Stochastic Phase Estimation and Unwrapping

Mara Pistellato, Filippo Bergamasco, Andrea Albarelli, Luca Cosmo, Andrea Gasparetto and Andrea Torsello

DAIS, Ca'Foscari University of Venice, Via Torino 155, Venice, Italy

Keywords: Phase Shift, Structured Light, 3D Reconstruction.

Abstract: Phase-shift is one of the most effective techniques in 3D structured-light scanning for its accuracy and noise resilience. However, the periodic nature of the signal causes a spatial ambiguity when the fringe periods are shorter than the projector resolution. To solve this, many techniques exploit multiple combined signals to unwrap the phases and thus recovering a unique consistent code. In this paper, we study the phase estimation and unwrapping problem in a stochastic context. Assuming the acquired fringe signal to be affected by additive white Gaussian noise, we start by modelling each estimated phase as a zero-mean Wrapped Normal distribution with variance $\bar{\sigma}^2$. Then, our contributions are twofolds. First, we show how to recover the best projector code given multiple phase observations by means of a ML estimation over the combined fringe distributions. Second, we exploit the Cramér-Rao bounds to relate the phase variance $\bar{\sigma}^2$ to the variance of the observed signal, that can be easily estimated online during the fringe acquisition. An extensive set of experiments demonstrate that our approach outperforms other methods in terms of code recovery accuracy and ratio of faulty unwrappings.

1 INTRODUCTION

The recent evolution of increasingly affordable and powerful 3D sensors and the consolidation of fast reconstruction algorithms enabled the widespread adoption of 3D data in both consumer (Han et al., 2013) and industrial (Luhmann, 2010) off-the-shelf devices. As a consequence of the resulting increase of general interest in the subject, 3D data capturing and processing has become a trending topic in recent Computer Vision research. The richness of information coming from 3D data have been exploited in several fields of application, from industrial inspection systems (Luhmann, 2010), robot and machine vision (Prez et al., 2016), pipe inspection (Bergamasco et al., 2012), medical (Cheah et al., 2003; Tikuisis et al., 2001; Pellot et al., 1994) and entertainment applications (Han et al., 2013; Winterhalter et al., 2015).

While consumer applications place more emphasis on speed and performance, in an industrial setup accuracy is of greater importance. To this end, a lot of effort has been put in reconstruction techniques trading design simplicity and speed for higher precision in 3D recovery and robustness to surface characteristics of captured objects. A wide range of different techniques have been proposed over the last decades.

Different approaches can be usually categorized on the basis of the exploited physical principle and on the design of the adopted sensor. For instance, a time-of-flight setup combines a pulsating light emitter with a sensor which measures the round-trip time of the signal. Then, given the distance between the sensor and the artifact, a depth map can be computed and used for reconstruction (Lange and Seitz, 2001). Despite this technology have been proven reliable in some specific application, it suffers from two major drawbacks. First, the sensor does not usually provide a high resolution response. Moreover, it is particularly sensitive to signal interferences and surface response (*i.e.* artifact material). For these reasons, when resolution is of greater importance, triangulation-based approaches, especially when paired with high-end hardware, proper sensor calibration and advanced signal processing techniques, are better suited. Among the approaches, passive 3D sensing ones employ different cameras which are used to capture artifact's images at different angles in order to triangulate the single material points whose projection on the different images has been matched on the basis of purely photometric information. Differently, active 3D sensing technology exploits the projection of some structured pattern of known spatio-temporal structure onto the object, and recovers depth information by means of triangulation

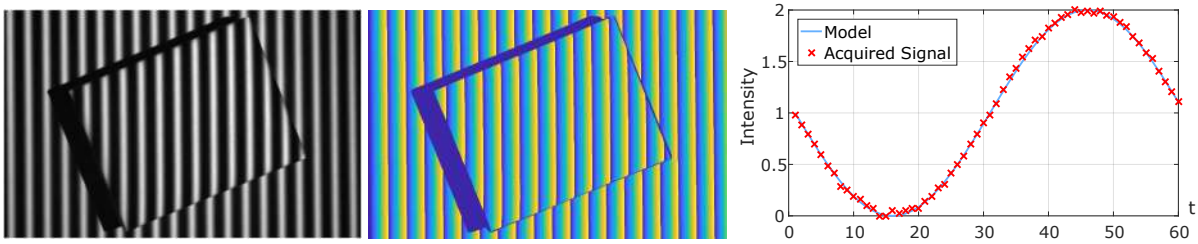

Figure 1: Left: Vertical sinusoidal pattern with a period length of 17 px. The pattern is shifted from right to left so that each pixel observe one complete period after m samples. Center: The recovered phase for each pixel. Since the period is shorter than the with of the projected image, we observe a phase ambiguity among the fringes. Right: Real-world example of the acquired signal compared to the projected ideal signal for a pixel with phase 0. Note how the samples are slightly noisy especially in the lower portion of the sine period.

between the rays emitted by the pattern projector and the corresponding image points, matched by means of signal decoding. Digital projector-based structured light is an example of such approaches. The main idea behind this technique is to decode from the image the corresponding projector coordinates (in pixels) to establish a map between the observed scene and the projector itself (camera-projector setup) or between different acquisitions of the scene (multi camera setup). In this paper we will not deal with the triangulation or matching problems, rather we will focus on the accurate recovery and decoding of the projected signal.

In literature, several different coding strategies have been proposed, each with a specific goal in mind. For instance, some approaches (der Jeught and Dirckx, 2016) allow to use a reduced number of patterns in order to increase the speed of the reconstruction and to enable real time operations. Other methods attain high speed by exploiting the interaction between the projected signal and the natural texture appearing in the scene (Vo et al., 2016). In (Fanello et al., 2016), the authors propose to infer the depth from the signal itself using a learning technique, dropping the actual triangulation step. Despite the abundance of recent specialized techniques, the most popular one is the long standing phase shift method (Srinivasan et al., 1984). This is particularly true in the industrial setting, since this approach allows high accuracy, resilience to noise and surface texture and great flexibility. The basic idea underlying phase-shift is indeed quite simple. The projected patterns are sine wave intensity frames which are periodic along one direction of the digital projector frustum. Several patterns, with different shift in the signal phase, are projected and retrieved over time, then the initial phase (i.e. the phase at frame 0) is reconstructed for each image pixel by means of convolution. Unfortunately, due to the periodic nature of the pattern in space, many pixels will be characterized by the same initial phase, thus a disambiguation step must be employed. This step takes the name of phase unwrapping.

The method introduced in this paper deals with both the phase recovery and unwrapping problems at once by casting them in a stochastic context. We assume that the acquired fringe signal is affected by additive white Gaussian noise so that we can model each estimated phase as a zero-mean Wrapped Normal distribution with variance $\bar{\sigma}^2$. We recover the most likely projector code, given multiple phase observations, by means of a ML estimation over the combined fringe distributions. Additionally, we exploit the Cramér-Rao bounds to relate the phase variance $\bar{\sigma}^2$ to the variance of the observed signal, that can be easily estimated online during the fringe acquisition. The validity of the assumptions and the effectiveness of our approach are then tested thoroughly with both real and synthetic experiments.

2 PHASE SHIFT

Phase-shift technique consists in generating a sinusoidal pattern spanning either the horizontal or vertical extent of the projector image plane (Fig. 1, Left). During the acquisition, the pattern is spatially shifted so that each projector pixel encompasses an entire period after m subsequent shifts. Together with m, the period length λ (measured in pixels) of the sine pattern is chosen a-priori. It affects both the phase difference of two neighbouring pixels along with the spatial ambiguity of all pixels in the image. Indeed, two adjacent pixels will exhibit a phase-difference of $\frac{2\pi}{\lambda}$. Consequently, points which are λ pixels away will be characterized by the same phase value.

The choice of λ is driven by two opposing needs. On one hand, a longer period will reduce the phase ambiguity along the image, with the extreme case of λ larger than the projector size causing no ambiguity at all. On the other hand, it is well known that the phase localization of the signal is proportional to the frequency, so we need to keep it small to increase the

accuracy. Usually, λ is kept in the order of dozens of pixels (Fig. 1, Center) and an unwrapping step is used to distinguish different fringes. The disambiguation technique is based on the idea of projecting multiple different sinusoidal signals (at different periods $\lambda_1 \dots \lambda_n$) that can be combined to get a unique code for each projector's pixel.

The problem of the aforementioned multi-phase shift approach is that the signal acquired by the camera is perturbed by different noise sources, causing an imperfect phase estimation. Possible sources are, for example: (I) The thermic white noise of the camera, especially at high gains or very short exposure times; (II) The non-linear interactions between neighbouring pixels due to intrinsic properties of each material (micro-facets and reflections causing a wrong signal response); (III) An imperfect mechanical and/or electronic functioning of the projecting device (particularly true for laser projectors), (IV) External noise sources like indoor ambient lighting subject to the oscillatory nature of the power outlet. Whatever the reason, a noisy phase estimation affects not only the accuracy of the reconstructed surface, but may cause a completely wrong phase unwrapping. Indeed, methods not particularly tolerant to phase errors produce a lot of erroneous codes for challenging materials, like brushed metals or glossy plastic.

2.1 Signal Model

Before describing our method, we start by formalizing the mathematical details of the projected signal and the corresponding noise model.

For any projector pixel P at coordinates (ξ, v), we project over time the following sinusoidal signal, discretized as a sequence of m subsequent samples:

$$Z(t) = \sin\left(2\pi\frac{t}{m} + \phi_P\right), \quad t = [0 \dots m). \tag{1}$$

Assuming a vertical pattern, the ideal pixel phase ϕ_P depends both on the period λ and its horizontal position on the image plane, according to the following simple relation:

$$\phi_P = 2\pi\left(\frac{\xi}{\lambda} - \left\lfloor\frac{\xi}{\lambda}\right\rfloor\right). \tag{2}$$

The value of ϕ_P cannot be measured directly, but it can be estimated from observed samples. Such values are subject to different noise sources affecting the acquisition process. As commonly performed in signal theory applications (Rife and Boorstyn, 1974; Macleod, 1998), we derive our observation model assuming to acquire a set of noisy samples:

$$z(t) = \sin\left(2\pi\frac{t}{m} + \phi_P\right) + w[t] \tag{3}$$

where $w[t]$ is a zero-mean white Gaussian noise of variance σ^2. Albeit not completely motivated by physical considerations, the assumption we made on the statistical nature of the noise represents a good trade-off between the simplicity of the model and the observed behaviour. Such convenience is supported by our experimental evidence (see Sec. 4 for details).

Given the acquired signal $\mathbf{z}$, the Maximum Likelihood estimator of the phase is defined as:

$$\hat{\phi_P} = \underset{\phi_P}{\operatorname{argmax}}\ p(\phi_P; \mathbf{z}) \tag{4}$$

where, considering our assumption on the statistical nature of the noise, we have:

$$p(\phi_P; \mathbf{z}) = \frac{1}{\sigma(2\pi)^{m/2}} e^{-\frac{1}{2\sigma^2}\sum_{t=0}^{m-1}\left(z(t) - \sin(2\pi\frac{t}{m} + \phi_P)\right)^2}. \tag{5}$$

Switching to the natural logarithm of the likelihood and arranging the terms, we obtain the common non-linear least-squares minimization:

$$\hat{\phi_P}(\mathbf{z}) = \underset{\phi_P}{\operatorname{argmin}} \sum_{t=0}^{m-1}\left(z(t) - \sin\left(2\pi\frac{t}{m} + \phi_P\right)\right)^2. \tag{6}$$

Since the least squares best fit based of a lower order Fourier series is exactly equivalent to the truncated DFT, the phase of the acquired single-tone signal can be easily recovered as:

$$\hat{\phi_P}(\mathbf{z}) = atan_2(x, y) + \pi \tag{7}$$

with

$$x = \sum_{i=0}^{m-1} cos\left(\frac{2\pi}{m}i\right)z(i); \qquad y = \sum_{i=0}^{m-1} sin\left(\frac{2\pi}{m}i\right)z(i)$$

and where $atan_2 : \mathbb{R}^2 \to [-\pi, \pi]$ evaluates the correct arctangent angle of x/y selecting the appropriate quadrant based on the arguments signs.

2.2 Stochastic Code Estimation

After the estimation of the phase of projected signal, our final goal is to recover the pixel horizontal (or vertical) coordinate ξ, usually referred as "*projector code*". To simplify the notation, let $\varphi(P) = \frac{\phi(P)}{2\pi}$ be the fractional part inside each projected fringe. In the typical case when λ is less than the projector's width

ξ_{max}, we observe multiple fringes along the image. Such stripes can be sequentially enumerated following the ordering induced by the projector's pixel grid. The fringe number associated to each coordinate ξ can be simply computed as:

$$\eta(\xi) = \left\lfloor \frac{\xi}{\lambda} \right\rfloor \in \mathbb{N} \tag{8}$$

so that each projector coordinate can be expressed in terms of the normalized phase and fringe number as:

$$\xi = (\eta(\xi) + \varphi(\xi))\lambda. \tag{9}$$

Values of $\varphi(\xi)$ are recovered as in (7), while λ is the constant value of the chosen period length. Thus, $\eta(\xi)$ is the only element that allows us to disambiguate among points with the same observed phase.

Similarly to other multi-phase approaches, we use n sinusoidal patterns with different co-prime period lengths $\lambda_1 \dots \lambda_n$ to disambiguate the fringes. In this setting, for each pixel P we estimate a sequence of n normalized phases $\hat{\varphi} = (\hat{\varphi}_1 \dots \hat{\varphi}_n)$ using the ML estimator shown in (6). Therefore, each $\hat{\varphi}_i$ is a sample from a random variable which we consider distributed as a Wrapped Normal of support $[0,1]$, unknown mean φ_i and variance $\bar{\sigma}_i{}^2$:

$$\hat{\varphi}_i \sim N_w(\varphi_i, \bar{\sigma}_i{}^2). \tag{10}$$

Note that, with a slight abuse of notation, with $\bar{\sigma}_i, i = 1 \dots n$ we denote the standard deviation of the random variable associated to the estimated phase of period λ_i (eq. 10), while with σ we express the standard deviation of $w[t]$, which is not related to the period length but only to the signal-to-noise ratio of the physical device. Albeit being different, the two standard deviations are related as described in Section 3.

Since $\bar{\sigma}_i$ is in general very small, we can approximate the Wrapped Normal efficiently with a standard Gaussian distribution (Kurz et al., 2014).

Therefore we can express the Probability Density Function relative to the observed phase in the i^{th} pattern as:

$$p_i^{\varphi}(\hat{\varphi}_i; \varphi_i, \bar{\sigma}_i) = \frac{1}{\sqrt{2\pi}\bar{\sigma}_i} e^{-\frac{d_c(\hat{\varphi}_i, \varphi_i)^2}{2\bar{\sigma}_i{}^2}} \tag{11}$$

where d_c is the signed circular distance (defined in the range $[-0.5, 0.5]$) of two normalized phases:

$$d_c(\varphi_1, \varphi_2) = sign(\varphi_1 - \varphi_2) min\{|\varphi_1 - \varphi_2|, 1 - |\varphi_1 - \varphi_2|\}$$

The PDF in (11) expresses the probability of observing a certain phase $\hat{\varphi}_i$ given the true (unknown) projected phase φ_i and the standard deviation on the phase observation $\bar{\sigma}_i$.

Since the projected phase φ_i is uniquely identified by the projector code ξ through equation (2), we can rewrite (11) in the following way:

$$p_i^{\xi}(\hat{\varphi}_i; \xi, \bar{\sigma}_i) = \frac{1}{\sqrt{2\pi}\bar{\sigma}_i} e^{-\frac{d_c\left(\hat{\varphi}_i, \frac{\xi}{\lambda_i} - \left\lfloor \frac{\xi}{\lambda_i} \right\rfloor\right)^2}{2\bar{\sigma}_i{}^2}} \tag{12}$$

This time, p_i^{ξ} models the probability of observing a certain phase $\hat{\varphi}_i$ at a specific projector location ξ.

When looking at all the n observed phases together, we can consider the components of vector $\hat{\varphi}$ as independent samples drawn from distribution (10) with means $\varphi = (\varphi_1 \dots \varphi_n)$ and fixed standard deviations $\bar{\sigma} = (\bar{\sigma}_1 \dots \bar{\sigma}_n)$. Similarly to what we did before with the signal, we consider ξ as a parameter and define the likelihood function describing the plausibility of observing a projector code ξ given the estimated $\hat{\varphi}$ and a pre-defined $\bar{\sigma}$ as

$$\begin{aligned} \mathcal{L}(\xi; \hat{\varphi}, \bar{\sigma}) &= \prod_{i=1}^{n} p_i^{\xi}(\xi; \hat{\varphi}_i, \bar{\sigma}_i) \\ &= \prod_{i=1}^{n} \frac{1}{\bar{\sigma}_i\sqrt{2\pi}} e^{-\frac{d_c\left(\hat{\varphi}_i, \frac{\xi}{\lambda_i} - \left\lfloor \frac{\xi}{\lambda_i} \right\rfloor\right)^2}{2\bar{\sigma}_i{}^2}} \end{aligned} \tag{13}$$

Given phase values estimated from the acquired signal, we can compute the code that most likely generated such phases by maximizing the Likelihood

$$\hat{\xi} = argmax_{\xi \in [0, \xi_{max})} \quad \mathcal{L}(\xi; \hat{\varphi}, \bar{\sigma}) \tag{14}$$

2.3 Our Proposed Method

Following the theoretical discussion made so far, we now summarize our proposed stochastic code recovery approach. Before starting the acquisition, one must choose how many periods to project and their period length in pixels.

To avoid ambiguity along the projector's codes space, all period lengths must be co-prime and their *LCM* smaller than the projector width (or height, if we project horizontal patterns). Moreover, one must specify the expected standard deviations of estimated phases, *i.e.* the vector $\bar{\sigma}$.

It is worth spending a couple of words here, because $\bar{\sigma}$ is in fact the only required parameter of our algorithm. First, we formulated the solution assuming a possibly different $\bar{\sigma}_i$ for each projected period. This accounts the fact that, for the microscopic characteristics of the surface, different sine frequencies may exhibit different signal noises. Moreover, longer periods are less localizable in space, so it is a good practice to use more samples m for long periods than for

the short ones. Second, this approach allows us to not exclude a-priori any value of the estimated $\hat{\varphi}$. Indeed, the vector $\bar{\sigma}$ is essentially a weight acting on all the acquired patterns. One may choose to keep extremely short or long periods, or to use very few samples, and modify the values of the relative $\bar{\sigma}_i$ accordingly.

This is an important aspect when we compare this method with other unwrapping techniques. Indeed, approaches like (Lilienblum and Michaelis, 2007) fail when facing noisy observations. We prefer to accept such values and compute the best possible code rather than having no code at all. Our goal, in fact, is to exploit measured phases in the best possible way.

Once $\bar{\sigma}$ is chosen and the n sinusoidal patterns (one for each period) are acquired, for each pixel of the camera we estimate the phases $\hat{\varphi}_1 \dots \hat{\varphi}_n$ using equation (7).

At this point, for each camera pixel, we search for an initial integer estimate of the projector code by taking the value of $\dot{\xi} \in \{0, 1, \dots \xi_{max}\}$ for which $\mathcal{L}(\dot{\xi}; \hat{\varphi}, \bar{\sigma})$ is maximum. Due to the different spatial resolution between camera and projector, it is unlikely that each camera pixel observe exactly one projector pixel. Indeed, one of the strengths of phase-shifting is that one can precisely recover the projector coordinate with sub-pixel precision.

In other words, the code $\hat{\xi}$ that maximizes the Likelihood can assume any real value between 0 and ξ_{max}.

Unfortunately, due to the signed circular distance, we cannot give a direct analytical solution for the global maximum $\hat{\xi}$. However, if we restrict the search in a small neighborhood of $\dot{\xi}$ so that the circular distance never wraps, and take the logarithm of the Likelihood, $\mathcal{L}(\xi; \hat{\varphi}, \bar{\sigma})$ becomes a parabola for which we can compute the maximum by sampling three distinct points on it.

Therefore, once we identified the integer part of the projector code $\dot{\xi}$, we recover the sub-pixel maximum $\hat{\xi}$ with the following formula:

$$\hat{\xi} = \frac{\mathcal{L}(\dot{\xi}+1; \hat{\varphi}, \bar{\sigma}) - \mathcal{L}(\dot{\xi}-1; \hat{\varphi}, \bar{\sigma})}{4\mathcal{L}(\dot{\xi}; \hat{\varphi}, \bar{\sigma}) - 2\big(\mathcal{L}(\dot{\xi}+1; \hat{\varphi}, \bar{\sigma}) + \mathcal{L}(\dot{\xi}-1; \hat{\varphi}, \bar{\sigma})\big)} \tag{15}$$

3 $\bar{\sigma}$ LOWER-BOUNDS

Now that we described the general algorithm, we still need to clarify how to provide reasonable values for the vector $\bar{\sigma}$.

One simple way is to empirically measure the phase error on a set of repeated experiments and compute its standard deviation.

There are two drawbacks in this approach. First, this sort of "calibration" must be performed every time a new object is acquired or any other condition changes the expected signal-to-noise ratio of acquired sinusoids. Second, the operation cannot be performed along with the acquisition because the exact phase of each acquired signal is not known, as it depends on the scene geometry.

To overcome that, the only way is to project the same known phase for all the projector pixels (essentially setting $\lambda = 1$) to collect statistics on its distribution. Unfortunately, this special pattern is useless for 3D reconstruction so it will just consume projector time to calibrate the parameter needed for our method.

Instead of directly calibrate $\bar{\sigma}$, we can empirically measure the variance of the signal noise $w[t]$ and relate it to the variance of the phase estimator.

Since ϕ_P is an unknown deterministic parameter of some probability function via the unbiased estimator $\hat{\phi_P}$, the variance of the estimator is subject to the Cramér-Rao bound:

$$\mathrm{var}(\hat{\phi_P}) \geq \frac{1}{J(\phi_P)} \tag{16}$$

where $J(\phi_P)$ is the Fisher information defined as:

$$\begin{aligned} J(\phi_P) &= E\left[\left(\frac{\partial \ln p(\mathbf{z}; \phi_P)}{\partial \phi_P}\right)^2\right] \\ &= -E\left[\frac{\partial^2 \ln p(\mathbf{z}; \phi_P)}{\partial^2 \phi_P}\right] \end{aligned} \tag{17}$$

By substituting (5) in (17) we obtain:

$$\begin{aligned} \ln p(\mathbf{z}; \phi_P) \quad = \quad & \ln\left(\frac{1}{\sigma(2\pi)^{m/2}}\right) \\ & -\frac{1}{2\sigma^2}\sum_{t=0}^{m-1}\left(z(t) - \sin\left(2\pi\frac{t}{m} + \phi_P\right)\right)^2 \end{aligned} \tag{18}$$

$$
\begin{aligned}
\frac{\partial^2 \ln p(\mathbf{z};\phi_P)}{\partial^2 \phi_P} &= -2K\Bigg[\sum_{t=0}^{m-1}\cos^2\left(2\pi\frac{t}{m}+\phi_P\right) \\
&\quad -\sum_{t=0}^{m-1}\sin^2\left(2\pi\frac{t}{m}+\phi_P\right) \\
&\quad +\sum_{t=0}^{m-1}\sin\left(2\pi\frac{t}{m}+\phi_P\right)z(t)\Bigg] \\
E\left[\frac{\partial^2 \ln p(\mathbf{z};\phi_P)}{\partial^2 \phi_P}\right] &= -2K_1\Bigg[\sum_{t=0}^{m-1}\cos^2\left(2\pi\frac{t}{m}+\phi_P\right) \\
&\quad -\sum_{t=0}^{m-1}\sin^2\left(2\pi\frac{t}{m}+\phi_P\right) \\
&\quad +\sum_{t=0}^{m-1}\sin\left(2\pi\frac{t}{m}+\phi_P\right)E[z(t)]\Bigg]
\end{aligned} \tag{19}
$$

where $K = -\frac{1}{2\sigma^2}$. Since

$$
E[z(t)] = \sin\left(2\pi\frac{t}{m}+\phi_P\right) + E[w[t]] \tag{20}
$$

and $E[w[t]] = 0$ by definition, we obtain:

$$
\begin{aligned}
E\left[\frac{\partial^2 \ln p(\mathbf{z};\phi_P)}{\partial^2 \phi_P}\right] &= -2K\sum_{t=0}^{m-1}\cos^2\left(2\pi\frac{t}{m}+\phi_P\right) \\
\mathrm{var}(\hat{\phi_P}) &\geq \frac{\sigma^2}{\sum_{t=0}^{m-1}\cos^2\left(2\pi\frac{t}{m}+\phi_P\right)}
\end{aligned} \tag{21}
$$

Equation (21) states that the variance of our phase estimator must be greater than a value proportional to the variance of the Gaussian white noise σ^2 and inversely proportional to the number of samples m. Indeed, since $\cos^2(x) \leq 1$, we have:

$$
\sum_{t=0}^{m-1}\cos^2(2\pi\frac{t}{m}+\phi_P) \leq m. \tag{22}
$$

This makes sense, as we expect to reduce the estimation error either by acquiring more samples or by increasing the signal-to-noise ratio. With this new formulation, the value of σ can be empirically estimated online with the acquisition of the sinusoids, just by computing the observation errors with respect to our ideal projected pattern.

Even if we did not solved the problem completely, we experimentally observed that the empirically estimated phase standard deviation is actually very close to the theoretical lower-bound (21). Hence, by just using a value of $\bar{\sigma}$ equal to the lower bound, we almost always obtain satisfactory results.

4 EXPERIMENTAL EVALUATION

We designed a set of synthetic and real-world experiments to verify the correctness of the assumptions we made in the previous sections and to compare our method with other multi phase-shift approaches.

The real-world experiments were performed with a structured-light scanner developed in our lab. The device is composed by a single MatrixVision BlueFox3 5Mpix camera and an LG projector with a resolution of 1920×1080 px. Camera and projector relative pose was calibrated using the fiducial markers descried in (Bergamasco et al., 2011). We manually corrected the projector gamma to obtain a camera signal response as linear as possible. Once calibrated, the defined gamma settings remained unchanged throughout the experiments.

4.1 Phase and Signal Standard Deviation

We performed a first set of experiments to assess the validity of our statistical noise model on a real scene. We considered three different kinds of planar surfaces, namely: matte white plastic, a brushed aluminum metal sheet and a dark coloured cardboard. The goal is to compare the acquired signal standard deviation σ with the standard deviation of the phase estimator $\bar{\sigma}_i$. To do that, we generated a pattern composed by a variable number of samples m with a period equal to one. This way, each pixel P in a scene observed exactly the same true phase $\phi_P = 0$, allowing us to empirically estimate the standard deviation of the phase error among pixels. Similarly, by knowing the phase, we computed the error of acquired signal with respect to the projected ideal sinusoid. In this way we can test if our zero-mean white Gaussian noise assumption is supported by experimental evidence.

In Figure 2 we show the results of this experiment for the three different materials. On the top row we plotted the histogram of the error of raw acquired signal for a sinusoidal pattern composed by $m = 60$ samples.

In all cases, the signal was analyzed by collecting the errors in a ROI of size 400×300 pixels selected to cover the central region of the projected area. We can see that the empirical distributions follow quite well our supposed Gaussian model with a mean close to zero. Especially for the cardboard material, we report a little positive bias that we guess is caused by blooming effect. In fact, blooming clearly bias the acquired intensity by always overshooting the actual value. Nevertheless, we think that the proposed as-

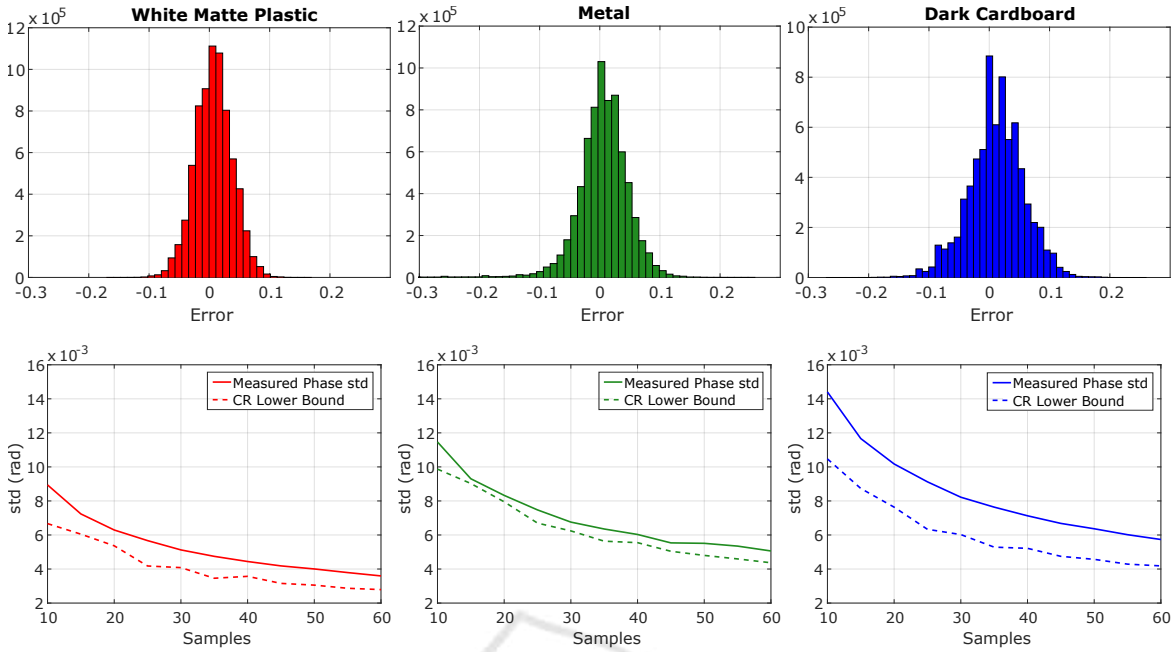

Figure 2: Relation between the acquired signal distribution and standard deviation σ (top row) with the theoretical lower-bound and the empirical phase standard deviation $\bar{\sigma}$ (bottom row) for the three different chosen materials.

sumptions are fair enough to approximate the real behaviour of the phenomenon. For each material, we computed the empirical standard deviation σ (from the $m = 60$ samples case) to compute the lower-bound according to (21).

In the second row of Figure 2 we plotted both the lower-bound and the empirically measured phase standard deviation varying the number of samples m for the three different materials. As expected, the empirical standard deviation of the estimated phase is proportional with the standard deviation of the signal. In fact, the white matte plastic exhibits a lower error due to its nearly Lambertian nature. Metal sheet is more noisy due to small reflections and dark cardboard is the worst since the low albedo is heavily decreasing the signal-to-noise ratio. In all the cases the estimated phase standard deviation is slightly above the theoretical lower bound computed from the signal. This validates our claim that the given bound can be effectively used to give a good estimation of the unknown $\bar{\sigma}$.

4.2 Code Estimation Comparisons

We performed a set of synthetically generated tests to compare the code recovery accuracy of our method with respect to a famous multi-phase shift approach proposed by Lilienblum and Michaelis (Lilienblum and Michaelis, 2007) and a single phase-shift approach using Graycoding to disambiguate the fringes and just one periodic pattern to obtain the code.

In all tests we generated 3 patterns with $\lambda_1 = 17$, $\lambda_2 = 23$ and $\lambda_3 = 27$ pixels and simulated a 1920×1080 pixels projector. Then we additively perturbed the synthetic phases of each pixel with a zero-mean Gaussian with different standard deviations. For each experiment we recovered the projector codes with the different methods and compute the RMS with respect to the ground truth.

In the first row of Fig.3 we plotted the comparison as RMS and percentage of outlier codes. For both methods, we considered a pixel as outlier if the distance wrt. the ground-truth was greater than half the minimum period length. From this test we can observe how our method consistently provides lower code RMS, less variability in the estimation and a lower number of outliers.

In the case of single phase-shift disambiguated with Graycoding (Fig.3, second Row), we observe how the recovered codes are less accurate than the multi-phase approach as we consider only a single pattern (and hence the output coding error is almost linearly related to the amount of phase perturbation introduced). The outlier ratio is consistently lower than the multi phase-shift approaches as we assumed that the Graycoding technique never fails to disambiguate the correct fringe. Nevertheless, the good inlier ratio does not justify its adoption for the higher overall RMS exhibited.

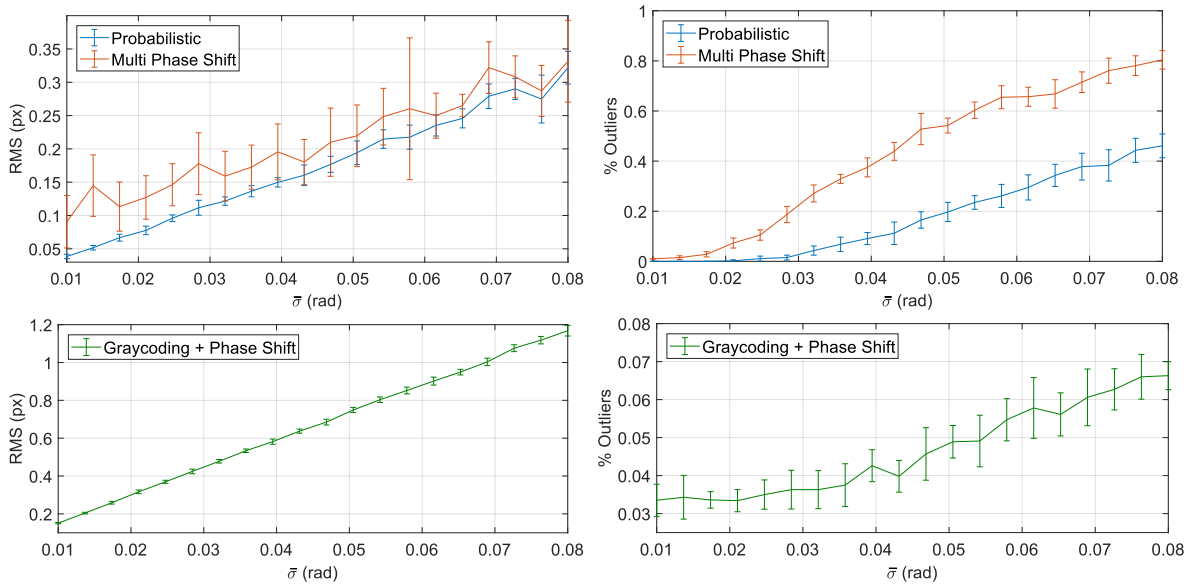

Figure 3: Comparison of our approach with respect to (Lilienblum and Michaelis, 2007) (top row) and single phase-shift disambiguated with Graycoding (bottom row). First and second column show respectively the coding RMS and the code outlier percentage varying the input phase error σ.

4.3 Projector Delay Recovery

One of the key strengths of our method is the resilience against noisy phase estimations. To demonstrate that, we tested the approach with the challenging task of recovery the correct code even if the projector and camera are not synchronized.

This can be useful in practice because ensuring a proper synchronization of the camera frames with the projected patterns usually requires a custom electronics that increases the cost and complexity of the scanner.

For simplicity, we assumed that the exposure time of the camera is far lower than the display time of each projected pattern. This way, all the acquired samples $z(t)$ are affected by a fixed unknown integer shift but the camera never acquires in-between two projector frames.

We generated a synthetic 3-periods signal composed by $m = 60$ samples (as in previous experiments) perturbed with a variable zero-mean noise with variance σ^2. Then, we shifted all the patterns by a random shift $0 \leq k \leq m$. Afterwards, we run our method for each possible circular shift k of the signal collecting, for each pixel, the value of the maximum likelihood obtained. Pixel wise, we compared all the maximum likelihood values to select the shift corresponding to the higher one. The result is that each pixel votes for the shift producing the higher maximum likelihood of the codes.

Since the data was synthetically generated, we marked as inliers all the pixels that actually voted for the correct shift and plotted the inlier percentage against the noise in Fig.5.

The experiment shows an inlier percentage greater than 50% (sufficient to correctly recover the unknown shift) for input signal noise std. up to 0.45, which is more than 5 times the noise levels we measured for all the three different materials in Fig.2. Therefore, we are confident that in the vast majority of real world cases the synchronization may be avoided with no severe consequence on the unwrapping result.

4.4 Qualitative Evaluation

Finally, we qualitatively evaluated the 3D triangulated meshes obtained both with our proposed method and the number-theoretical approach (Lilienblum and Michaelis, 2007).

In Fig.4 we show the range-maps obtained while reconstructing a metallic kitchen sink (top row) and a matte chalk bas-relief (bottom). As pre-processing, before triangulation we just filtered all the codes with a value outside the range $0 \ldots \xi_{max}$.

Especially in the more challenging metallic object, our method provides a denser and less noisy reconstruction. In the optimal scenario of the Lambertian white artefact, our method outperforms (Lilienblum and Michaelis, 2007) in high curvature regions where self-reflections usually lead to higher coding errors.

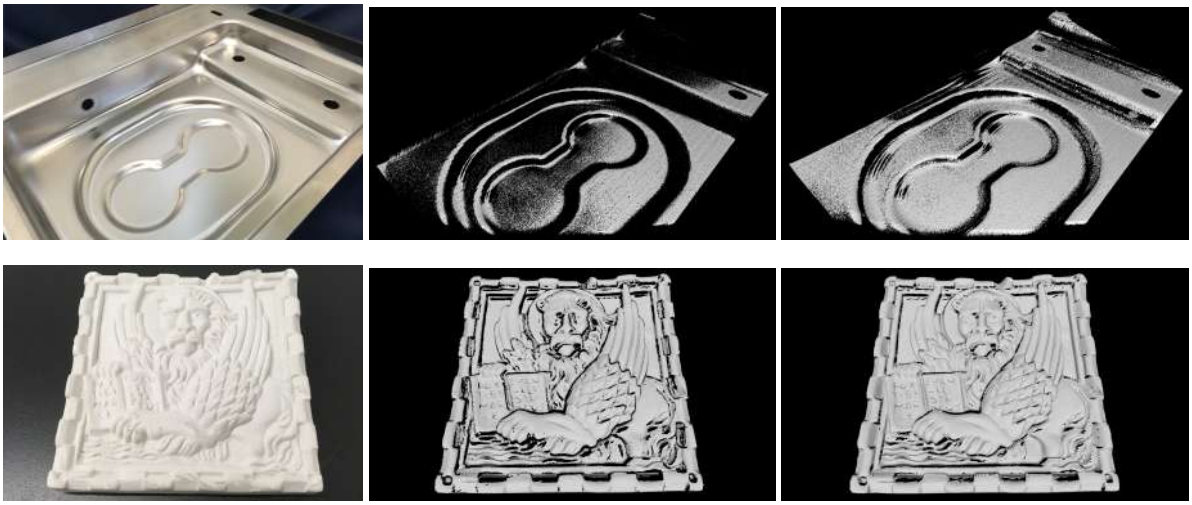

Figure 4: Examples of reconstructions for two different objects. The second column shows the triangulated mesh using (Lilienblum and Michaelis, 2007); while the third column shows the mesh obtained applying our method.

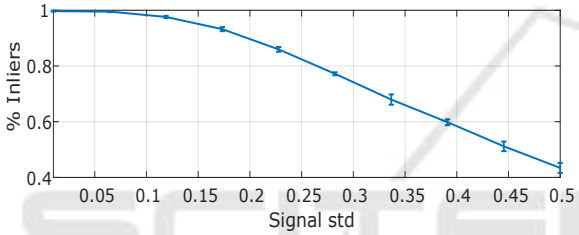

Figure 5: Projector delay recovery: percentage of correctly estimated shifts (inliers) against the acquired noise standard deviation.

5 CONCLUSIONS

In this paper we proposed a novel phase-shift technique to simultaneously disambiguate among different phase periods and estimate the final projector code. Our method is based on statistical inference on the observed phases which we assume being affected by zero-mean Gaussian white noise. Under this assumption, we derived a Maximum Likelihood estimator for the final projector code taking into account all the observations in an optimal way.

Additionally, we exploit the Cramér-Rao bound to derive a lower-bound for the phase estimator variance according to the variance of acquired signal. This way, the estimation of the vector $\bar{\sigma}$ (the only parameter required by our method) can be efficiently computed on-line with the acquisition of the sinusoidal patterns.

We tested our method both synthetically and with a real camera-projector structured light setup. From the the experiments, we observed how the empirically estimated phase variance is close to the theoretical lower-bound, suggesting that we can effectively use that value for the subsequent code recovery. Comparisons show how our method can outperform Graycoding and number theoretical multi-phase shift methods not only in terms of accuracy but also considering the number of correctly estimated codes. This leads to coded images that are in general denser, even for challenging materials.

REFERENCES

Bergamasco, F., Albarelli, A., and Torsello, A. (2011). Image-space marker detection and recognition using projective invariants. pages 381–388.

Bergamasco, F., Cosmo, L., Albarelli, A., and Torsello, A. (2012). A robust multi-camera 3d ellipse fitting for contactless measurements. pages 168–175.

Cheah, C.-M., Chua, C.-K., Tan, K.-H., and Teo, C.-K. (2003). Integration of laser surface digitizing with cad/cam techniques for developing facial prostheses. part 1: Design and fabrication of prosthesis replicas. *International Journal of prosthodontics*, 16(4).

der Jeught, S. V. and Dirckx, J. J. (2016). Real-time structured light profilometry: a review. *Optics and Lasers in Engineering*, 87(Supplement C):18 – 31.

Fanello, S. R., Rhemann, C., Tankovich, V., Kowdle, A., Escolano, S. O., Kim, D., and Izadi, S. (2016). Hyperdepth: Learning depth from structured light without matching. pages 5441–5450.

Han, J., Shao, L., Xu, D., and Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Transactions on Cybernetics*, 43(5):1318–1334.

Kurz, G., Gilitschenski, I., and Hanebeck, U. D. (2014). Efficient evaluation of the probability density function of a wrapped normal distribution. In *2014 Sensor Data*

Fusion: Trends, Solutions, Applications (SDF), pages 1–5.

Lange, R. and Seitz, P. (2001). Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics*, 37(3):390–397.

Lilienblum, E. and Michaelis, B. (2007). Optical 3d surface reconstruction by a multi-period phase shift method. *JCP*, 2(2):73–83.

Luhmann, T. (2010). Close range photogrammetry for industrial applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):558–569.

Macleod, M. D. (1998). Fast nearly ml estimation of the parameters of real or complex single tones or resolved multiple tones. *IEEE Transactions on Signal Processing*, 46(1):141–148.

Pellot, C., Herment, A., Sigelle, M., Horain, P., Maître, H., and Peronneau, P. (1994). A 3d reconstruction of vascular structures from two x-ray angiograms using an adapted simulated annealing algorithm. *IEEE transactions on medical imaging*, 13(1):48–60.

Prez, L., Rodrguez, I., Rodrguez, N., Usamentiaga, R., and Garca, D. (2016). Robot guidance using machine vision techniques in industrial environments: A comparative review. *Sensors (Switzerland)*, 16(3).

Rife, D. and Boorstyn, R. (1974). Single tone parameter estimation from discrete-time observations. *IEEE Transactions on Information Theory*, 20(5):591–598.

Srinivasan, V., Liu, H. C., and Halioua, M. (1984). Automated phase-measuring profilometry of 3-d diffuse objects. *Appl. Opt.*, 23(18):3105–3108.

Tikuisis, P., Meunier, P., and Jubenville, C. (2001). Human body surface area: measurement and prediction using three dimensional body scans. *European journal of applied physiology*, 85(3):264–271.

Vo, M., Narasimhan, S. G., and Sheikh, Y. (2016). Texture illumination separation for single-shot structured light reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2):390–404.

Winterhalter, W., Fleckenstein, F., Steder, B., Spinello, L., and Burgard, W. (2015). Accurate indoor localization for rgb-d smartphones and tablets given 2d floor plans. volume 2015-December, pages 3138–3143.