Sparse Multi-Path Corrections in Fringe Projection Profilometry

Yu Zhang
Nanjing University

Daniel Lau
University of Kentucky

David Wipf
Amazon

Abstract

Three-dimensional scanning by means of structured light illumination is an active imaging technique involving projecting and capturing a series of striped patterns and then using the observed warping of stripes to reconstruct the target object's surface through triangulating each pixel in the camera to a unique projector coordinate corresponding to a particular feature in the projected patterns. The undesirable phenomenon of multi-path occurs when a camera pixel simultaneously sees features from multiple projector coordinates. Bimodal multi-path is a particularly common situation found along step edges, where the camera pixel sees both a foreground and background surface. Generalized from bimodal multi-path, this paper examines the phenomenon of sparse or N-modal multi-path as a more general case, where the camera pixel sees no fewer than two reflective surfaces, resulting in decoding errors. Using fringe projection profilometry, our proposed solution is to treat each camera pixel as an underdetermined linear system of equations and to find the sparsest (least number of paths) solution by taking an application-specific Bayesian learning approach. We validate this algorithm with both simulations and a number of challenging real-world scenarios, demonstrating that it outperforms state-of-the-art techniques.

1. Introduction

In a digital camera, each pixel on the focal plane array records the average light measured across the pixel's field of view, and, in some cases, the pixel simultaneously records light from multiple paths, such as when a pixel sees the edge of the foreground object along with the non-occluded background object. In this situation, the collected intensity/color will be a weighted sum of the foreground and background intensity/color, depending on how much of the foreground object takes up the pixel's field of view (Fig. 1). Separating out the foreground and background contributions plays a fundamental role in a range of classic imaging tasks, such as image matting, where the alpha channel represents the weighting of the foreground and background compo-

Figure 1. Illustrations of the causes of sparse multi-path and its representations in both 2D and 3D imaging. From left to right are the FoV of a pixel on the sensor array during the imaging procedure, a classic image matting task [30], and an ideal step edge compared with a reconstructed step edge under the effect of multi-path in the structured light or time-of-flight imaging techniques [55].
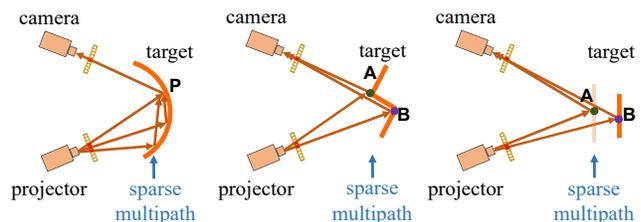
Figure 2. Illustration of sparse multi-path scenarios in (left) a curved reflective surface, (middle) a step edge, and (right) a semi-transparent surface in front of an opaque target with foreground pixel **A** and background pixel **B**.

nents [47, 31]. In 3D imaging, the collecting of foreground and background light paths along a step edge can result in either a predictable smoothing of the step edge in the 3D reconstruction or a catastrophic, unpredictable displacement of edge points [36, 54]. Fig. 2 illustrates three kinds of challenging sparse multi-path scenarios, including a curved reflective surface, a step edge, and a semi-transparent surface in front of an opaque target. Regardless of the outcome, the multi-path problem is unavoidable in any scanning situation, including structured light [23, 53] and time-of-flight (ToF) [36, 34, 16], but it is an especially prevalent problem in structured light as very little research has been performed on the topic, and it has been largely ignored by the computer vision and computational imaging community until recently.

With the rapid development of deep learning in com-

puter vision, there is now a diverse range of neural network topologies [15, 42, 41, 57] that can process 3D data. This has likewise driven a demand for large model databases that include a diverse range of labeled objects, such as body scans SMPL-X [40], face scans Florence [1], object scans PartNet [35], Redwood [7], indoor scene scans ScanNet [9], and outdoor road and vehicle scans KITTI [17]. As more and more work is done building AI models that can recognize and reconstruct these existing objects, there will be an increasing demand for 3D scanners to acquire similarly diverse objects, perhaps containing specularities and other hard-to-scan features. In order to address the need for a large volume of these objects, scanners will be required to handle these hard-to-scan features in a fully automated way.

A common active scanning technique for building high resolution, 3D object databases is structured light imaging (SLI), which involves projecting and capturing a series of striped or structured patterns and reading the patterns as they reflect off a target object's surface [3, 19, 32, 21, 23, 50, 39, 44]. Because structured light scanners, including single pattern techniques [49, 18, 56] and multiple pattern techniques [19, 32, 5, 27], need to actively project patterns onto the target surface, they often suffer from multi-path issues in which a camera pixel sees light from multiple projector positions, making accurate scanning impossible. It is a common and unavoidably important issue in many scanning situations where a target surface includes a step edge or a specular surface. Multi-path results, at best, in easily identifiable outliers like spikes along edges, which can be removed by filtering or, at worst, in imperceptible changes to the surface like a blurring of edges. These small errors can be catastrophic in certain applications, such as digital dentistry where digital scans are replacing plastic molds for the manufacturing of replacement crowns.

With regard to multi-path interference and its cancellation, many papers have been devoted to ToF sensing using unique hardware that needs to be customized or modified. Structured light scanners, on the other hand, can be constructed from commodity components [53] and are, therefore, widely studied; however, very few solutions exist for the multi-path problem, and relevant theories to address multi-path have not been firmly established. In this paper, we present an inexpensive and practical approach to address this N-modal multi-path issue by treating each pixel as an underdetermined linear system of equations and finding the sparsest (least number of paths) solution. We do so without any hardware modifications [39] or additional requirements for novel/customized patterns [8, 22]. As a result, it is easy for our algorithms to be integrated into existing popular and broadly used phase shifting structured light scanners. As part of this process, we combine a powerful sparse Bayesian learning (SBL) framework with novel application-specific constraints. In aggregate, these contributions lead to state-of-the-art results in addressing multi-path problems regarding fringe projection profilometry.

2. Related Work

Sparse Multi-Path While multi-path has long been a concern in studies of structured light, very few solutions exist short of relying on an alternate imaging technique like deflectometry [25]. Nevertheless, multi-path is gaining attention in the research literature. Dedrick [10] is an early documented example of examining multi-path interference in SLI but without an effective algorithm for extracting the absolute paths. Alternatively, Courture *et al.* [8] designed special patterns that, while resistant to inter-reflections, would not be recognized as traditional phase shifting patterns. Zhang *et al.* [54] recently proposed a solution for bimodal multi-path relying on scanning a target with increasing pattern frequencies in order to measure the constructive and destructive interference in the captured patterns.

By taking a number of scans at various spatial frequencies, Zhang *et al.* [54] showed that it is possible to separate the signals from the foreground and background surfaces with the potential to produce two decoupled depth estimates for each pixel seeing exactly two surfaces. However, Zhang *et al.*'s technique requires both a large number of unique spatial frequencies to separate paths and long sequences of video. Therefore, while a traditional SLI scanner accumulates three spatial frequencies and perhaps 24 frames of video, one following Zhang's approach might collect 10 or more spatial frequencies with 80 or more frames of video. Yu *et al.* [53] avoided the extra spatial frequencies by limiting their filtering to detecting multi-path rather than reconstructing the separate paths. In addition, neither Zhang *et al.* nor Yu *et al.* attempted to address more than two light paths, and even these two paths were limited to the field of view of a particular pixel. Hence, these approaches could not reconstruct a path, for example, that came from a specular reflection from the left or right of the projector's field of view.

Dense Multi-Path The majority of multi-path research on more than two paths for any 3D imaging modality is based on light transport modeling [45, 26, 20, 6], which is a dense multi-path model where the light observed by a camera pixel is derived from both a direct path from projector to surface to camera [4] and an ambient plus dense multi-path component where all reflections in the scene, specular and diffuse, collectively combine into a single, constant or DC signal [43, 28]. Accordingly, the goal of dense multi-path correction is to separate the direct light path from the indirect illumination/global illumination [24, 38]. To this end, Nayar *et al.* [37] separated the direct and global illumination components of a scene by means of projecting a high frequency binary illumination pattern and its complement.

The assumption is that the multi-path reflections of the high frequency pattern will be so diverse that their combination will be the same for both the pattern and its complement. With regard to choosing the best frequency, higher frequencies are better as long as they are not so high as to be blurred by the modulation transfer functions of the projector and camera. From Nayar *et al.* [37], Gupta and Nayar [23] developed a structured light scanning technique, called micro phase shifting, to reconstruct unambiguous phase images without the low frequency phase, thereby making the scans resilient to the effects of dense multi-path. The actual micro phase shifting is not responsible for removing multi-path, but is rather just a way to perform phase unwrapping by employing only high spatial frequency patterns, which are assumed to be immune to dense multipath. As such, Gupta and Nayar's technique fails for sparse multi-path [54], where high frequency spatial patterns are equally susceptible to sparse multi-path interference. Overall, sparse multi-path, especially the $N \geq 2$ multiple path case, remains a challenging and unsolved problem in the field.

3. N-Modal Multi-Path Model

In fringe pattern projection [58], a series of phase-shifted sinusoidal patterns are projected onto the object with the pixels of the projected patterns generalized as:

$$x_t = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi f \frac{m}{M} - \frac{2\pi t}{T}\right), \qquad (1)$$

where x_t is the intensity of the pixel in the mth of M-rows; f is the spatial frequency of the fringe pattern that is equal to the number of sinusoidal periods from top to bottom in the projection area; and t is the frame index of the pattern in its sequence of T frames numbered from 0 to $T-1$. From the camera's perspective, a captured pixel is generalized as:

$$y_t = a + r\cos\left(\theta - \frac{2\pi t}{T}\right), \qquad (2)$$

where y_t is the intensity of the pixel; a is the average intensity of the pixel over the T patterns; and r is the observed amplitude of the projected sinusoid seen by the camera pixel. The magnitude, r, is then derived according to:

$$r = \left\{ \left(\sum_{t=0}^{T-1} y_t \sin(\frac{2\pi t}{T})\right)^2 + \left(\sum_{t=0}^{T-1} y_t \cos(\frac{2\pi t}{T})\right)^2 \right\}^{\frac{1}{2}}, \qquad (3)$$

while the angle, $\theta = 2\pi f \frac{m}{M}$, is derived according to:

$$\theta = \arctan \frac{\sum_{t=0}^{T-1} y_t \sin(\frac{2\pi t}{T})}{\sum_{t=0}^{T-1} y_t \cos(\frac{2\pi t}{T})}, \qquad (4)$$

where the projector coordinate, m, visible to the camera pixel is extracted from θ using an appropriate phase unwrapping scheme [58]. Once m is uniquely identified, we can derive the corresponding world Cartesian coordinate using calibrated lens models for the camera and projector.

In order to introduce the N-modal multi-path model, it is instructive to examine the basis for the bimodal model. In Zhang *et al.* [54], the authors explained how paths interact as a function of the spatial frequency, f, to produce standing waves of constructive and destructive interference and how to decouple two light paths by taking advantage of this interesting cue. However, in this section, we drop the assumption that the number of observed multi-paths is limited to two unique paths in order to incorporate the general $N \geq 2$ multi-path model. To start, assume we perform a sequence of K scans using a projector with M rows, each row being defined by the sinusoidal frequency $\{f_k : k = 1, 2, ..., K\}$ and the normalized projector row coordinate $\{w_m = \frac{m}{M} : m = 1, 2, ..., M\}$.

For each projected sequence, a pixel in the camera records a magnitude, r_k, and a phase, θ_k, such that after all K scans, each camera pixel is defined by the vector $\boldsymbol{y} = [r_1 e^{-j2\pi f_1 \theta_1}, r_2 e^{-j2\pi f_2 \theta_2}, ..., r_K e^{-j2\pi f_K \theta_K}]^{\mathsf{T}}$, modeled according to:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}, \qquad (5)$$

where $\boldsymbol{\Phi}$ represents the dictionary matrix holding all possible light paths equal to:

$$\begin{bmatrix} e^{-j2\pi f_1 w_1} & e^{-j2\pi f_1 w_2} & \cdots & e^{-j2\pi f_1 w_M} \\ e^{-j2\pi f_2 w_1} & e^{-j2\pi f_2 w_2} & \cdots & e^{-j2\pi f_2 w_M} \\ e^{-j2\pi f_3 w_1} & e^{-j2\pi f_3 w_2} & \cdots & e^{-j2\pi f_3 w_M} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi f_K w_1} & e^{-j2\pi f_K w_2} & \cdots & e^{-j2\pi f_K w_M} \end{bmatrix} \qquad (6)$$

and $\boldsymbol{x} = [x_1, x_2, ..., x_M]^{\mathsf{T}}$ represents the solved light path in which the element, x_m, represents the total amount of light coming from all pixels in the mth projector row.

To solve for $\boldsymbol{x}$, we formulate the problem in terms of finding an unknown signal of interest from an incomplete or limited number of observations. When the relationship between the signal of interest and the observations is given by an underdetermined system of linear equations, the system has an infinite number of possible solutions. In general, accurate estimates of the signal of interest that are consistent with the measurements require additional regularity constraints. In the case of the ℓ_0 norm, we want to find the solution to the underdetermined set of equations given by:

$$\boldsymbol{x}_{opt} = \arg\min_{\boldsymbol{x}} \left\{ ||\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{x}||_2^2 + \lambda ||\boldsymbol{x}||_0 \right\} \qquad (7)$$

where $\boldsymbol{x}_{opt}$ is the sparsest solution, i.e., the solution with the least number of light paths between camera and projector. Note that feasible (or nearly feasible) solutions to $\boldsymbol{y} \approx \boldsymbol{\Phi}\boldsymbol{x}$

are unique if they are sufficiently sparse, meaning a sufficiently small ℓ_0 norm [13]. Hence, as long as the light paths are sparse, then solving Eq. 7 is likely to find them.

4. N-Modal Multi-Path Reconstruction

In this section, we first introduce the basic sparse estimation algorithm that underpins the proposed multi-path reconstruction technique and then explain our extension technique to handle application-specific constraints. There is no existing reference that we are aware of that implements these same real, non-negative constraints within a broader complex sparse Bayesian learning framework. Nor is there any reference that explicitly highlights the need for a method that can handle correlated columns within the present application context.

All of the algorithm-related sections are simply concerned with minimizing Eq. 7 over x under various conditions. More specifically, Section 4.1 focuses on minimizing Eq. 7 with no constraints on x, and Section 4.2 adds the constraints that x must be real and non-negative (one of our main contributions; see Algorithm 1).

4.1. Basic Sparse Estimation Model

Because solving the unconstrained case Eq. 7 is NP-hard, it is a common practice to replace the non-convex, discontinuous ℓ_0 regularizer with the convex ℓ_1 norm relaxation $\|x\|_1 = \sum_i |x_i|$ [14, 12]. Unfortunately, though, when Φ has highly correlated columns, as will necessarily be the case for the application considered herein, then this convex formulation will not represent a good approximation to Eq. 7. Likewise, greedy approaches for at least locally minimizing Eq. 7, e.g., orthogonal matching pursuit (OMP), are highly susceptible to being trapped at poor extrema, especially with correlated designs.

Consequently, instead of problematic convex relaxations or greedy methods, we invoke a Bayesian modeling paradigm [48] for solving both Eq. 7 and constrained alternatives involving complex data. For this purpose, let us simplify the problem by assuming $\mathbf{y}$ and Φ are purely real and define the Gaussian likelihood as

$$p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left[-\tfrac{1}{2\lambda}\|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2\right], \qquad (8)$$

where, for now, we assume that λ is a known noise variance parameter. Next, we assume a parameterized, zero-mean, Gaussian distribution as a prior over x. Specifically, we have

$$p(\boldsymbol{x};\boldsymbol{\gamma}) \propto \exp\left[-\tfrac{1}{2}\boldsymbol{x}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{x}\right], \quad \boldsymbol{\Gamma} \triangleq \mathrm{diag}[\boldsymbol{\gamma}], \quad (9)$$

where $\boldsymbol{\gamma}$ indicates a vector of unknown variance hyperparamters. Because both likelihood and prior are Gaussian, the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\gamma})$ is also Gaussian with

mean $\boldsymbol{\mu_x}$ and covariance $\boldsymbol{\Sigma_x}$ given by

$$
\begin{aligned}
\boldsymbol{\mu_x} &= \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top\boldsymbol{\Sigma_y}^{-1}\boldsymbol{y} \\
\boldsymbol{\Sigma_x} &= \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top\boldsymbol{\Sigma_y}^{-1}\boldsymbol{\Phi}
\end{aligned}
\qquad (10)
$$

respectively, where

$$\boldsymbol{\Sigma_y} \triangleq \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^\top + \lambda\boldsymbol{I} \qquad (11)$$

is the conditional covariance of $\boldsymbol{y}$ given $\boldsymbol{\gamma}$ obtained by marginalization over $\boldsymbol{x}$. From these expressions, it is clear that if $\boldsymbol{\gamma}$ is a sparse vector with mostly zero-valued entries, then by virtue of its diagonal positioning and lefthand-side multiplication in Eq. 10, the estimator $\boldsymbol{\mu_x}$ will have a matching sparsity profile or support pattern (i.e., the indeces of nonzero values will be the same). Of course, for this framework to be a successful strategy for producing sparsity, we require some way of determining a reasonable estimate for $\boldsymbol{\gamma}$ that favors pushing many or most elements towards zero while preserving a good posterior fit for the data.

A Type-II maximum likelihood approach to this problem involves treating $\boldsymbol{x}$ as nuisance variables that can be marginalized out of the model [33]. The resulting objective then only depends on $\boldsymbol{\gamma}$ and, furthermore, because the resulting convolution of the Gaussians integral is available in closed-form [48], we can equivalently minimize the negative log-likelihood expression:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\gamma}) &= -\log \int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x};\boldsymbol{\gamma})d\boldsymbol{x} \\
&\equiv \boldsymbol{y}^\top\boldsymbol{\Sigma_y}^{-1}\boldsymbol{y} + \log|\boldsymbol{\Sigma_y}|.
\end{aligned}
\qquad (12)
$$

Although this objective is non-convex in $\boldsymbol{\gamma}$, the expectation-maximization (EM) algorithm can be applied for minimization purposes, treating $\boldsymbol{x}$ as the hidden data [11, 48]. For the E-step, this involves simply computing the posterior moments of $p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\gamma})$ as given by Eq. 10, while the M-step reduces to the update:

$$\gamma_i \leftarrow (\boldsymbol{\mu_x})_i^2 + (\boldsymbol{\Sigma_x})_{ii}, \quad \forall i. \qquad (13)$$

Given that we only require the i-th diagonal element of $\boldsymbol{\Sigma_x}$ (as opposed to the full matrix) for this update, the overall computational complexity of Eqs. 10 and 13 is linear in the dimension of $\boldsymbol{x}$ and quadratic in the dimension of $\boldsymbol{y}$, where generally $\dim[\boldsymbol{y}] \ll \dim[\boldsymbol{x}]$ by construction. Once we have a hyper-parameter estimate, $\hat{\boldsymbol{\gamma}}$, we can compute a final estimator, $\hat{\boldsymbol{x}}$, as the posterior mean from Eq. 10 evaluated at $\hat{\boldsymbol{\gamma}}$. We refer to this as our baseline ℓ_0 norm sparse estimator. Additionally, if analogous derivations are carried through assuming that $\boldsymbol{y}, \boldsymbol{x}$, and $\boldsymbol{\Phi}$ are complex (with complex Gaussian underlying distributional assumptions), then we only need to convert transposes to Hermitian transposes to obtain the requisite EM updates. In the next section, we

will present an alternative formulation of this sparse learning algorithm that serves to both elucidate why optimizing Eq. 12 can lead to better sparse estimates and provide a natural entry point for introducing practically-useful constraints in the complex domain.

4.2. Application-Specific Constraints via Alternative Formulation

Thus far, we have essentially deferred a direct search for a sparse estimator x to the indirect search for an ostensibly sparse γ, which can then later produce a sparse x. But the mechanism by which this occurs is not at all obvious, especially given the assumption of a Gaussian prior that typically favors diverse or non-sparse solutions. However, if we apply [51, Theorem 2], we can convert the estimation problem of minimizing $\mathcal{L}(\gamma)$ in γ-space to an equivalent problem in x-space, thereby facilitating transparent analysis and a clear pathway for introducing useful application-specific constraints. In particular, it can be shown that minimizing Eq. 12 to find some $\hat{\gamma}$ and then computing $\hat{x} = \mu_x$ using Eq. 10 evaluated at $\hat{\gamma}$ is equivalent to minimizing:

$$\mathcal{L}(x) \triangleq \|y - \Phi x\|_2^2 + \lambda f(x; \Phi, \lambda), \qquad (14)$$

where

$$f(x; \Phi, \lambda) \triangleq \inf_{\gamma \succcurlyeq 0} \sum_{i \in \mathcal{S}} \frac{x_i^2}{\gamma_i} + \log |\Phi \Gamma \Phi^\top + \lambda I| \qquad (15)$$

is a penalty function, parameterized by Φ and λ, and $\mathcal{S} = \{i : x_i \neq 0\}$ represents the set of indices, of elements in x, that are not equal to zero. Note also that if any $x_i = 0$, then the corresponding optimal γ_i will also equal zero, given that the log-determinant term in Eq. 15 is a concave, non-decreasing function of γ. And with complex data, the only difference is that, again, a Hermitian transpose must be substituted.

Interestingly, although there is generally no closed-form solution for f itself, it nonetheless can be shown to be a strictly concave, non-decreasing function of each coefficient magnitude, $|x_i|$, for all $\lambda \geq 0$, and, hence, it naturally favors exactly sparse solutions [51], meaning many $x_i = 0$. Additionally, given the general determinant identity:

$$\log \left| \Phi \Gamma \Phi^\top + \lambda I \right| = \log \left| \tfrac{1}{\lambda} \Phi^\top \Phi + \Gamma^{-1} \right| + \dots$$
$$\dots + \log |\Gamma| + \log |\gamma I|, \quad (16)$$

when we optimize Eq. 15, the effective penalty on x will explicitly depend on $\Phi^\top \Phi$ through the action of a volumetric log-det measure. This effect allows regularization via f to compensate for high correlations (i.e., large off-diagonal elements in $\Phi^\top \Phi$) in producing higher quality sparse estimates [52]. Therefore, for present purposes, where by design Φ will necessarily have highly correlated columns, we

are essentially replacing the ℓ_0 norm with f, as opposed to a standard convex ℓ_1 norm alternative that often fails with correlated designs.

Furthermore, despite the implicit, variational expression for f, it is nonetheless possible to optimize Eq. 14 via a convenient iterative reweighted ℓ_1-norm minimization process [52]. In brief, this process requires iterating:

$$x \leftarrow \arg\min_x \|y - \Phi x\|_2^2 + \lambda \sum_i w_i |x_i|,$$

$$\gamma_i \leftarrow |x_i| w_i^{-\frac{1}{2}}, \quad \forall i \qquad (17)$$

$$w \leftarrow \mathrm{diag} \left[\Phi \left(\lambda I + \Phi \Gamma \Phi^\top \right)^{-1} \Phi \right]^{\frac{1}{2}},$$

which is guaranteed, based on properties of majorization-minimization algorithms [46], to reduce or leave unchanged Eq. 14 until a local minimum (or other critical point) is reached. Standard convex solvers can, therefore, be used for implementing the x update, and if complex data is assumed, only the w update needs be changed by simply substituting a Hermitian transpose as before (the convex solver for the x update will, of course, also need appropriate modification).

Additionally, there is another key advantage of this revised formulation that is especially salient for present purposes. Assuming y and Φ are complex-valued instead of real, then the optimal $\hat{x}$ obtained from minimizing either Eq. 12 or 14 will likely be an unconstrained complex vector as well. However, we know *a priori* that x should be a non-negative, real number by virtue of the physical problem constraints for the present structured light illumination problem. Introducing such constraints into the original Bayesian formulation from Section 4.1 is not possible because the marginalization required by type-II maximum likelihood is no longer tractable. In contrast, optimizing Eq. 14 with constraints on x is much more straightforward. Essentially, we need only modify the convex x update from Eq. 17 to include these constraints, which in many cases involves quite simple modifications.

For example, proximal gradient methods are commonly applied to solving weighted ℓ_1-norm minimization problems such as the x update [2]. In brief, this approach is based on constructing a decoupled quadratic upper bound on the term $g(x) \triangleq \|y - \Phi x\|_2^2$ via

$$g(x) \leq g(\tilde{x}) + (x - \tilde{x})^H \nabla g(\tilde{x}) + \tfrac{L}{2}\|\tilde{x} - x\|_2^2, \quad (18)$$

where $\tilde{x}$ is an arbitrary parameter vector of the bound, $L \in \left[\|\Phi^H \Phi\|, \infty \right)$, and

$$\nabla g(\tilde{x}) = \Phi^H (y - \Phi \tilde{x}). \qquad (19)$$

This is because $\nabla g(x)$ is Lipschitz continuous with the Lipschitz constant $\|\Phi^H \Phi\|$, in which case such a quadratic

bound can always be constructed. Therefore, for a fixed value of $\tilde{\boldsymbol{x}}$, if we replace $g(\boldsymbol{x})$ with this upper bound, then after some algebraic manipulations, solving

$$\min_{\boldsymbol{x}} \ (\boldsymbol{x} - \tilde{\boldsymbol{x}})^H \nabla g(\tilde{\boldsymbol{x}}) + \frac{L}{2}\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_2^2 + \lambda \sum_i w_i |x_i| \quad (20)$$

can be replaced with i independent/decoupled problems of the form

$$\min_{x_i} \ \frac{L}{2}\left|x_i - \left(\tilde{x}_i - \frac{1}{L}\nabla g(\tilde{\boldsymbol{x}})\right)\right|^2 + \lambda w_i |x_i|. \quad (21)$$

Solving Eq. 21 has a simple closed-form solution, including when real, non-negative constraints are enforced on each x_i. In this case, the optimal solution is easily shown to be

$$
\begin{aligned}
x_i^{imag} &\leftarrow 0, & (22)\\
x_i^{real} &\leftarrow \max\left[\text{real}\left(\tilde{x}_i - \frac{1}{L}\nabla g(\tilde{\boldsymbol{x}})\right) - \frac{\lambda w_i}{L}, 0\right], & \forall i.
\end{aligned}
$$

After computing a new $\boldsymbol{x}$ with Eq. 22, we can then set

$$\tilde{\boldsymbol{x}} \leftarrow \boldsymbol{x} \quad (23)$$

to update the bound and repeat. By the basic properties of proximal gradient algorithms [2], iterating Eqs. 19, 22, and 23 is guaranteed to converge to an optimal solution of the weighted ℓ_1-norm minimization problem in Eq. 17 when real, non-negative constraints are imposed. Therefore, these iterations form an inner loop for solving the stated constrained sub-problem, while Eq. 17 serves as an outer loop for the overall task. In aggregate, the combined iterations, even if not executed until convergence, are guaranteed to reduce or leave unchanged Eq. 14 subject to $\boldsymbol{x}^{imag} = \boldsymbol{0}$ and $\boldsymbol{x}^{real} \geq \boldsymbol{0}$.

The pseudo-code of our core sparse Bayesian learning algorithm for minimizing Eq. 7 with real and non-negative constraints is shown in Algorithm 1. Overall, the algorithm is quite simple to implement. Basically, it involves iterating the steps of Eq. 17, where the argmin over x is computed by iterating Eqs. 19, 22, and 23 as an inner loop.

5. Experiments

To evaluate our method, we consider a classic compressive sensing problem in which the goal is to reconstruct sparse signals with N nonzero elements from K observations. Here, $\boldsymbol{\Phi}$ is defined by Eq. 6, where we use spatial frequencies $\{f_k : k = 1, 2, ..., K, K = 60\}$ and set M to 1000. We then perform a series of experiments, where N ranges from 1 to 12. In each experiment, we perform 500 trials with each trial (i) randomly assigning N nonzero elements in $\boldsymbol{x}$, in the range [0.2, 1.2], and the rest to 0; (ii) reconstructing $\boldsymbol{x}$; and (iii) calculating the corresponding error.

The chamfer error [15] is shown in Fig. 3 (left) on the log scale for the original signal with N nonzero elements

---

Algorithm 1 SBL Algorithm with Constraints

Input: $\boldsymbol{\Phi}, \boldsymbol{y}, L, \lambda$
Output: $\boldsymbol{x}$
   $\boldsymbol{w} \leftarrow 1$
   $\boldsymbol{x}^{imag} \leftarrow 0$
   $\boldsymbol{x}^{real} \leftarrow max(real(\boldsymbol{\Phi}^H \boldsymbol{y}), 0)$
   while *not converged* **do**
      $\gamma_i \leftarrow \frac{|x_i|}{\sqrt{w_i}}, \forall i$
      $\boldsymbol{w} \leftarrow \text{diag}\left[\boldsymbol{\Phi}^H \left(\lambda \boldsymbol{I} + \boldsymbol{\Phi} diag(\boldsymbol{\gamma})\boldsymbol{\Phi}^H\right)^{-1} \boldsymbol{\Phi}\right]^{\frac{1}{2}}$
      while *not converged* **do**
         $\nabla g(\boldsymbol{x}) = \boldsymbol{\Phi}^H(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x})$
         $x_i^{real} \leftarrow \max\left[\text{real}\left(x_i - \frac{1}{L}\nabla g(\boldsymbol{x})\right) - \frac{\lambda w_i}{L}, 0\right],$
  $\forall i$
      end while
   end while
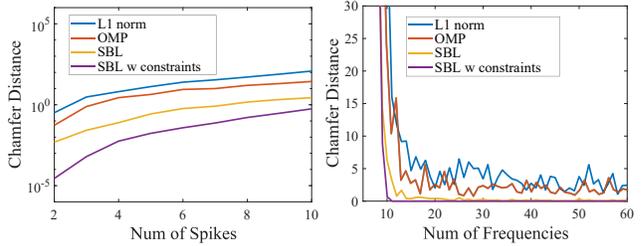
Figure 3. The reconstruction error plotted against (left) a varying number of impulses, N, in $\boldsymbol{x}$ and (right) a varying number of spatial frequencies, K.

and the estimated signal obtained from Eq. 7 using (1) the convex ℓ_1 norm; (2) orthogonal matching pursuit (OMP); (3) baseline SBL; and (4) our enhanced version of the SBL with constraints. The proposed SBL-based method outperforms the ℓ_1 norm and OMP at locating the spikes and shows a significantly lower chamfer error. In Fig. 3 (right), we plot the evolution of the reconstruction accuracy (evaluated by chamfer distance) versus different numbers of frequencies. The performances of these SBL-based methods look quite similar as the reconstruction error of the proposed method decreases dramatically and stays within a limited range when the number of scan frequencies is greater than 10 while OMP, which produces a local minimizer of the ℓ_0 cost, often gets stuck at bad local optima regardless of the number of scan frequencies. ℓ_1 fails because the convex ℓ_1 norm is not a good approximation of the ℓ_0 norm for this case.

Now, for a demonstration of multi-path separation, we use the real-world scenarios as shown in Fig. 4 and evaluate different processing approaches on them. The sinusoidal wavelength range is an integral multiple of 8 pixels on the projector with 480 rows from top to bottom. We shift 8 frames per frequency and capture 480 frames of video with
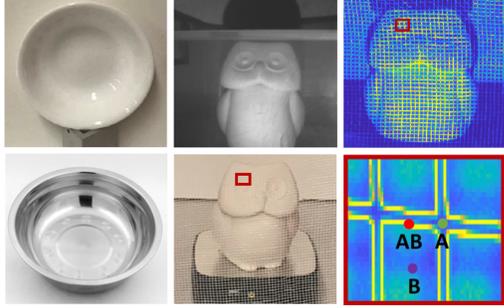
Figure 4. The experimental scenarios include (left) a white, porcelain bowl and a specular, stainless steel bowl; and (middle) a white plaster owl figurine sitting behind a semi-transparent plastic plate and behind a polyester mesh fabric [54]. The (right) close-up region shows the magnitude of the observation in the owl and mesh scene with foreground pixel **A**, background pixel **B**, and multipath pixel **AB**.

60 scan frequencies in total, namely $\{f_n = 480/(8n), n = 1, 2, \ldots, K, K = 60\}$, which are sorted from high to low. The proposed inexpensive and practical approach can simultaneously extract the phases and magnitudes of each light path without any hardware modifications or additional requirements for customized patterns. The first/strongest light path is selected to plot the point clouds/phases. The first interesting experimental task is to scan a white plaster owl figurine through a polyester mesh fabric with the same number of multi-frequency patterns. Some pixels on the image sensor will see through the holes of the mesh, while others will see both the foreground mesh and background owl. This represents an unusually difficult multi-path scenario because a large number of camera pixels receive light from multiple projector positions. Fig. 5 illustrates the point cloud reconstructions of the owl figurine, showing the effects of using the proposed methods with (top) 10 and (bottom) 11 frequencies. Fig. 5 (a) represents the traditional phase unwrapping with obviously devastating errors in the 3D reconstruction, while (b) and (c) show the current state-of-the-art bimodal multi-path processing [54] and OMP on sparse multi-path corrections with ghost-layers due to assuming an insufficient number of measured frequencies. It is evident that the reconstruction improves from left to right and top to bottom as the number of scan frequencies increases. The proposed sparse Bayesian learning with constraints, as shown in (d), outperforms all others.

As a further demonstration of the sparse multi-path approach, the point cloud reconstructions of an owl sitting behind a semi-transparent plastic plate are shown in the front, side, and top view in Fig. 6 with (left) traditional unwrapped phase reconstruction, (middle) bimodal multi-path reconstruction, and (right) the proposed sparse multi-path reconstruction. We note that both the bimodal and sparse models can decouple the light paths; however, in observing

the magnitude images from this owl and semi-transparent plate scan, the proposed model achieves much greater detail, as multiple tiny scratches are visible in Fig. 7 (right) compared to (left), which shows the magnitude image of the first/strongest light path from the prior bimodal model [54].

Figures 8 and 9 illustrate the different methods of multipath correction on the curved reflective surface of a white, porcelain bowl and a metal bowl. As shown in Fig. 8, the state-of-the-art bimodal multi-path procedure [54] is not completely immune to issues caused by specularities on the target surface; however, our proposed method greatly improves the artifacts introduced by multi-reflection. Reconstructing the metal bowl as shown in Fig. 9 is a particularly challenging task as both micro phase shifting [23] and bimodal multi-path processing [54] fail to reconstruct the surface, especially at the 12 and 6 o'clock positions. The reconstruction from micro phase shifting based on a scan under 7 frequencies contains large holes and errors due to the high-frequency specular interreflections. Compared to the dense multi-path approach of micro phase shifting, bimodal multi-path can, at least, reconstruct a secondary path, but it fails in the presence of three or more paths. In this case, the first/strongest light path may not always be the direct path on such a curved and highly reflective surface. Thus, we need to select the correct light path from a candidate set (decoupled light paths) based on a piece-wise planar assumption via discrete optimization [29]. Also of note is the fact that any SLI system needs to be calibrated over a range from some Z_{min} to some Z_{max}; we would, therefore, not want to reconstruct points outside this range. The evolution of the recomposed phase image versus an increasing number of decoupled light paths is shown in Fig. 9.

6. Limit on the Number of Separable Paths

Generally speaking, when searching for sparse solutions to an underdetermined linear system of equations (with more columns than rows), the maximum number of nonzeros we can expect to uniquely resolve will be less than the number of measurements/rows. However, if the rank of the measurement matrix turns out to be less than the number of rows, then the *effective* number of measurements will also be less. In this regard, we computed the singular value decomposition of $\mathbf{\Phi}$ given by Eq. 6 in order to examine how many actually independent measurements are available. We found that there are about 17 significant singular values of $\mathbf{\Phi}$, and hence no possible algorithm can reliably estimate more than 17 light paths.

The above analysis assumes that the sparse solutions are unconstrained; however, at least in principle, real and/or non-negativity constraints can potentially allow for the recovery of more than 17 nonzeros by eliminating some proportion of confounding solutions.
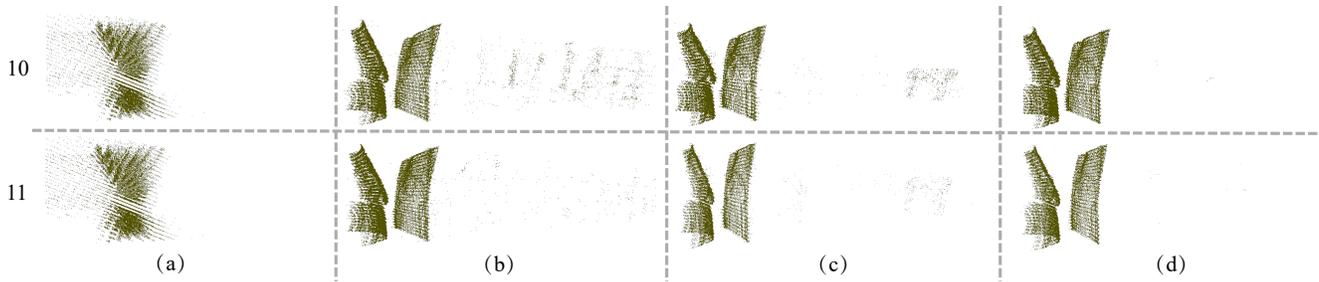
Figure 5. Point cloud reconstructions shown in side view of the owl figurine sitting behind the polyester mesh fabric using (top) 10 frequencies and (bottom) 11 frequencies with reconstructions acquired from (a) traditional phase unwrapping; (b) bimodal multi-path processing [54]; (c) OMP; (d) our proposed SBL with constraints.
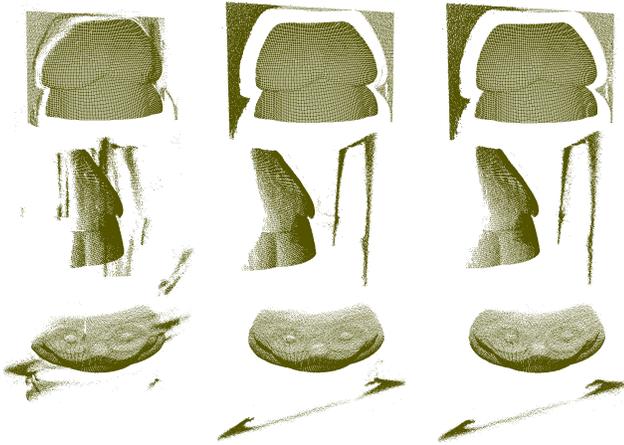
Figure 6. Illustration of the reconstructions of the owl figurine sitting behind a semi-transparent plastic plate using (left) traditional unwrapped phase, (middle) bimodal multi-path processed phase, and (right) the proposed sparse multi-path processed phase.
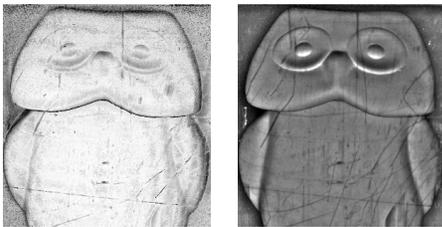
Figure 7. The extracted magnitude image of (left) the primary light path from the bimodal multi-path processing and (right) the first/strongest light path from the sparse multi-path processing.
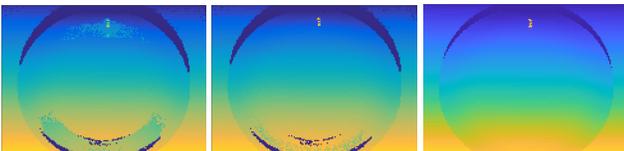
Figure 8. Reconstruction of a white, porcelain bowl. From left to right are traditional phase, bimodal multi-path processed phase, and the proposed SBL with constraints processed phase.
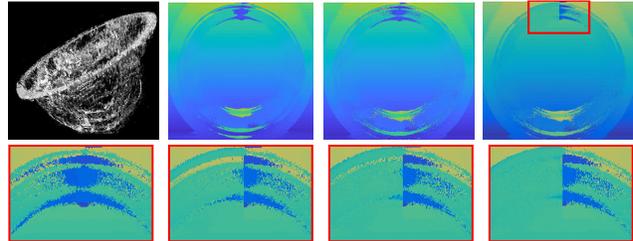
Figure 9. (Top) The reconstruction of a shiny, steel bowl showing: (left) MicroPS from [23], (center-left) bimodal multi-path from [54], (center-right) the first/strongest light path based on proposed SBL with constraints, and (right) the recomposed phase image in the left half of the red region of interest via the top thirteen decoupled light paths. (Bottom) Close-up regions of the top-right scan where the right side of each red box is the unprocessed phase, while the left side shows the evolution of recomposed phase image versus an increasing number of decoupled light paths with (left) the first two paths, (center-left) the first three paths, (center-right) the first five paths, and (right) the first thirteen paths.

7. Conclusion and Future Work

This paper has elucidated the problem of multi-path in the structured light method. The problem of mixing multiple rays in one pixel was formulated as an under-determined linear system. Furthermore, a mathematical method for estimating the contained light rays under the assumption of sparseness was described. The proposed modified version of the sparse Bayesian learning method to handle constraints is potentially useful for other problems as well. As we have demonstrated, the new technique outperforms the bimodal multi-path solution when dealing with sharp edges, inter-reflection, or semi-transparent objects. Moreover, it was shown that the number of separable paths depends on the frequency set. In the future, we may seek to obtain some ideas from [14] and try computing the SVD of $\mathbf{\Phi}$ constructed by different frequencies to investigate the optimal frequency setting for sparse multi-path corrections.

References

[1] Andrew D. Bagdanov, Alberto Del Bimbo, and Iacopo Masi. Florence faces: A dataset supporting 2d/3d face recognition. In *International Symposium on Communications, Control and Signal Processing*, pages 1–6, 2012. 2

[2] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1), 2009. 5, 6

[3] Kim L Boyer and Avinash C Kak. Color-encoded structured light for rapid active ranging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):14–28, 1987. 2

[4] David Burke, Abhijeet Ghosh, and Wolfgang Heidrich. Bidirectional importance sampling for direct illumination. *Rendering Techniques*, 5:147–156, 2005. 2

[5] Tongbo Chen, Hans-Peter Seidel, and Hendrik PA Lensch. Modulated phase-shifting for 3d scanning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2

[6] Naoya Chiba and Koichi Hashimoto. Ultra-fast multi-scale shape estimation of light transport matrix for complex light reflection objects. In *IEEE International Conference on Robotics and Automation*, pages 6147–6152, 2018. 2

[7] Sungjoon Choi, Qian-Yi Zhou, Stephen Miller, and Vladlen Koltun. A large dataset of object scans. *arXiv preprint arXiv:1602.02481*, 2016. 2

[8] Vincent Couture, Nicolas Martin, and Sebastien Roy. Unstructured light scanning to overcome interreflections. In *IEEE International Conference on Computer Vision*, pages 1895–1902, 2011. 2

[9] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2

[10] Eric Dedrick. Improving sli performance in optically challenging environments. PhD thesis, University of Kentucky, 2011. 2

[11] Arthur P Dempster, Donald B Rubin, and Robert K Tsutakawa. Estimation in covariance components models. *Journal of the American Statistical Association*, 76(374):341–353, 1981. 4

[12] David L Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 4

[13] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003. 4

[14] Candès Emmanuel, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006. 4, 8

[15] Haoqiang Fan, Hao Su, and Leonidas J. Guibas. A point set generation network for 3d object reconstruction from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2463–2471, 2017. 2, 6

[16] Daniel Freedman, Yoni Smolin, Eyal Krupka, Ido Leichter, and Mirko Schmidt. Sra: Fast removal of general multipath for tof sensors. In *European Conference on Computer Vision*, pages 234–249. Springer, 2014. 1

[17] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 2

[18] Jason Geng. Rainbow three-dimensional camera: new concept of high-speed three-dimensional vision systems. *Optical Engineering*, 35(2):376–383, 1996. 2

[19] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011. 2

[20] Mohit Gupta. Scene recovery and rendering techniques under global light transport. PhD thesis, Carnegie Mellon University, 2010. 2

[21] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. Structured light 3d scanning in the presence of global illumination. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 713–720, 2011. 2

[22] Mohit Gupta and Nikhil Nakhate. A geometric perspective on structured light coding. In *European Conference on Computer Vision*, pages 87–102, 2018. 2

[23] Mohit Gupta and Shree K Nayar. Micro phase shifting. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 813–820, 2012. 1, 2, 3, 7, 8

[24] Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics*, 34(5):156, 2015. 2

[25] Lei Huang, Mourad Idir, Chao Zuo, and Anand Asundi. Review of phase measuring deflectometry. *Optics and Lasers in Engineering*, 107:247–257. 2

[26] Henrik Wann Jensen, Stephen R Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Conference on Computer graphics and interactive techniques*, pages 511–518, 2001. 2

[27] Chufan Jiang, Beatrice Lim, and Song Zhang. Three-dimensional shape measurement using a structured light system with dual projectors. *Applied optics*, 57(14):3983–3990, 2018. 2

[28] Jorge Jimenez, Károly Zsolnai, Adrian Jarabo, Christian Freude, Thomas Auzinger, Xian-Chun Wu, Javier von der Pahlen, Michael Wimmer, and Diego Gutierrez. Separable subsurface scattering. *Computer Graphics Forum*, 34(6):188–197, 2015. 2

[29] Nikos Komodakis, Philip Torr, Vladimir Kolmogorov, and Yuri Boykov. Discrete optimization in computer vision. ICCV 2007 Tutorial. 7

[30] Jizhizi Li, Jing Zhang, Stephen J Maybank, and Dacheng Tao. End-to-end animal image matting. *arXiv preprint arXiv:2010.16188*, 2020. 1

[31] Jinlin Liu, Yuan Yao, Wendi Hou, Miaomiao Cui, Xuansong Xie, Changshui Zhang, and Xian-Sheng Hua. Boosting semantic human matting with coarse annotations. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2020. 1

[32] Kai Liu, Yongchang Wang, Daniel L Lau, Qi Hao, and Laurence G Hassebrook. Dual-frequency pattern scheme for high-speed 3-d shape measurement. *Optics Express*, 18(5):5229–5244, 2010. 2

[33] David JC MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992. 4

[34] Julio Marco, Quercus Hernandez, Adolfo Munoz, Yue Dong, Adrian Jarabo, Min H Kim, Xin Tong, and Diego Gutierrez. Deeptof: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. *ACM Transactions on Graphics*, 36(6):219, 2017. 1

[35] Kaichun Mo, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 909–918, 2019. 2

[36] Nikhil Naik, Achuta Kadambi, Christoph Rhemann, Shahram Izadi, Ramesh Raskar, and Sing Bing Kang. A light transport model for mitigating multipath interference in time-of-flight sensors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 73–81, 2015. 1

[37] Shree K Nayar, Gurunandan Krishnan, Michael D Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics*, 25(3):935–944, 2006. 2, 3

[38] Matthew O'Toole, Felix Heide, Lei Xiao, Matthias B Hullin, Wolfgang Heidrich, and Kiriakos N Kutulakos. Temporal frequency probing for 5d transient analysis of global light transport. *ACM Transactions on Graphics*, 33(4):87, 2014. 2

[39] Matthew O'Toole, John Mather, and Kiriakos N Kutulakos. 3d shape and indirect appearance by structured light transport. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3246–3253, 2014. 2

[40] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2

[41] Charles R. Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J. Guibas. Frustum pointnets for 3d object detection from RGB-D data. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 918–927, 2018. 2

[42] Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 77–85, 2017. 2

[43] Tobias Ritschel, Thorsten Grosch, and Hans-Peter Seidel. Approximating dynamic global illumination in image space. In *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games*, pages 75–82, 2009. 2

[44] Vishwanath Saragadam, Jian Wang, Mohit Gupta, and Shree Nayar. Micro-baseline structured light. In *IEEE International Conference on Computer Vision*, pages 4049–4058, 2019. 2

[45] Steven M Seitz, Yasuyuki Matsushita, and Kiriakos N Kutulakos. A theory of inverse light transport. In *IEEE International Conference on Computer Vision*, pages 1440–1447, 2005. 2

[46] Bharath K Sriperumbudur and Gert RG Lanckriet. A proof of convergence of the concave-convex procedure using zangwill's theory. *Neural computation*, 24(6):1391–1407, 2012. 5

[47] Jingwei Tang, Yagiz Aksoy, Cengiz Oztireli, Markus Gross, and Tunc Ozan Aydin. Learning-based sampling for natural image matting. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2019. 1

[48] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 1:211–244, 2001. 4

[49] Piet Vuylsteke and André Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):148–164, 1990. 2

[50] Yongchang Wang, Kai Liu, Qi Hao, Xianwang Wang, and Laurence G Hassebrook. Robust active stereo vision using kullback-leibler divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):548–563, 2012. 2

[51] David Wipf, Bhaskar Rao, and Srikantan Nagarajan. Latent variable bayesian models for promoting sparsity. *IEEE Transactions on Information Theory*, 57(9):6236–6255, 2011. 5

[52] Bo Xin, Yizhou Wang, Wen Gao, and David Wipf. Building invariances into sparse subspace clustering. *IEEE Transactions on Signal Processing*, 66(2):449–462, 2017. 5

[53] Ying Yu, Daniel L. Lau, Matthew P. Ruffner, and Kai Liu. Dual-projector structured light 3d shape measurement. *Applied Optics*, 59(4):964–974, 2020. 1, 2

[54] Yu Zhang, Daniel L. Lau, and Ying Yu. Causes and corrections for bimodal multi-path scanning with structured light. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4431–4439, 2019. 1, 2, 3, 7, 8

[55] Yu Zhang, Mao Ye, Dinesh Manocha, and Ruigang Yang. 3d reconstruction in the presence of glass and mirrors by acoustic and visual fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1785–1798, 2018. 1

[56] Zonghua Zhang. Review of single-shot 3d shape measurement by phase calculation-based fringe projection techniques. *Optics and Lasers in Engineering*, 50(8):1097–1106, 2012. 2

[57] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4490–4499, 2018. 2

[58] Chao Zuo, Lei Huang, Minliang Zhang, Qian Chen, and Anand Asundi. Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review. *Optics and Lasers in Engineering*, 85:84–103. 3