Rapid Hand Shape Reconstruction with Chebyshev Phase Shifting

Daniel Moreno Wook Yeon Hwang Gabriel Taubin
Brown University, Providence, RI, USA
{daniel_moreno,wookyeon_hwang,taubin}@brown.edu

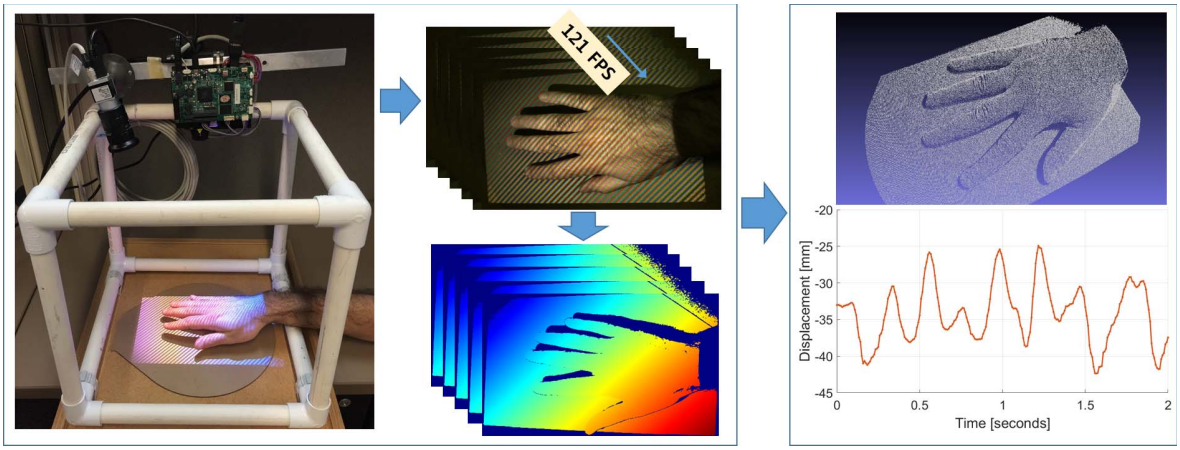

Figure 1: (Left) Sensor prototype for the measurement of a rapidly trembling hand shape and color: 14" cube structure delimiting the measurement volume, digital projector and camera above. (Middle Above) A sequence of color images is captured at 121 FPS while projecting CPS patterns (continuously repeating sequence of three RGB patterns). (Middle Below) Sliding window of three images is used to continuously generate range and color images at 121 FPS. (Right Above) Example of a one of the range images rendered as a very detailed point cloud. (Right Below) Vertical hand displacement plot at a control point is generated with a resolution of 8.3ms.

Abstract

Human hand motion and shape sensing is an area of high interest in medical communities and for human interaction researchers. Measurement of small hand movements could help professionals to quantize the stage of conditions like Parkinson's Disease (PD) and Essential Tremor (ET). Similar data is also useful for designers of human interaction algorithms to infer information about hand pose and gesture recognition. In this paper we present a structured light sensor capable of measuring hand shape and color at 121 FPS. Our algorithm uses a novel structured light method developed by us, called Chebyshev Phase Shifting (CPS). This method uses a digital projector and a camera to create high-resolution color 3D models from sequences of color images. We show how to encode CPS patterns in three RGB images for a reduced acquisition time, enabling high speed capture. We have built a prototype to measure rapid trembling hands. Our results show our prototype accurately captures fast tremors similar to those of PD patients. Color 3D model sequences recorded at high speed with our sensor will be used to study hand kinematic properties in a future.

1. Introduction

Accurate shape measurement of static and moving human bodies has been in the past years, and still is, an area of great interest in research communities with applications in diverse fields such us medicine, virtual reality, and gesture recognition. In this work, we are interested on accurate measurement of hand shape, color, and motion patterns. This kind of data is requested in the study of hand kinematics, and particularly by those trying to record and quantize hand motions due to conditions like Parkinson's Disease (PD) and Essential Tremors (ET). The motor phenotype of PD is characterized by spontaneous movement, rigidity, and tremors occurring at a frequency between 4 and 6 Hz [7]. The capture and analysis of such motions may contribute to a reliable and objective clinical assessment, and progression rate evaluation of these diseases [17]. Real time capture of data with these characteristics could also serve as input for algorithms performing tasks like hand pose determination, gesture recognition, and hand motion modeling. To enable collection of such data we have built a structured light sensor prototype capable of capturing accurate hand shape and color at 121 FPS. We made this possible with the application of a novel structured light method which we called Chebyshev Phase Shifting (CPS).

978-1-5090-5407-7/16 $31.00 © 2016 IEEE
DOI 10.1109/3DV.2016.24

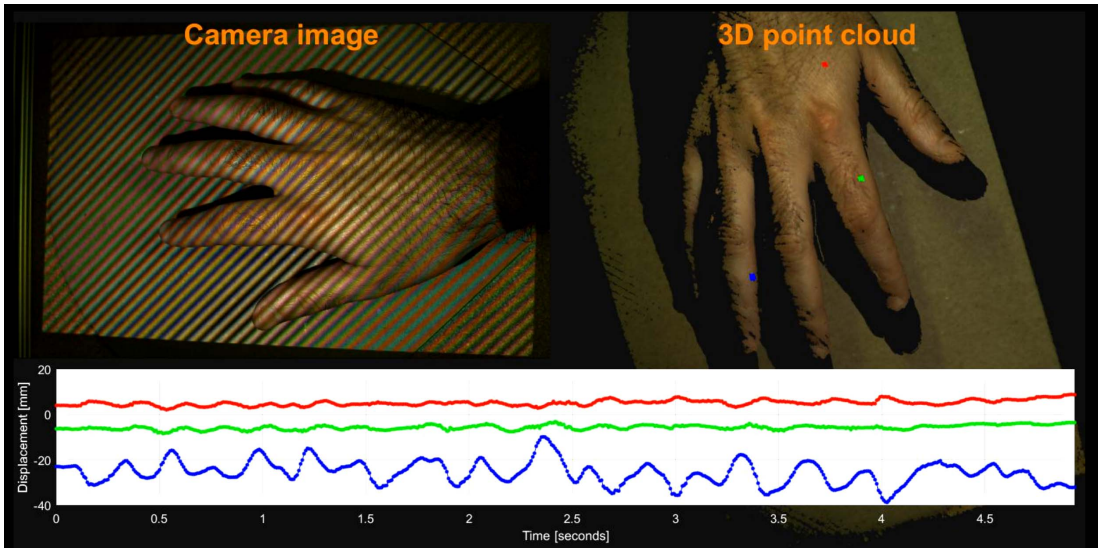

Figure 2: Screenshot of System for Rapid Hand Motion Measurement. (Left) Current camera image. (Right) Current point cloud. Control points marked in red, green, and blue. (Bottom) Plot of vertical displacement of the control points for 5 seconds with a time resolution of 8.3ms (600 data points). Prototype succeeded in capturing the blue marker tremors similar to those in PD (between 4 and 6 Hz).

Structured Light (SL) methods are preferred for accurate shape measurement. These are active methods that illuminate a scene of interest in the visible or invisible light spectrum with specially designed patterns. Images are captured of the scene under this illumination, with one or more cameras, from which the scene shape is inferred. The active light source makes results from SL consistent across different scene properties, such as color, texture, and materials. On the contrary, passive methods rely only on scene content and may fail to solve ambiguities in areas with no relevant features. Digital imaging technology can reliably project and capture high resolution images at high speed and with great precision. SL methods make use of such standard technology to generate very accurate 3D measurements. A competing technology is that of Time-of-Flight (ToF) sensors. They calculate depth by measuring the time of a signal to travel back and forth between the sensor and scene. ToF is promising but, at the time of this writing, existing sensors are either low-resolution or low-speed. As example consider a TI OPT8241 3D ToF sensor with a resolution of 320×240 at 60 FPS, and a Basler ToF Camera with a resolution of 640×480 at 15 FPS. Both with much lower resolution and speed than those of regular digital projectors and cameras. Our prototype has joint requirements of capture speed and detail, which moved us to develop a new sensor.

Most properties of any SL method is determined by the characteristics of the sequence of patterns being projected. The main objective of the pattern sequence is to encode information that permits to relate camera pixel locations to coordinates in the projection image plane by analyzing scene images captured by the camera. It is common for patterns to encode projector image column or row coordinates because, combined with geometric calibration, they allow to identify 3D scene point locations by calculating the intersection of a ray through a camera pixel and a projector plane of light, method known as optical triangulation. SL patterns are classified as discrete or continuous [14], depending on whether they permit to identify discrete or continuous positions at the projector coordinate space. Continuous patterns could potentially identify a different scene point for each camera pixel location, whereas, discrete patterns only identify discrete projector positions. In other words, 3D models resulting from the projection of continuous patterns may have as many points as camera pixels, but they may have as many as projector pixels if discrete patterns are used. In practice, continuous patterns impose stricter projection requirements, as the ability to accurately project different colors and intensity levels, for which photometric calibration becomes important to correctly match projected and observed light intensities. In comparison, discrete patterns usually are B&W requiring no photometric calibration. These binary patterns are also more robust to image noise and variations in scene color. Discrete examples are Binary Patterns [11] and Binary Reflected Gray Code (BRGC) [6]. Many continuous SL methods, including ours, are based on Phase Shifting (PS) [16] which uses the phase of sinusoidal patterns to continuously encode a projector direction (e.g. x-axis). PS pattern sequences are shorter than sequences of BRGC patterns, resulting in a faster acquisition time. In this work, we have chosen PS in favor of high resolution measurements and fast acquisition over robustness to image noise. Moreover, PS patterns behave well in the presence of illumination defocus that naturally occurs with the use of digital projectors having large optical apertures.

Our main contribution is the proposal of a new structured light method called Chebyshev Phase Shifting which accurately captures highly detailed models at high speed, the construction of a prototype for recording three-dimensional hand shape and color varying in time, and its application to the measurement of hand tremors to be used in the study of involuntary motion conditions.

2. Background

Phase shifting (PS) 3D scanners measure scene shape by projecting sinusoidal patterns of particular spatial frequencies and capturing images of the scene under such illumination. The application of optical triangulation, by means of ray-plane intersection, to a geometrically calibrated system permits to calculate the position of a subset of 3D points on the scene of interest. As result, we create a dense set of 3D points, called 'point cloud', with scene geometric information and photometric data, as scene color. PS methods, as the general group of SL methods, rely on light reflected by the scene to measure its shape. Therefore, scene surface properties affect the result in various ways. For instance, transparent and mirror-like surfaces cannot be measured with standard methods. A common assumption made by SL methods is that all surfaces are Lambertian, *i.e.* they reflect light equally in every direction. In truth, the majority of scenes are not perfectly Lambertian nor mirror-like, but something in between those extreme cases. Our goal is to measure human hands which may have body hair and are covered with semi-translucent skin that allows light to scatter internally. Pattern inter-reflections due to the scene geometry is, in addition to the intrinsic surface properties, a source of measurements errors because they change the light intensity perceived at the cameras, as illustrated in Figure 3. Previous research [2] has shown that under these circumstances, measurement errors are minimized if the spatial frequency of the projected patterns are in narrow band, opposite to some commonly used patterns.

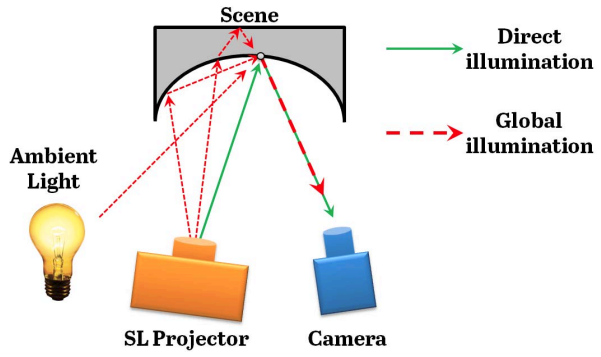

Figure 3: Global illumination, caused by inter-reflections, subsurface scattering, and external light sources, adds up to the direct illumination at the camera and may cause SL decoding errors.

2.1. Structured Light Systems

SL systems can be divided in three basic steps:

- **Encode**: a sequence of SL patterns is created taking into account the chosen method, user parameters, and the projector capabilities.
- **Acquisition**: the pattern sequence is projected onto the scene while the cameras capture images of them, usually synchronization is required.
- **Decode**: the sequence of captured images is processed with knowledge of the projected patterns to recover a map between camera and projector coordinates.

Encode, Acquisition, and Decode steps are shown in Figure 4. They are present in every SL system, however, they may not be explicitly separated and executed sequentially in actual implementations (e.g. a complete pattern may be created and stored beforehand, or they might be created on-the-fly during capture). Additional steps often follow to create range images, point clouds, or mesh models, possibly integrating several decoded coordinate maps.

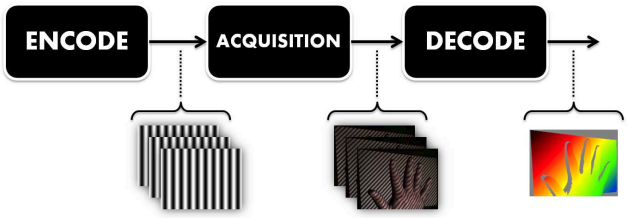

Figure 4: SL System Steps. Additional processing may follow as triangulation, multiple point cloud registration, and mesh creation.

2.2. Phase Shifting

We consider that projectors have a continuous image plane with coordinates in one direction varying from 0 to x_{max}. PS assigns to each pixel (x, y) in a pattern P an intensity value as described by Equation 1, where o_p and a_p are constants chosen to fit the projector dynamic range, f is the pattern frequency, and θ is the phase shift.

$$P(x, y) = o_p + a_p \cos(f\, x + \theta) \tag{1}$$

The intensity of a camera pixel (u, v) observing a scene illuminated by P is given by a similar relation, Equation 2, where o_c and a_c are constants of unknown value. x is the unknown phase of the projector pixel illuminating the scene point imaged at (u, v).

$$I(u, v) = o_c + a_c \cos(f\, x + \theta) \tag{2}$$

We use the formula of the cosine-of-a-sum to rewrite Equation 2 and isolate the known pattern shift θ from the

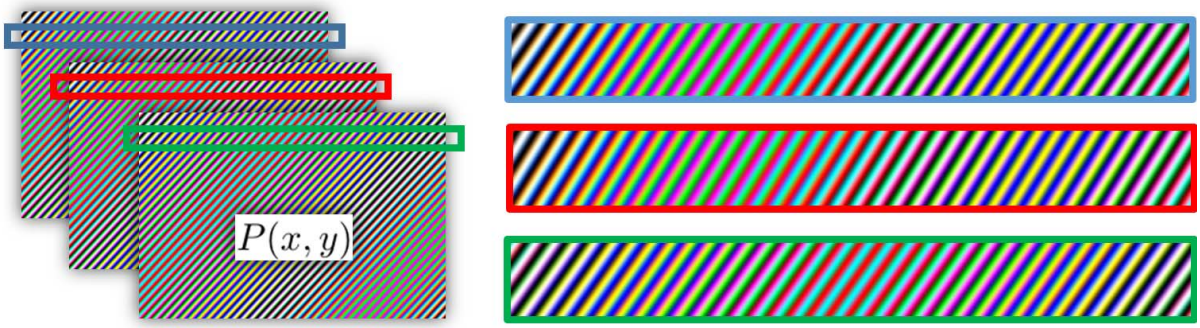

Figure 5: Sequence of three RGB CPS patterns: green channel $f_{105} = 1/28.20$, red channel $f_{106} = 1/27.94$, and blue channel $f_{107} = 1/27.68$. Each image corresponds to a different phase shift θ. They have a diagonal orientation to accommodate the diamond pixel configuration in the DLP LightCrafter 4500. Frequencies were chosen to create patterns with a period of about 28 pixels because this frequency range works well with human skin which is semi-translucent and causes subsurface light scattering.

unknown phase value x,

$$I(u,v) = o_c + a_c \cos\theta \cos(fx) - a_c \sin\theta \sin(fx), \quad (3)$$

$$I(u,v) = \begin{bmatrix} 1 & \cos\theta & -\sin\theta \end{bmatrix} \begin{bmatrix} o_c \\ a_c \cos(fx) \\ a_c \sin(fx) \end{bmatrix}. \quad (4)$$

Let be $\mathbf{u} = [o_c, a_c\cos(fx), a_c\sin(fx)]^T$ a 3-vector, where $\mathbf{u}(i)$ is its i^{th}-component. We find $\mathbf{u}$, the only unknown in Equation 4, projecting additional patterns with the same frequency but different phase shifts θ, and solving the resulting system of linear equations. Then, the unknown variables are recovered as follows

$$\begin{cases} o_c = \mathbf{u}(1) \\ a_c = \sqrt{\mathbf{u}(2)^2 + \mathbf{u}(3)^2} \\ \hat{x} = \text{atan2}(\mathbf{u}(3), \mathbf{u}(2)) \end{cases} \quad (5)$$

where $\hat{x}$ is a 'relative phase'. The original 'absolute phase' x is given by

$$x = \frac{2\pi n + \hat{x}}{f}, \quad (6)$$

where n is a non-negative integer. Equation 6 results from the periodicity of the trigonometric functions. The function 'atan2' returns a value between 0 and 2π. A trivial solution $n = 0$ exists if $f \geq 2\pi/x_{max}$. Otherwise n needs to be estimated in a separate process called 'phase unwrapping'.

2.3. Phase Unwrapping

Phase unwrapping finds the correct integer n which relates relative and absolute phases in Equation 6. Unwrapping may be performed using additional information such as a reference plane, or scene continuity assumptions, but we prefer to relay only on information extracted from the captured data. Some PS methods project patterns of several frequencies and recover a group of relative phases $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\}$ from them. If some of these phases correspond to the frequency $f_{unit} = 2\pi/x_{max}$, called 'unit frequency', then a simple unwrapping formula exists. Let us assume that $\hat{x}_i$ corresponds to a frequency f_i, that the frequency set meets the following relation

$$f_1 < f_2 < \cdots < f_N, \quad (7)$$

and that $f_1 \leq f_{unit}$. Then, the set of relative phases is unwrapped sequentially from x_1 to x_N with the formula,

$$\begin{cases} n_1 = 0 & \text{for } i = 1 \\ n_i = \left\lfloor \dfrac{x_{i-1} f_i - \hat{x}_i}{2\pi} \right\rceil & \text{for } i > 1 \end{cases}, \quad (8)$$

where the operator $\lfloor \cdot \rceil$ rounds its argument to the nearest integer, and the absolute phase x_i is calculated with Equation 6 and the unwrapping constant n_i.

3. Related Work

Salvi *et al*. [14, 15] survey several SL codification strategies. Discrete and continuous patterns are analyzed, as well as, spatial and time multiplexing techniques. We have not considered discrete codings, nor spatial multiplexing, because of their limited resolution. Instead, we have made a compromise and chosen a PS method with a small number projected patterns, so that we could create higher resolution models. PS methods have been extensively studied in the past. Here, we review only a few to describe some particular properties of them. Every PS method is based on the principles described in Section 2.2 but they differ from each other on the number of frequencies used, the number of projected patterns, the relation between the projected frequencies, and how these frequencies and estimated relative phases are used to recover a final absolute phase. Single-frequency PS (SPS) uses a single frequency in all patterns [16], consequently, f_{unit} must be chosen to

avoid unwrapping, or additional data is required. Shape measurements performed with f_{unit} are strongly affected by global illumination. Multiple-frequency PS (MPS) have patterns of several frequencies [13], including f_{unit}, and it applies the unwrapping equations in Section 2.3. Micro PS [3] is a method robust to global illumination that uses frequencies varying within a very narrow spectrum. It unwraps relative phases with a linear search in a look-up table (LUT) containing known pattern values at integer projector column positions. The LUT is necessary because f_{unit} is missing from its high-frequency patterns, but unwrapping is slow because the search must be executed for every pixel.

A group of PS methods robust to global illumination is based on the so-called 'number theoretical method' (NTM) [12]. They use a set of coprime high-frequencies meeting the requirements of the Chinese Remainder Theorem (CRT), which they apply to find f_{unit}. Embedded PS (EPS) [8] uses high frequency patterns within a very narrow frequency spectrum and it recovers f_{unit} using a property of the sum-of-cosines. EPS and methods based on NTM do relative phase unwrapping as in Section 2.3. Other methods are usually similar to one of these. Different are those methods that take advantage of color information. Color PS techniques have been researched for surface contouring of 3D objects [5, 10]. A color-encoded fringe pattern whose RGB components comprise three phase shifts is generated and projected onto the object. A color camera is used to capture one image which is separated into its RGB components creating three phase-shifted grayscale images. These images are then used as in SPS. Wust and Capson [18] proposed a color fringe projection method for surface profile measurement of neutral colored surfaces. Another methods utilizing color fringe projection can be found in various studies Zhang *et al*. [19], Zhang [20], and Ha *et al*. [4].

4. Chebyshev Phase Shifting

We propose a novel method called Chebyshev Phase Shifting (CPS) which projects sinusoidal phase shifting patterns of three different high frequencies, and from the images of these patterns recovers the absolute phase corresponding to f_{unit}. This knowledge allow us to unwrap the relative phases of each pattern and generate three projector-camera coordinate maps which can be combined to create an accurate 3D model of the scene. CPS is related to Chebyshev polynomials in that it uses the same recurrence. A Chebyshev polynomial of degree $n \geq 0$ is defined as,

$$T_n(x) = cos(n \arccos x) \quad \text{with} \quad x \in [-1, 1]. \tag{9}$$

Chebyshev polynomials can also be generated from the following recurrence relation:

$$\begin{cases} T_0(x) = 1, \quad T_1(x) = x, \\ T_{n+1} = 2xT_n(x) - T_{n-1}(x), \quad \text{for } n \geq 1 \end{cases} \tag{10}$$

Returning to Chebyshev PS, let $f_1 = \pi / x_{max}$ and $f_n = nf_1$, then $f_{unit} = f_2$ (f_1 is half the unit frequency). From the Chebyshev recurrences above we have:

$$\begin{cases} \cos(f_n x) = 2\cos(f_1 x)\cos(f_{n-1} x) - \cos(f_{n-2} x) \\ \sin(f_n x) = 2\cos(f_1 x)\sin(f_{n-1} x) - \sin(f_{n-2} x) \end{cases} \tag{11}$$

CPS has a positive integer n as the single parameter to adjust the frequency range of the projected patterns. The value of n uniquely defines three frequencies $\{f_n, f_{n-1}, f_{n-2}\}$ which are used to create PS patterns. Different applications may choose to create sequences with different number of patterns. For instance, if three or more phase shifts are available per frequency a separate vector $\mathbf{u}$ could be solved for each one, or alternatively, the offset o_c may be considered constant across all images to solve a single linear system of equations. Each one similar to Equation 4 but $\mathbf{u}$ is extended 7 components to include the sine and cosine of the three frequencies. In either case, the solution gives the following values:

$$\begin{cases} s_0 = \sin(f_n x), & c_0 = \cos(f_n x) \\ s_1 = \sin(f_{n-1} x), & c_1 = \cos(f_{n-1} x) \\ s_2 = \sin(f_{n-2} x), & c_2 = \cos(f_{n-2} x) \end{cases} \tag{12}$$

We reorder Equations 11 as follows:

$$\begin{cases} \cos(f_1 x) = \dfrac{\cos(f_n x) + \cos(f_{n-2} x)}{2\cos(f_{n-1} x)} = \dfrac{c_0 + c_2}{2c_1} \\ \cos(f_1 x) = \dfrac{\sin(f_n x) + \sin(f_{n-2} x)}{2\sin(f_{n-1} x)} = \dfrac{s_0 + s_2}{2s_1} \end{cases} \tag{13}$$

CPS exploits Equations 13 to estimate the phase corresponding to f_{unit} by noting that,

$$\hat{x} = \frac{1}{f_1} \arccos \frac{c_0 + c_2}{2c_1}, \tag{14}$$

$$\hat{x} = \frac{1}{f_1} \arccos \frac{s_0 + s_2}{2s_1}, \tag{15}$$

where c_1 and s_1 are never simultaneously zero. Equation 14 is applied if $|c_1| \geq |s_1|$, Equation 15 otherwise. Observe that 'arccos' returns values in the range $[0, \pi)$ ensuring that $\hat{x} \in [0, x_{max})$ as desired. Therefore, the value $\hat{x}$ computed by CPS is the phase value corresponding to f_{unit} and the unwrapping method described in Section 2.3 can be used to unwrap the patterns relative phases $\hat{x_0}$, $\hat{x_1}$, and $\hat{x_2}$:

$$\begin{cases} \hat{x_0} = & \dfrac{1}{f_n} & \text{atan2}(s_0, c_0) \\ \hat{x_1} = & \dfrac{1}{f_{n-1}} & \text{atan2}(s_1, c_1) \\ \hat{x_2} = & \dfrac{1}{f_{n-2}} & \text{atan2}(s_2, c_2) \end{cases} \tag{16}$$

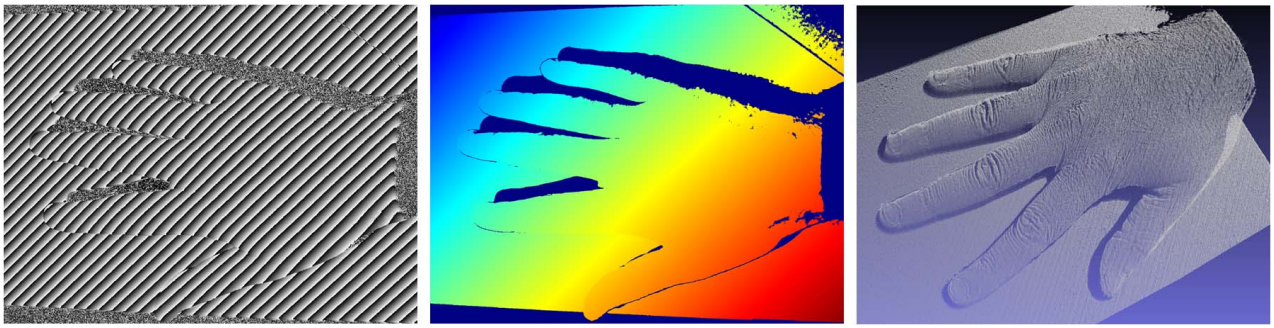

Figure 6: (Left) Relative phase image. (Middle) Unwrapped phase image: each color represents a different projector light plane, dark blue corresponds to pixels ignored because they are occluded or have low amplitude. (Right) Point cloud 3D triangulation result.

The absolute phases $\{x_0, x_1, x_2\}$ at each image pixel provide three separate but similar mappings between camera and projector pixel coordinates which can be combined into a single map with less measurement noise than each map individually. The combined coordinate map and the system geometric calibration permit to create a 3D model of the scene as in other SL methods.

Frequency Selection Example

We now illustrate how frequencies are chosen for a projector with 1024 columns: $x_{max} = 1024$ and $f_1 = \pi/1024$. If we set $n = 129$, the we have the following frequencies:

$$f_{129} = \frac{2\pi}{15.876},\ f_{128} = \frac{2\pi}{16.000},\ f_{127} = \frac{2\pi}{16.126}. \quad (17)$$

The value n was chosen to have patterns with a period of about 16 pixels. Increasing or decreasing n we adjust the frequencies as demanded by each specific application. We could create a 7 pattern sequence (the minimum possible for this method) by creating two patterns with f_{127}, three with f_{128}, and two with f_{129}. Longer frequencies are possible.

4.1. Difference from other Phase Shifting methods

CPS differs from SPS [16] in that it uses three frequencies instead of one. CPS does not impose shape constraints on the scene. Compared to MPS methods [13], CPS differs in how pattern frequencies are chosen, neither of them being f_{unit}. Different from Micro PS [3], CPS recovers f_{unit} and performs relative phase unwrapping without referring to a LUT, nor performing any search. It is also different from NTM methods [12] in how pattern frequencies are chosen (no co-prime frequency constraint) and how f_{unit} is recovered without the application of CRT. Pattern frequencies in CPS are chosen with a completely different formula than that of EPS [8] because different trigonometric properties are used. CPS exploits the sine and cosine multiple-angle formulas to recover the phase corresponding to f_{unit}. As far as the authors know, these relations have never been used in the context of PS, or other SL method for shape sensing.

5. Case study: Rapid Hand Shape and Color Reconstruction

Up to here, we described a general version of CPS which can be applied to measure regular scenes in replacement of other PS methods. Now, we describe how we have applied CPS to create 3D color models of a rapid trembling hand.

5.1. Prototype Description

Our prototype is shown in Figure 1. It consists of a structure in the shape of a 14" cube delimiting a working volume. A digital camera and projector are mounted on top. Our projector is a TI DLP LightCrafter 4500 with "diamond pixel" configuration (squared pixels rotated $45°$) and our camera is a Basler acA1920-155uc USB 3.0. Both devices are synchronized with hardware triggering. Image resolution is 1920×1200 and pattern resolution is 914×1140 pixels.

5.2. Encoding

Chebyshev Phase Shifting needs a minimum of 7 patterns to successfully unwrap the relative phases. This number is appropriate to capture static scenes, but their acquisition consumes an amount of time unsuitable for dynamic scenes. We have shortened the acquisition time to less than half by assigning each frequency to a different color channel. Effectively packing 9 CPS patterns (3 frequencies, 3 phase shifts) in three RGB patterns. Color allows us to pack more data in fewer images but color calibration is required to correctly match projected and observed color intensities. Decoding must also be adapted to consider the calibrated photometric model. We have set CPS parameter $n = 107$ to create diagonal patterns with $x_{max} = 1481$. Figure 5 shows the complete pattern sequence consisting of three RGB patterns: green channel $f_{105} = 1/28.20$, red channel $f_{106} = 1/27.94$, and blue channel $f_{107} = 1/27.68$. Each image corresponds to a different phase shift θ. The value of n was chosen so that the pattern periods were about 28 pixels, as this range is appropriate to measure with human skin. The skin surface is semi-translucent and produces subsurface light scattering.

5.3. Color calibration

Color calibration is required to correctly match the color on the captured images to those from the projected patterns. The three projected colors influence each camera color channel in a proportion a priori unknown. A phenomenon known as color crosstalk. We assume an affine model, as illustrated in Figure 7, to describe the relation between pattern and image colors (similar to [1]),

$$\mathbf{c} = \mathbf{A}\mathbf{p} + \mathbf{b}, \tag{18}$$

where $\mathbf{A} \in \mathbb{R}^{3\times3}$ is called "cross-talk matrix", $\mathbf{p}$ is a 3-vector containing the projected color intensities, $\mathbf{c}$ is a 3-vector with the colors as observed by the camera, and $\mathbf{b}$ is 3-vector of offsets that relates projector and camera intensity levels. The calibration identifies the values of $\mathbf{A}$ and $\mathbf{b}$. To solve the unknowns, we project a "black" pattern (corresponding to $\mathbf{p} = [0, 0, 0]^T$) and we use samples $\mathbf{c}_i$ from the observed camera intensities to calculate $\mathbf{b}$ by minimizing $\|\mathbf{c}_i - \mathbf{b}\|^2$ in the least squares sense,

$$\mathbf{b} = \frac{1}{N}\sum_i \mathbf{c}_i. \tag{19}$$

We project additional patterns to obtain projected-observed color correspondences $\mathbf{c}_i \leftrightarrow \mathbf{p}_i$. We rewrite Equation 18 as

$$\mathbf{c}_i - \mathbf{b} = \begin{bmatrix} r_i & 0 & 0 & g_i & 0 & 0 & b_i & 0 & 0 \\ 0 & r_i & 0 & 0 & g_i & 0 & 0 & b_i & 0 \\ 0 & 0 & r_i & 0 & 0 & g_i & 0 & 0 & b_i \end{bmatrix} \mathbf{a} \tag{20}$$

$$\mathbf{a} = [a_{11}, a_{21}, a_{31}, ..., a_{33}]^T,\ \mathbf{p}_i = [r_i, g_i, b_i]^T \tag{21}$$

We stack multiple instances of Equation 20, one for each color correspondence, and we solve $\mathbf{a}$ by least squares. The entries of $\mathbf{a}$ determine the calibration matrix $\mathbf{A}$.

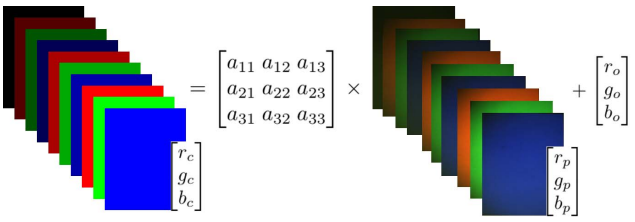

Figure 7: Color calibration: an affine model relates projected color intensities (r_p, g_p, b_p) to observed camera colors (r_c, g_c, b_c).

5.4. Acquisition

We project and capture images at 121 FPS. Our projector is configured to continually and cyclically project the sequence of three patterns stored in its firmware. In order to reach the required projection rate we reduced the quantization to 6 bits per pattern, so that, the complete sequence fits in the projector's frame buffer. Projection exposure was set to 7.5ms per pattern (2.5ms per color channel), as shown in Figure 8. Our camera continuously captures a stream of 8 bit RGB images. The system acquisition frame rate is appropriate to capture 6 Hz hand tremors.

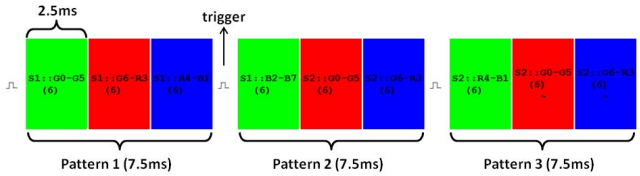

Figure 8: Projector timings: each pattern is projected in 7.5ms by assigning 2.5ms to each color channel. The complete sequence, quantized to 6 bits, fits within the projector's frame buffer.

5.5. Decode

We use a three-image sliding window to decode the incoming images and produce a range and a color image at each position in the image stream, as shown in Figure 9. We now reformulate Equation 4 to account for the color channel cross-talk present in the new acquisition scheme. We extend $\mathbf{u}$ to include three offset values (o_r, o_g, o_b), one for each color channel, and propose a new decoding equation:

$$I_{rgb}(u, v) = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{a}_1\mathbf{S}_1 & \mathbf{a}_2\mathbf{S}_2 & \mathbf{a}_3\mathbf{S}_3 \end{bmatrix} \mathbf{u}, \tag{22}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix},\ \mathbf{S}_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \end{bmatrix}. \tag{23}$$

Equation 22 models that each image intensity has contributions from all three projected colors, according to matrix $\mathbf{A}$. Equations 22 and 4 are equivalent if $\mathbf{A}$ is the identity matrix. We stack one Equation 22 for each image in the sequence and solve $\mathbf{u}$. Offsets (o_r, o_g, o_b) form an image without trace of the projected patterns, useful to assign colors to the 3D models. The other components in $\mathbf{u}$ are used to estimate relative and absolute phases as in Section 4.

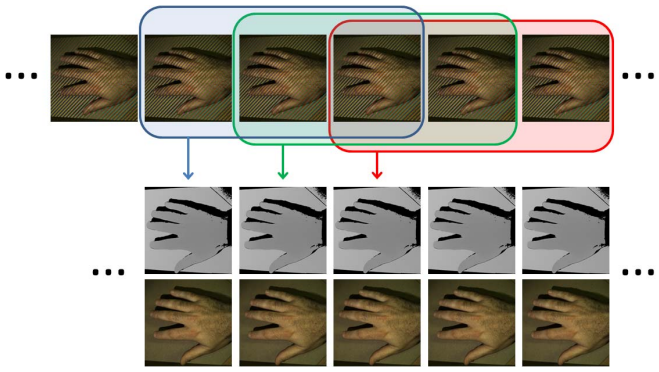

Figure 9: A three-image sliding window produces depth and color at the frame rate of the input stream: 121 FPS.

6. Results

We performed all experiments with our prototype, as described in Section 5.1. We used the projector-camera calibration tool [9] for geometric calibration. The system was color calibrated with a sequence of 10 constant color patterns, their color is listed in Table 1. Captured images were cropped to a region including only the projected color and 10000 randomly chosen samples from each were used to calculate the cross-talk matrix $\mathbf{A}$ shown in Equation 24. Observe that it is a diagonally dominant matrix. Values across rows or columns must not necessarily add up to 1.

$$\mathbf{A} = \begin{bmatrix} \mathbf{0.82} & 0.11 & 0.02 \\ 0.22 & \mathbf{0.71} & 0.08 \\ 0.02 & 0.14 & \mathbf{0.51} \end{bmatrix}. \quad (24)$$

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
r	0	85	0	0	170	0	0	255	0	0
g	0	0	85	0	0	170	0	0	255	0
b	0	0	0	85	0	0	170	0	0	255

Table 1: Projected constant color calibration patterns.

Figure 10 compares 3D models created with 7 grayscale 8 bit patterns projected at 60 FPS and with 3 RGB patterns at 121 FPS. Both models are highly detailed, but the model created from color patterns has more noise due to a smaller projection time (2.5ms each color versus 16.6ms a grayscale image) and because of higher quantization of the color 6 bit patterns compared to the 8 bit grayscale ones.

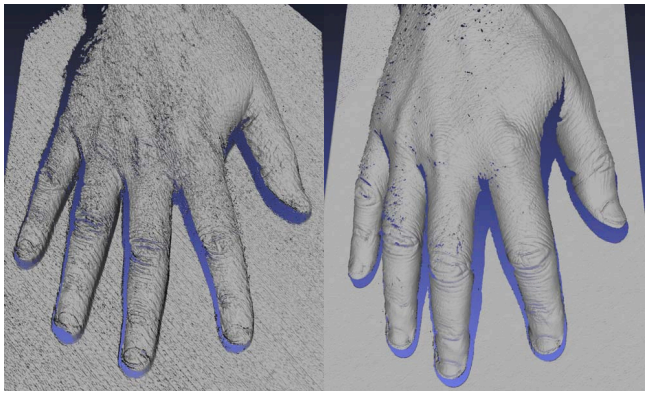

Figure 10: Color (left) and grayscale (right) point cloud comparison for static acquisition. Both models are highly detailed, however, the one created from color patterns presents more noise due to a smaller projection time and because of higher quantization of the projected 6 bit patterns versus the 8 bit grayscale ones.

Figure 11 shows the scene colors estimated during decoding. Note that all captured images have a projected pattern imposed. CPS succeeds in estimating a new image as seen if no pattern were projected. Colors found by CPS can be assigned to point cloud vertexes or rendered as 2D color textures. Figure 6 shows side-by-side relative and unwrapped phase images. CPS recovered an absolute phase map and created a highly detailed point cloud.

Figure 11: Color estimation: offset values (o_r, o_g, o_b) are the scene color at each camera pixel. They are used to create a color 3D point cloud (left), or may be rendered as a 2D texture (right).

Figure 2 is a screenshot of our system for rapid hand motion measurement. It shows the current camera image, the current point cloud with three control points, and a plot of the vertical displacement of the control points over 5s. The plot has a resolution of 8.3ms (600 data points). Rapid motions of the finger with the blue marker were captured.

7. Limitations

We described CPS and we showed that it successfully measures high speed hand models with great accuracy. However, the reader must be aware of some limitations that may be relevant when applied to other applications. As described, the fast acquisition of each sequence takes 7.5ms, it is not "instantaneous". If a scene were to change faster than that, the result would have additional errors, or CPS may fail. In addition, if a scene does not reflect light for some of the projected colors, then the images would not have information of the three projected frequencies.

8. Conclusion

We have presented a novel PS method called "Chebyshev Phase Shifting" that is substantially different from previous PS methods. CPS can be used to measure static scenes similar to other methods, but in this work we focused on how to apply CPS to measure dynamic scenes. We have discussed its applications in a case study to capture highly detailed high-speed 3D color models of rapid human hand movements. Finally, we have built a prototype which serves to collect data to be used in the study of PD and kinematics.

9. Acknowledgments

The research described in this paper was primarily funded by the National Science Foundation through grant No. IIP-1500249. The authors want to specially thank David Michael and Cognex Corp. for partially supporting the research that lead to this project.

References

[1] Dalit Caspi, Nahum Kiryati, and Joseph Shamir. Range imaging with adaptive color structured light. *IEEE Transactions on Pattern analysis and machine intelligence*, 20(5):470–480, 1998.

[2] M. Gupta, A. Agrawal, A. Veeraraghavan, and S.G. Narasimhan. Structured light 3D scanning in the presence of global illumination. In *CVPR 2011*, pages 713–720, June 2011.

[3] M. Gupta and S.K. Nayar. Micro Phase Shifting. In *CVPR*, pages 1–8, Jun 2012.

[4] Hyowon Ha, Jaesik Park, and In So Kweon. Dense depth and albedo from a single-shot structured light. In *3D Vision (3DV), 2015 International Conference on*, pages 127–134. IEEE, 2015.

[5] Peisen S Huang, Qingying Hu, Feng Jin, and Fu-Pen Chiang. Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Optical Engineering*, 38(6):1065–1071, 1999.

[6] Seiji Inokuchi, Kosuki Sato, and Fumio Matsuda. Range imaging system for 3D object recognition. In *Proceedings of the ICPR*, volume 48, pages 806–808, 1984.

[7] Joseph Jankovic. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4):368–376, 2008.

[8] D. Moreno, K. Son, and G. Taubin. Embedded phase shifting: Robust phase shifting with embedded signals. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2301–2309. IEEE, 2015.

[9] D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *3DIMPVT 2012*, pages 464–471. IEEE, 2012.

[10] Jiahui Pan, Peisen S Huang, and Fu-Pen Chiang. Color phase-shifting technique for three-dimensional shape measurement. *Optical Engineering*, 45(1):013602–013602, 2006.

[11] J.L Posdamer and M.D Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1–17, 1982.

[12] T. Pribanić, H. Dźapo, 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.

[13] Henrik O Saldner and Jonathan M Huntley. Temporal phase unwrapping: application to surface profiling of discontinuous objects. *Applied optics*, 36(13):2770–2775, 1997.

[14] J. Salvi, S. Fernandez, T. Pribanić, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666 – 2680, 2010.

[15] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004.

[16] V. Srinivasan, H. C. Liu, and Maurice Halioua. Automated phase-measuring profilometry: a phase mapping approach. *Appl. Opt.*, 24(2):185–188, Jan 1985.

[17] Ana Lisa Taylor Tavares, Gregory SXE Jefferis, Mandy Koop, Bruce C Hill, Trevor Hastie, Gary Heit, and Helen M Bronte-Stewart. Quantitative measurements of alternating finger tapping in parkinson's disease correlate with updrs motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Movement disorders*, 20(10):1286–1298, 2005.

[18] Clarence Wust and David W Capson. Surface profile measurement using color fringe projection. *Machine Vision and Applications*, 4(3):193–203, 1991.

[19] Li Zhang, Brian Curless, and Steven M Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*, pages 24–36. IEEE, 2002.

[20] Song Zhang. 3D range data compression with a virtual fringe projection system. In *SPIE Sensing Technology+ Applications*, pages 948902–948902. International Society for Optics and Photonics, 2015.