accurate 3D scanning results than state-of-the-art methods for challenging surface materials, with an equal or smaller number of projected patterns and at lower computational cost.

1. Introduction

The need for better 3D scanners that simplify the creation of high quality 3D models from everyday physical objects, is driven nowadays in a good measure by recent improvements in 3D printing technologies, and the increasing availability of low cost 3D printers. Structured light 3D scanners, comprising a digital camera and a digital projector, are well suited for this task because they can produce dense 3D point cloud models with as many points as camera pixels, the 3D scanning device has no moving parts, and the

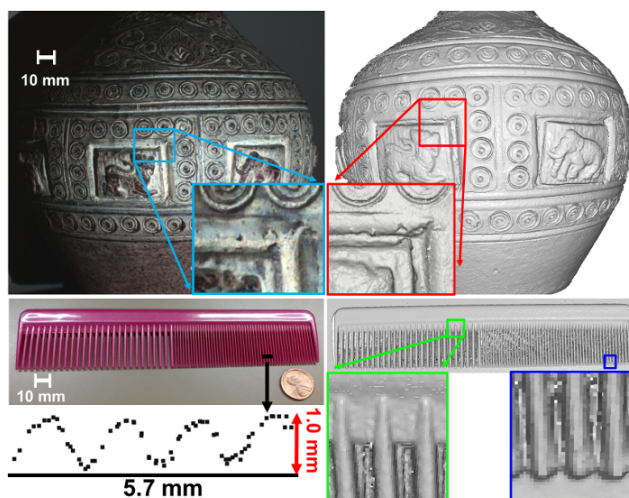


Figure 1. Embedded PS. Models generated from 9 images with no post processing such as filtering or mesh reconstruction. Fine structures below 1mm depth in a 180mm object are correctly measured. Total decoding time with a single thread Matlab implementation takes about 1.2s, image resolution is 1600x1200.

scanning process usually requires a relatively small number of projected patterns and corresponding captured images (a minimum of five in our method). As a comparison, laser based scanners, comprising a digital camera and a laser line projector, also need a mechanism to adjust the location and orientation of the projected line with respect to the object after each image is captured. And since a single line is recovered from each image, they need to capture hundreds of images to build a dense 3D model. In this paper we present the *Embedded Phase Shifting (Embedded PS)* method for structured light 3D scanning. Embedded PS generates very accurate high resolution dense 3D point cloud models projecting only high spatial frequency sinusoidal patterns. Different from alternative prior-art methods, Embedded PS embeds a low frequency signal in the *phase* of the high frequency patterns. The method is robust to global illumination effects because its pattern frequencies are constrained to a small range. Its computational complexity is low because decoding the embedded low frequencies reduces to a simple

subtraction operation, and the relative phase unwrapping is trivial once the embedded frequencies have been recovered. Finally, it provides redundant absolute phase measurements which are used for denoising by a voting scheme or by averaging, without feature-destroying spacial smoothing.

In traditional two-camera stereo systems, pixel correspondences are established by time consuming and error prone pixel matching processes. Structured light 3D scanning systems, comprising a digital projector and a single digital camera, use the data projector to illuminate the scene with specially designed patterns, allowing for example to identify which data projector pixel column is illuminating the scene point being imaged by each camera pixel with no search, independently of the scene content. Mathematically, both the projector and the camera are modeled as pinhole cameras. The position of 3D points is estimated using the principle of triangulation, where each projector column corresponds to a plane in 3D, each camera pixel corresponds to a ray in 3D, and the intersection of the plane and the ray results in the 3D point coordinates. The straightforward extension to multiple cameras is not discussed here.

The details of how the correspondences between projector columns and image pixels are determined depend on the coding strategy used in the projected patterns. For example, the Binary Coding method represents each projector column index number as a sequence of bits, which are encoded as a sequence of black and white patterns. For each projected pattern an image is captured. The code is reconstructed by determining whether each camera pixel is black or white in each image. The resulting sequence of bits is the binary expansion of the corresponding column number. Although this formulation is very simple, the decision of whether a pixel is black or white is often non-trivial, because pixel intensities are a sum of three components: ambient illumination (considered fixed), direct illumination from the projector, and indirect illumination or global illumination effects. The latter is usually unknown and difficult to measure. In a seminal work, Nayar et al. [8] demonstrated how to separate direct and global illumination using high frequency structured light patterns, a principle that was used in [15] to robustly classify pixels either as black or white in binary patterns as the one described above. Since the projection of binary patterns permits to resolve only projector integer column numbers, the resolution of the estimated 3D point cloud model is limited by the projector resolution.

Many other coding strategies have been described in the literature, and a comprehensive list is provided by Salvi et al. [12]. In the context of this work, we are concerned only with methods based on phase shifting, where projector column coordinates are encoded as the phase of a continuous sinusoidal pattern. The continuous nature of these coding strategies results in 3D point cloud models generated at the resolution of the camera, which tends to be higher than the

resolution of the digital projectors these days.

Our phase shifting method is optimal according to the seven principles enumerated by Pribanić et al. [10]. The most important of these principles are: that each pixel correspondence is solved independently of the others, enabling high resolution reconstructions and robustness to sharp object discontinuities; and that simple image processing is performed allowing efficient implementations. Within the existing structured light coding strategies (e.g. spatial neighborhood, direct codification, Fourier transform profilometry) in general phase shifting is known to produce superior results for static scenes [10].

The main contributions of the *Embedded Phase Shifting* method are:

- It is a robust phase shifting method which works with as low as 5 projected patterns. It exploits low frequency signals embedded in high frequency patterns to find correspondences between projector and camera pixels. 3D scanning results are quantitatively and visually better than state-of-the-art methods using the same or smaller number of projected patterns.
- We present a mathematical framework which explains how to encode the low frequency periodic signals in the high frequency projected patterns, and how to decode them from the corresponding captured images. These theoretical results are shown to be applicable not only to sinusoids, but to a wide variety of families of periodic signals. We believe that these results could be used to exploit embedded frequencies in other phase shifting methods (e.g. Trapezoidal PS [4]).
- It produces multiple high-precision 3D measurements from a small number of projected patterns, enabling robust 3D imaging with no post processing such as smooth filtering, which removes fine details, resulting in the ability to reconstruct much finer object structures.
- It requires only simple image processing operations, resulting in a fast decoding algorithm. The algorithm complexity increases linearly with the number of camera pixels and it is independent of the number of projector pixels.

2. Related Work

Following the classification from [12] ours is a Multiple Phase Shifting method within the *continuous* coding techniques, same as [1, 3, 5, 9], meaning that sinusoidal signals of multiple frequencies are used for shape measurement and that the coding domain is continuous. Single Phase Shifting methods [13] assume that the surface being scanned

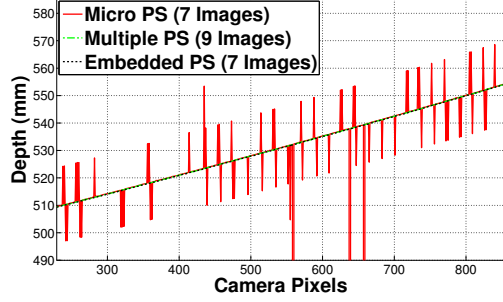


Figure 2. Metal plane profile measured with Micro PS (red), Multiple PS (green), and Embedded PS (black). We have used Micro PS with the 5 optimized frequencies from Gupta et al. [3] and yet it produces frequent unwrapping errors, opposite to the other two methods which use temporal phase unwrapping.

is smooth and without depth discontinuities (e.g. a human face) and use spatial phase unwrapping (fringe counting). Multiple Phase Shifting uses temporal phase unwrapping [11, 14] allowing to resolve each scene point independently of its neighbors being robust to surface discontinuities. Gupta et al. [2] observed that, unless the scene is perfectly Lambertian, structured light methods are sensitive to global illumination effects and showed that such effects are minimized when the spatial frequencies belong to a narrow range, in which case, explicit separation of direct and global illumination is unnecessary [8, 15]. They demonstrated this result by designing a frequency constrained Gray code. The same principled was applied by Micro PS [3] to create a phase shifting method with multiple sinusoids all within a very narrow frequency range. Micro PS is robust to global illumination effects and defocus of the illumination source but, because of the complete lack of low frequency signals, unwrapping of relative phase values is performed using a nearest neighbor search on a LUT for each pixel, which is both slow and error prone. Some errors are reduced to some extent by using a median filter on the phase map. Our method is inspired on Micro PS but with the help of the embedded frequencies we perform *temporal phase unwrapping* [11, 14] which is fast and robust to unwrapping errors without any extra filtering. On a different approach, Modulated PS [1] adds a high frequency carrier so they can explicitly separate the illumination components. Karpinsky et al. [5] extract one low frequency signal from two high frequency sinusoids which they use to reduce the total number of images projected, however their method does not generalize to more than two signals. Liu et al. [6] use amplitude modulation to add a low frequency signal to high frequency patterns reducing the effective dynamic range of each signal. As result their patterns have both high and low spatial frequencies, whereas ours embed a low frequency in the phase resulting *only* in high spatial frequencies.

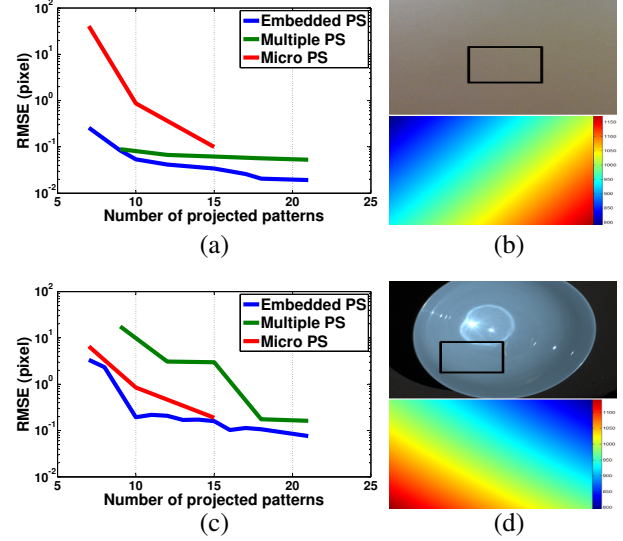


Figure 3. A planar surface (a) and a convex bowl (b) scanned with Multiple PS, Micro PS, and Embedded PS. Plot shows RMSE against ground truth as number of projected patterns increase. The plane is Lambertian while the bowl is a specular surface which causes strong interreflections. (b) and (d) display object images and ground truth phase maps.

3. Phase Shifting Methods

3.1. Single frequency phase shifting

Single frequency phase shifting (Single PS) [13] encodes projector column (or row) numbers as the phase of a sinusoidal which is shifted in phase a few times. More formally, let I_n be the n phase shifting pattern from a total of N , $n \in [0, N - 1]$, then the intensity of a projector pixel (x_p, y_p) is given by

$$I_n(x_p, y_p) = o + a \cos(\omega x_p + \theta_n) \quad (1)$$

where o and a are an offset and amplitude respectively, $\omega = 2\pi f$ depends on the frequency f , and $\theta_n = 2\pi n/N$ is the n -shift. The intensity at a camera pixel observing such patterns is given by

$$R_n(x, y) = o_\omega + a_\omega \cos(\omega x_p + \theta_n) \quad (2)$$

where variables o_ω , a_ω , and x_p are now unknown. By projecting $N > 2$ shifts a system of equations is built and the phase value is solved as

$$\hat{x}_p = \frac{1}{\omega} \tan^{-1} \frac{\sin(\omega x_p)}{\cos(\omega x_p)} \quad (3)$$

with $x_p = \hat{x}_p + kT$ due to the periodic nature of the sine function. The value \hat{x}_p is called *relative phase* and the goal is to recover the *absolute phase* x_p by finding the unknown integer k . The sine period is defined as $T = 1/f$. Unless T is equal or larger than the total number of columns x_{max} , finding the correct k is a non-trivial problem known as *phase unwrapping*.

3.2. Multiple frequency phase shifting

Phase unwrapping is impossible to solve in Single PS for general scenes where surface discontinuities are allowed, except in the trivial case where $k = 0$, because the projected patterns do not have enough information. To account for this problem, multiple frequency phase shifting (Multiple PS) repeats Single PS M times with frequencies f_1, f_2, \dots, f_M with $M > 1$, $f_m > f_{m+1}$, and $f_M \geq 1/x_{max}$. Given these conditions, absolute phases are computed using *temporal phase unwrapping* [11]. The minimum number of projected images for Multiple PS is $3M$ and most of the works in phase shifting, including ours, are a variant of this method.

3.3. Temporal phase unwrapping

Let ϕ_1, \dots, ϕ_M be relative phases corresponding to frequencies f_1, \dots, f_M , meeting the conditions stated in Section 3.2. Let u_m be their corresponding unwrapped phases, $u_m = \phi_m + k_m T_m$, and let x be the true unknown absolute phase. Note that if all values were exact then $x = \phi_m$ for all m . In practice, values ϕ_m are corrupted by noise, quantization, and measurement errors, and the relation is not exact. Temporal phase unwrapping is used to compute an absolute phase estimate \hat{x} by incrementally unwrapping phases ϕ_m from the lowest to the highest frequency as follows: 1) $k_M = 0$; 2) $k_m = \lfloor f_m(u_{m+1} - \phi_m) \rfloor$ for $m = M - 1, \dots, 0$, where $\lfloor \cdot \rfloor$ is a rounding function of its argument; and 3) $\hat{x} = \phi_1$. Temporal phase unwrapping is optimal [11] when periods of the sequence of frequencies increase exponentially (e.g. 16, 256, 4096, ...) rather than linearly.

3.4. Micro Phase Shifting

Multiple PS, similar to Binary and Gray codes, generates patterns with a wide spectrum of frequencies, from low to high, which makes them not robust to global illumination effects. To improve over this situation, Micro PS [3] constrains the patterns to a very narrow spectrum. Additionally, Multiple PS computes different o_ω and a_ω values (Equation 2) for each frequency, which is correct because they vary with ω , however, when frequencies are very close to each other the measured o_ω 's and a_ω 's across frequencies are very similar too. Based on this observation, Micro PS made the simplification of considering a single offset o and amplitude a for all the projected frequencies reducing the number of unknown variables. Therefore, reducing the minimum number of patterns in relation to the number of frequencies. Micro PS projects 3 shifts of one frequency, as Single PS, and one extra pattern for each additional frequency. The total number of projected patterns is $M + 2$ for M frequencies. Solving all the unknown in a single lin-

ear system they recover the following phase vector

$$\mathbf{u} = [\cos \phi_1, \sin \phi_1, \cos \phi_2, \sin \phi_2, \dots, \cos \phi_M, \sin \phi_M]^T \quad (4)$$

where only the relative phase ϕ_1 can be solved using Equation 3. Temporal phase unwrapping cannot be applied because the all-high frequencies do not follow the stated condition and because the values of ϕ_2 to ϕ_M cannot be extracted from \mathbf{u} . To overcome this limitation, a look-up-table (LUT) is built with the value of vector \mathbf{u} for integers from 0 to x_{max} (all projector columns), and unwrapping is done with a nearest neighbor search for each decoded pixel. LUT-unwrapping is slow and not very robust because measurement vector \mathbf{u} is corrupted with noise and because phase values ϕ_m are continuous, whereas the LUT has only a subset of them. As result, median filtering of the unwrapped phase map is almost mandatory for Micro PS which smooths out some fine structures and adds extra processing time, without necessarily removing all the errors. Figure 2 displays how frequent this unwrapping errors occur in Micro PS compared to Multiple PS and Embedded PS both of which use temporal phase unwrapping. In summary, Micro PS is robust to global illumination effects by design but produces frequent unwrapping errors.

4. Embedded Phase Shifting

Embedded PS is a new multiple frequency phase shifting method. Its key properties are that it is robust to global illumination effects, similar to Micro PS, and that uses temporal phase unwrapping, similar to Multiple PS, on the contrary, neither Micro PS nor Multiple PS have both properties. Our algorithm takes advantage of the fact that any combination of two or more relative phases could be used to compute a phase of a new signal of different frequency. In particular we use this property to generate a sequence of low frequency signals, from a sequence of high frequencies, which we use together with temporal phase unwrapping to recover the absolute phase for all frequencies. We now explain our method: Section 4.1 refers to pattern creation, Section 4.2 to the recovery of relative phases from images, and Section 4.3 to absolute phases and depths measurements.

4.1. Coding strategy

Let be $\{T_1, \dots, T_M\}$ a set of real numbers greater than 1, we define *embedded frequency* F_m as

$$F_m = \frac{1}{T_1} \cdots \frac{1}{T_m}. \quad (5)$$

They meet the relation $F_m > F_{m+1}$. We define a set of pattern frequencies $\{f_1, \dots, f_M\}$ as

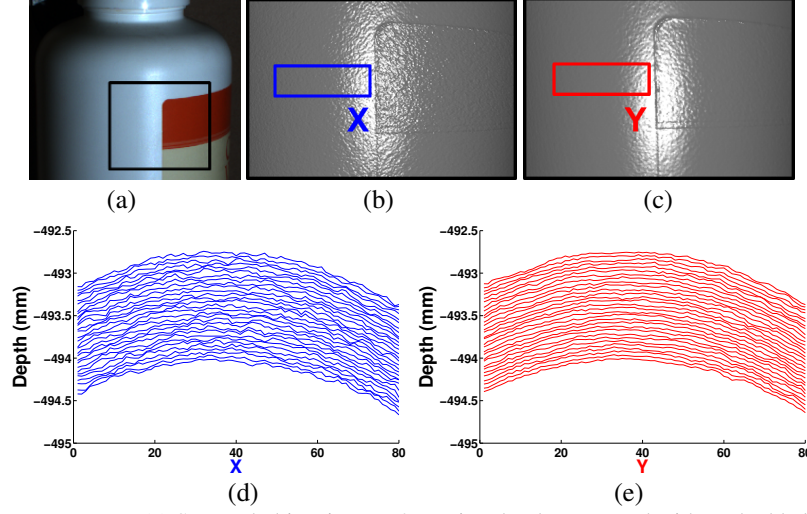


Figure 4. Multiple depth measurements. (a) Scanned object image. 3D point clouds generated with Embedded PS using (b) a single depth measurement and (c) 10 depth measurements. Averaging multiple depths removed randomized noise while preserving small discontinuities as the step at the paper label. (d) and (e) are magnified plots of the surface profile within the selected boxes.

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_M \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_M \end{bmatrix}. \quad (6)$$

They all meet the relation $f_1 \leq f_m \leq f_M$, being $f_M - f_1$ the pattern frequency range.

Parameters T_m define both the set of embedded (Equation 5) and pattern frequencies (Equation 6) and the user is free to choose their values with the requirement of $F_M \geq 1/x_{max}$, which ensures that the set of embedded frequencies meet the conditions from Section 3.2 so that temporal phase unwrapping can be used. By changing values T_m we can adjust the pattern frequency range as desired.

Let be $\{N_1, \dots, N_M\}$ a set of integers greater than 1. Now, use Equation 1 to create a set of phase shifting patterns with N_m shifts for each pattern frequency f_m . The total number of patterns created N is given by

$$N = \sum_{m=1}^M N_m \quad (7)$$

Integers N_m define the number of phase shifts $\theta_{m,n}$ projected for each frequency f_m . They must be chosen such as $N \geq 2M + 1$ to ensure that a linear system used for decoding (Equation 10) is not underdetermined. It is common to set $N_1 = 3$ and N_m equal 2 or 3 for $m > 1$.

It is assumed that $M > 1$ resulting in 5 as the minimum number of patterns for this method. Note that a total of $N_f = 2M - 1$ pattern and embedded frequencies are encoded with a minimum of $N = 2M + 1$ patterns,

or equivalently, $N_f + 2$ patterns are projected for N_f frequencies, same pattern-frequency count relation than that of Micro PS.

Coding Example: if $x_{max} = 1024$ we could set $T_1 = 16$ and $T_2 = T_3 = 8$. The embedded frequencies (Equation 5) are $F_1 = 1/16$, $F_2 = 1/128$, and $F_3 = 1/1024$. The pattern frequencies (Equation 6) are $f_1 = 1/16$, $f_2 = 1/14.22$, and $f_3 = 1/15.75$, being all high frequencies. We could set $N_1 = 3$ and $N_2 = N_3 = 2$; defining f_1 shifts as $\theta_{1,1} = 0$, $\theta_{1,2} = 2\pi/3$, and $\theta_{1,3} = 4\pi/3$; and f_2 and f_3 shifts as $\theta_{2,1} = \theta_{3,1} = 0$ and $\theta_{2,2} = \theta_{3,2} = 2\pi/3$. The total number of patterns (Equation 7) is $N = 7$. Note that F_m 's follow an exponential relation which is optimal for temporal phase unwrapping.

4.2. Pattern decoding

Each camera pixel location is decoded independently. Let \mathbf{r} be a radiance vector containing the values in the image sequence at the position being decoded, let be o an offset and a_m an amplitude value, and let be \mathbf{u} a phase vector defined as

$$\mathbf{u} = [o, c_1, s_1, \dots, c_M, s_M]^T, \quad (8)$$

$$c_m = a_m \cos(\omega_m \phi_m), \quad s_m = a_m \sin(\omega_m \phi_m). \quad (9)$$

We recover the phase vector by solving for solving for \mathbf{u} in Equation 10 and we use each pair (c_m, s_m) in Equation 3 to obtain the a relative phase ϕ_m .

$$\mathbf{u} = \arg \min_{\mathbf{u}} \|\mathbf{r} - \mathbf{A}\mathbf{u}\|, \quad (10)$$

$$\mathbf{A} = \begin{bmatrix} 1 & A_1 & & \\ \vdots & & \ddots & \\ 1 & & & A_M \end{bmatrix}, \quad (11)$$

$$A_m = \begin{bmatrix} \cos(\theta_{m,1}) & -\sin(\theta_{m,1}) \\ \vdots & \vdots \\ \cos(\theta_{m,N_m}) & -\sin(\theta_{m,N_m}) \end{bmatrix}, \quad (12)$$

A is the shift-matrix and is fixed for all pixels.

Finally, we compute a relative phase Φ_m for each embedded frequency F_m by simple subtraction

$$\begin{cases} \Phi_1 \equiv \phi_1 \\ \Phi_m = \phi_m - \phi_1 \text{ for } m > 1. \end{cases} \quad (13)$$

Relative phases ϕ_1 to ϕ_F must be unwrapped to recover depth information at the corresponding camera pixel. Relative phases Φ_1 to Φ_F provide the information required to perform the unwrapping.

4.3. Unwrapping with embedded frequencies

We apply temporal phase unwrapping to unwrap phases Φ_2 to Φ_F (we skip $\Phi_1 = \phi_1$). Since by design the embedded frequencies meet the conditions from Section 3.2. Unwrapped Φ_m 's are not directly used for depth estimation because they are sensitive to noise. Instead, unwrapped Φ_2 , corresponding to the highest embedded frequency, is accurate enough to unwrap *all* phases ϕ_m with little or no error. All unwrapped ϕ_m 's were extracted from high frequency patterns and provide reliable depth measurements. As an alternative, all pattern and embedded frequencies could be sorted from highest to lowest and unwrapped together in a single step, but we prefer to unwrap each ϕ_m independently using only Φ_2 to generate independent depth estimations.

4.4. Multiple depths estimations

Multiple PS and Micro PS use multiple frequencies to help the unwrapping procedure, but final depth measurements are computed from a single frequency (the highest one in Multiple PS and the first one in Micro PS), whereas, our method recovers multiple absolute phase values ϕ_m , all with comparable resolution given that frequencies f_m are all similar. We exploit this additional information to estimate a more accurate final depth measurement by averaging all ϕ_m 's which reduces white noise. In some cases other approaches might be preferred, for instance, outliers could be detected using a voting scheme between all phases.

4.5. Theoretical Foundations

We present simple theorems explaining how embedded frequencies are generated from pattern frequencies.

Theorem 1. *Let $a = a_1 + k_1T_1$ and $a = a_2 + k_2T_2$ where $0 \leq a_1 < T_1$, $0 \leq a_2 < T_2$; and k_1, k_2 are integers. Then $a = a_3 + k_3T_3$ with*

$$T_3 = \frac{T_1T_2}{T_2 + T_1}, \quad a_3 = T_3 \bmod \left(\frac{a_1}{T_1} + \frac{a_2}{T_2}, 1 \right), \quad (14)$$

where k_3 is an integer.

Theorem 1 applies to phase shifting as this: let a be an unknown absolute phase; let be a_1 and a_2 relative phase values of periodic signals, with period T_1 and T_2 respectively, recovered from the pattern images. Then, we can calculate another relative phase a_3 of a periodic signal with period T_3 without this signal being explicitly present in the original patterns. We calculate additional phases with the following generalization.

Theorem 2. *Let $a = a_1 + k_1T_1$ and $a = a_2 + k_2T_2$ where $0 \leq a_1 < T_1$, $0 \leq a_2 < T_2$; and k_1, k_2 are integers. If $\alpha T_2 + \beta T_1 > 0$ and α, β are integers, then $a = a_3 + k_3T_3$ with*

$$T_3 = \frac{T_1T_2}{\alpha T_2 + \beta T_1}, \quad a_3 = T_3 \bmod \left(\frac{\alpha a_1}{T_1} + \frac{\beta a_2}{T_2}, 1 \right), \quad (15)$$

where k_3 is an integer.

Theorem 2 permits to use two relative phase values to recover the phase of additional periodic signals by varying parameters α and β . Theorem 2 is further generalized by applying the same concepts to multiple periodic signals.

Theorem 3. *Let $a = a_m + k_mT_m$ with $m \in \{1, \dots, M\}$, where $\forall m : 0 \leq a_m < T_m$. Let be k_1, k_2, \dots, k_M integers. Then $a = a_3 + k_3T_3$ with*

$$T_3 = \left(\sum_{m=1}^M \frac{\alpha_m}{T_m} \right)^{-1}, \quad a_3 = T_3 \bmod \left(\sum_{m=1}^M \frac{\alpha_m a_m}{T_m}, 1 \right), \quad (16)$$

if $\frac{\alpha_m a_m}{T_m}$ are all positive integers, where k_3 is also integer.

Theorem 3 states that phases of additional periodic signals are found with any combination of the recovered relative phases. We call the new periodic signals *embedded signals* and their frequencies *embedded frequencies*. We note that Theorems 1, 2 and 3 are not restricted to sinusoids, they are valid for any periodic signal class. Therefore, they might be used to create an *embedded* version of other structured-light coding schemes.

5. Experiments

Our equipment is a DLP LightCrafter 4500 projector and a Point Grey GRAS-20S4C camera. The projector native resolution is 912x1140 and the camera is configured to capture 1600x1200 8bits RGB images saved as PNG files. We used the projector-camera calibration software [7] for geometric calibration. We implemented Embedded PS and Multiple PS in Matlab and we have used the original authors Matlab implementation of Micro PS, either with the optimized 5-frequency set [14.57, 16.09, 16.24, 16.47, 16.60] from Gupta et al. [3] or the frequencies Gupta et al. have used in the sample data available at their website.

We performed various experiments to verify the performance of Embedded PS, as seen in Figures 1 and 6. We applied our algorithm on challenging objects which introduce strong global illumination effects. A convex bowl with specular surface is difficult to scan with phase shifting algorithms due to the presence of severe interreflections. Embedded PS scanned it using 15 images and produced a visually correct model, whereas, models from Micro PS and Multiple PS display severe artifacts. Subsurface scattering is also a major challenge for structured light 3D scanning. Our algorithm produced an artifact-free model of an orange presenting subsurface scattering with only 6 images.

Figure 3 quantitatively compares the algorithms. We selected two simple objects: a plane and a convex bowl. These objects were scanned by the three methods with incrementally increasing number of patterns. Root Mean Squared Error (RMSE) was measured for each scan. Ground truth (GT) was obtained in an independent procedure where the objects were scanned many times while gradually increasing the number of patterns up to 123, point at which we saw no further improvement on the results which we considered as the most reliable measurement that could be made by the method. Micro PS was not used for GT computation because it is unclear how to select a large number of frequencies all being coprime and varying a fraction of a pixel within a very narrow band optimally. Figure 3 (a) shows a plot corresponding to the RMSE on the plane for each algorithm. Since almost no global illumination effects are present in this scene, Embedded PS and Multiple PS exhibit low error. Micro PS performs a little worse due to unwrapping errors. In contrast, a similar plot in Figure 3 (c) corresponding to the bowl, where a lot of interreflections were present due to its geometry and specular finish, shows how Multiple PS results are degraded, while Embedded PS and Micro PS maintain their performance. Our algorithm outperforms the other two in both cases.

Embedded PS, unlike Micro PS and Multiple PS, outputs multiple depth measurements for each pixel. These multiple depths allow to reduce randomized noise without applying spatial filtering which smooths out fine details. Figure 4 (b) shows a point cloud generated using a single depth estimate and Figure 4 (c) one generated from 10 depth estimates for each point. Both were created from the same 21 captured images. We observed that the point cloud from multiple depths is less noisy while the small step at the paper label is clearly preserved.

Figure 5 compares the computation time for Micro PS and Embedded PS versus the number of images processed. In both cases, time is directly proportional to the number of images, however, Embedded PS spent 1.1s to decode and unwrap 7 images and it slowly increases to 4.5s for 21 images. On the contrary, Micro PS spent 44.7 seconds only on the 7 image set.

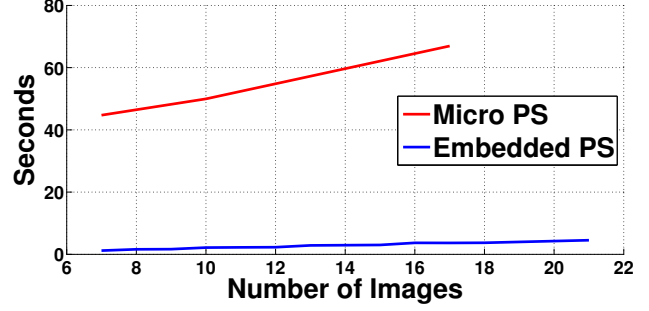


Figure 5. Computation time versus number of images

6. Discussion

We proposed a robust, accurate, and high precision phase shifting 3D scanning method. Low frequency signals embedded in high frequency patterns allow us (1) to only use high spatial frequencies and avoid severe global illumination interference, (2) to unwrap relative phases with simple closed-form equations without approximation nor quantization, (3) to measure multiple 3D depths per pixel which permits to denoise 3D measurements without spatial smoothing, and (4) to compute 3D measurements fast. We proved that embedded signals are not limited to sinusoidals, they could be extracted from other classes of periodic signals opening new possibilities for structured-light research.

Appendix

Proof of Theorem 1. We divide both equations with T_1 and T_2 respectively

$$\frac{a}{T_1} = \frac{a_1}{T_1} + k_1, \frac{a}{T_2} = \frac{a_2}{T_2} + k_2.$$

We find a by first adding both equations and next multiplying them by $T_3 = \frac{T_1 T_2}{T_2 + T_1}$

$$a = T_3 \left(\frac{a_1}{T_1} + \frac{a_2}{T_2} \right) + T_3 (k_1 + k_2),$$

$$a = T_3 \operatorname{mod} \left(\frac{a_1}{T_1} + \frac{a_2}{T_2}, 1 \right) + k' T_3 + (k_1 + k_2) T_3$$

Given that k_1 , k_2 and k' are all integers, $k_3 = k_1 - k_2 + k'$ is an integer too. Therefore,

$$a = T_3 \operatorname{mod} \left(\frac{a_1}{T_1} + \frac{a_2}{T_2}, 1 \right) + k_3 T_3,$$

Theorems 2 and 3 are proved similarly.

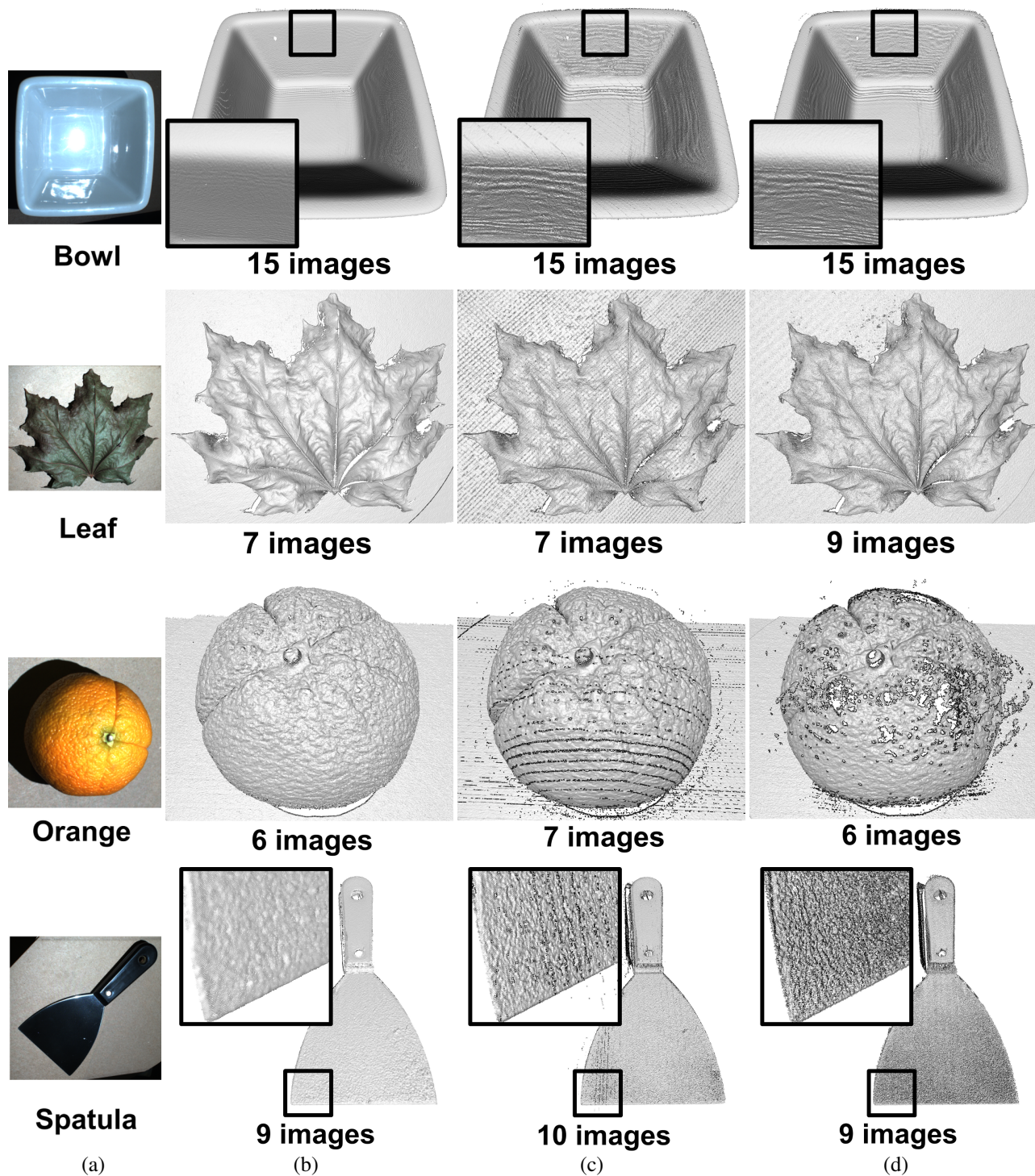


Figure 6. (a) Object images. (b) Proposed Embedded PS. (c) Micro PS. (d) Multiple PS. A bowl is a convex and Non-Lambertian surfaced object which introduces interreflection between surfaces. Embedded PS robustly scans the bowl with very small noise while results from the state-of-the-art carry severe noise. Fine structures on the leaf were clearly scanned by our algorithm using 7 images. Oranges are difficult to scan due to subsurface scattering, our algorithm produced an excellent result using just 6 images. The proposed Embedded PS scanned all objects robustly and precisely with a relatively small number of images even in presence of strong global illumination effects.

7. Acknowledgments

The authors want to thank David Michael for his support and Cognex Corp. for facilitating the equipment used in this work.

References

- [1] T. Chen, H.-P. Seidel, and H. Lensch. Modulated phase-shifting for 3D scanning. In *CVPR 2008*, pages 1–8, June 2008.
- [2] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. Narasimhan. Structured light 3D scanning in the presence of global illumination. In *CVPR 2011*, pages 713–720, June 2011.
- [3] M. Gupta and S. Nayar. Micro Phase Shifting. In *CVPR*, pages 1–8, Jun 2012.
- [4] P. S. Huang, S. Zhang, and F.-P. Chiang. Trapezoidal phase-shifting method for 3D shape measurement. volume 5606, pages 142–152, 2004.
- [5] N. Karpinsky, M. Hoke, V. Chen, and S. Zhang. High-resolution, real-time three-dimensional shape measurement on graphics processing unit. *Optical Engineering*, 53(2):024105–024105, 2014.
- [6] K. Liu, Y. Wang, D. L. Lau, Q. Hao, and L. G. Hassebrook. Dual-frequency pattern scheme for high-speed 3D shape measurement. *Optics express*, pages 5229–5244, 2010.
- [7] D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *3DIMPVT 2012*, pages 464–471. IEEE, 2012.
- [8] S. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Jul 2006.
- [9] T. Pribanić, H. Džapo, and J. Salvi. Efficient and low-cost 3D structured light system based on a modified number-theoretic approach. *EURASIP J Adv Sign Proc*, pages 1–11, 2010.
- [10] T. Pribanić, S. Mrvoš, and J. Salvi. Efficient multiple phase shift patterns for dense 3D acquisition in structured light scanning. *EURASIP IVC*, 28(8):1255–1266, 2010.
- [11] H. O. Saldner and J. M. Huntley. Temporal phase unwrapping: application to surface profiling of discontinuous objects. *Applied optics*, 36(13):2770–2775, 1997.
- [12] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666 – 2680, 2010.
- [13] V. Srinivasan, H. C. Liu, and M. Halioua. Automated phase-measuring profilometry: a phase mapping approach. *Appl. Opt.*, 24(2):185–188, Jan 1985.
- [14] J.-d. Tian, X. Peng, and X.-b. Zhao. A pitch-variation moiré fringes method of temporal phase unwrapping profilometry. *Optoelectronics Letters*, 3(3):215–218, 2007.
- [15] Y. Xu and D. G. Aliaga. Robust pixel classification for 3D modeling with structured light. In *Proceedings of Graphics Interface 2007*, GI ’07, pages 233–240, 2007.