

CLAMP: Majorized Plug-and-Play for Coherent 3D Lidar Imaging

Tony G. Allen, David J. Rabb,
Gregory T. Buzzard, *Senior Member, IEEE*, Charles A. Bouman, *Fellow, IEEE*.

Abstract—Coherent lidar uses a chirped laser pulse for 3D imaging of distant targets. However, existing coherent lidar image reconstruction methods do not account for the system’s aperture, resulting in sub-optimal resolution. Moreover, these methods use majorization-minimization for computational efficiency, but do so without a theoretical treatment of convergence.

In this paper, we present Coherent Lidar Aperture Modeled Plug-and-Play (CLAMP) for multi-look coherent lidar image reconstruction. CLAMP uses multi-agent consensus equilibrium (a form of PnP) to combine a neural network denoiser with an accurate surrogate forward model of coherent lidar. Additionally, CLAMP introduces a computationally efficient FFT-based method to account for the system’s aperture to improve resolution of reconstructed images. Furthermore, we formalize the use of majorization-minimization in consensus optimization problems and prove convergence to the exact consensus equilibrium solution. Finally, we apply CLAMP to synthetic and measured data to demonstrate its effectiveness in producing high-resolution, speckle-free, 3D imagery.

Index Terms—lidar, coherent imaging, majorization-minimization, plug-and-play, consensus equilibrium, deep neural networks.

I. INTRODUCTION

COHERENT lidar is a method for three-dimensional (3D) imaging in which a target is flood-illuminated by a frequency-modulated laser source. In multi-look lidar, multiple snapshots of the reflected wave are recorded on a 2D detector array by a spatial-heterodyne interferometric system [1]. This process yields a collection of 2D measurements over the duration of the chirped waveform. A 3D complex-valued volume can be recovered by a demodulation and filtering process [2]–[4]. This volume can be further processed to produce a 3D image of the surfaces in the scene.

There are several challenges in creating high-quality 3D images from multi-look lidar data. Recent efforts have focused primarily on correcting phase aberrations caused by non-concurrent measurements [5], object motion [6], or atmospheric turbulence [7]–[10]. However, beyond addressing

phase aberrations, reconstruction algorithms also need to reduce speckle, noise, and aperture-induced blur.

The most straightforward method to reconstruct multi-look coherent lidar images, often called speckle averaging, involves averaging the squared magnitude of back-projected, demodulated data from multiple looks. Assuming the speckle decorrelates between looks, this averaging reduces speckle in the final reconstruction. However, this method does not incorporate regularization or priors.

In complex-valued image reconstruction problems, researchers generally have used model-based iterative reconstruction (MBIR) methods that regularize the complex-valued reflectance of the target, assuming the scene reflectance to have locally correlated phase and magnitude [11], [12]. However, in many scenarios, the phase of the reflectance is well-modeled as random, without spatial or temporal correlation [13]–[15]. This observation can help simplify the estimation problem by reducing its effective dimensionality.

In a related 3D near-field radar imaging problem, the authors of [16] leverage this observation to regularize only the magnitude of the reflectance, leaving the phase unregularized. In contrast, the papers of Pellizzari et al. [17]–[21] introduce an approach for 2D coherent lidar that uses a Bayesian framework to directly estimate the speckle-free, real-valued reflectivity of the target. The real-valued reflectivity roughly corresponds to the image seen using broad-spectrum illumination; since the reflectivity is relatively smooth, it can be more strongly regularized to further reduce the effective dimensionality of the problem. While estimating the reflectivity introduces a seemingly intractable, nonlinear relationship between measured data and the estimated image, Pellizzari et al. overcomes this problem by using the Expectation-Maximization (EM) algorithm [22]–[24].

A primary benefit of these methods is that they enable the use of advanced real-valued image priors through Plug-and-Play (PnP) algorithms [25]–[27]. Despite the success of advanced priors in 2D image reconstruction [28], relatively little work has been done using 3D priors in lidar image reconstruction or related modalities. It is important to use 3D priors instead, as the images involve complex 3D structures and spatial dependencies that cannot be adequately captured by 2D models. Early research made a simplifying assumption that there is only a single surface depth at each 2D pixel [29], [30], while more recent works used Total Variation (TV) [31] and mixed-norm ($\ell_{2,1}$) regularization [32], deep point cloud denoiser [33], or a 3D U-Net denoiser [16] to jointly exploit range and cross-range correlations and account for multiple

Tony G. Allen is with Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37830, USA, but this work was completed while at the Air Force Research Laboratory, WPAFB, OH 45433, USA, and Purdue University, West Lafayette, IN 47907, USA (email: allentg@ornl.gov).

David J. Rabb is with the Sensors Directorate, Air Force Research Laboratory, WPAFB, OH 45433, USA (email: david.rabb@us.af.mil).

Gregory T. Buzzard is with the Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA (email: buzzard@purdue.edu) and was partially supported by NSF CCF-1763896.

Charles A. Bouman is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA (email: bouman@purdue.edu) and was partially supported by the Showalter Trust.

(Corresponding author: Tony G. Allen)

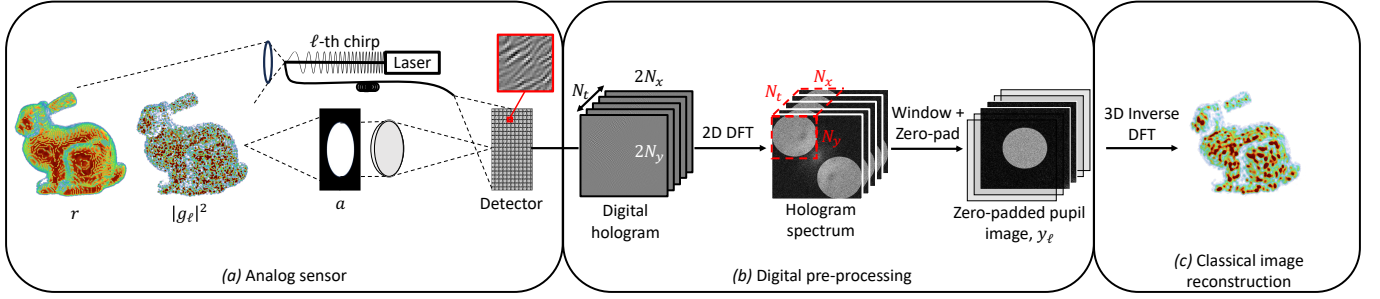


Fig. 1: The measurement process and conventional data processing pipeline for a 3D coherent lidar imaging system using image plane recording geometry. (a) Multiple chirped laser pulses, or looks, illuminate a target. Each look is assumed to generate an independent speckle pattern, g_ℓ . Reflected light propagates to an optical system with aperture, a , where a lens focuses an image onto a focal plane array. The image is interfered with a delayed local reference beam to form a hologram, which is sampled in time to create N_t frames of size $2N_x \times 2N_y$ for each look. (b) The 2D DFT of each hologram frame is taken to separate the frequency components of the target and reference beam. The pupil images are then windowed to size $N_x \times N_y \times N_t$, centered, and zero-padded by a factor, q , to form a 3D stack of complex fields, y_ℓ for the ℓ^{th} look. (c) A noisy 3D image of the target can be reconstructed from y_ℓ by the inverse 3D DFT and averaging over $\ell = 1, \dots, L$.

surface depths at each cross-range pixel. However, these methods have not been applied to 3D coherent lidar imaging.

Another challenge in lidar reconstruction is to efficiently correct the blur caused by the finite extent of the system's aperture. The aforementioned MBIR methods [17]–[21] make a simplifying assumption that the forward model operator is an orthogonal matrix. This greatly reduces computational complexity, but effectively disregards the system aperture. On a similar problem in polarimetric synthetic-aperture radar, [34] uses the same EM algorithm approach, but addresses the aperture problem by using a greedy graph-coloring-based matrix probing method to do an approximate matrix inversion.

Additionally, from a theoretical point of view, a variety of lidar reconstruction algorithms combine the EM algorithm with PnP or ADMM [35]; however, convergence results for this type of combined algorithm are limited. In [36], the authors prove an ergodic convergence rate when using general majorization-minimization [37]–[39] principles within ADMM, but they do not prove non-ergodic convergence of the iterates.

In this paper, we present Coherent Lidar Aperture Modeled Plug-and-Play (CLAMP), a majorized PnP algorithm for multi-look coherent lidar imaging that accounts for the imaging system's aperture, which builds on our previous publication [40]. Using the Multi-Agent Consensus Equilibrium (MACE) framework [41], CLAMP combines multi-look coherent lidar data with a deep 3D image prior model to produce a 3D reconstruction with reduced noise, speckle, and increased resolution relative to traditional reconstruction methods. Specifically, we make the following novel contributions:

- Accurate lidar aperture modeling using a computationally efficient FFT-based method;
- Incorporation of a deep neural network prior in 3D multi-look coherent lidar reconstruction;
- Formulation of a unified framework for majorization-minimization within MACE with theoretical guarantees of convergence to an exact MACE solution;
- Results on simulated and measured data that show CLAMP yields 3D images with reduced speckle and improved resolution.

A key component of CLAMP is the use of majorization-minimization, or surrogate function optimization, in the MACE framework. We use convex surrogate functions in place of the proximal agents in MACE and show that this method converges to the exact MACE solution. This surrogate-based MACE applies to many imaging problems and modalities and encompasses several common approximation methods and existing algorithms as special cases.

II. BACKGROUND

In this section, we provide an overview of the multi-look coherent lidar imaging problem, develop the models used in our approach, and introduce the MACE framework. We strive to provide a comprehensive background on the problem and the methods used in our approach, but for a more detailed discussion of 3D coherent lidar imaging, we refer the reader to [2]–[4], [7].

A. Coherent Lidar Imaging Measurement System

Figure 1(a) illustrates a typical coherent lidar imaging system with an image-plane recording geometry. In this system, a series of L chirped laser pulses, or looks, illuminate a target with reflectivity, r . Each look generates a speckle pattern, g_ℓ , which is a complex-valued field that represents the target's reflectance. Throughout this work, we assume that the target is stationary and that the speckle realizations and measurement noise samples are statistically independent from look to look. For each look, $\ell = 1, \dots, L$, the corresponding complex reflectance, g_ℓ propagates to the aperture, a , where a thin lens forms an image of the target on a focal plane array. The resulting image is mixed with a delayed, off-axis, local reference beam, which results in an interference pattern, or hologram, on the focal plane array.

As illustrated in Figure 1(b), the hologram is sampled in time to create a series of N_t frames of size $2N_x \times 2N_y$ for each look. Intuitively, each of these hologram frames corresponds to a slightly different laser wavelength due to the chirp slope. This stack of 2D holograms will can be used to recover 3D image of the complex reflectance of the target. For information

on the principles of 3D holographic imaging, we refer the reader to [2]–[4].

Figure 1(b) and (c) illustrate the conventional processing steps to reconstruct a 3D image of the target from this stack of N_t holograms. First, the 2D discrete Fourier transform (DFT) is taken for each frame in a look. Since the frames are real valued, each DFT results in a complex image that contains two circular disks corresponding to the complex conjugate circular apertures in the pupil plane. One of the two circular apertures is then embedded in a bounding square array (shown with a red dotted line) and windowed to give the pupil image. This image is zero-padded to increase the effective sampling rate in the space and time domains and yields a larger 3D stack of pupil images, y_ℓ . Specifically, we zero-pad the pupil image by a factor of q in each of the three dimensions; so the final array, y_ℓ , is q^3 times the size of the un-padded images. The last step in conventional reconstruction is to take the inverse 3D DFT of each y_ℓ and average over ℓ , which yields a candidate reconstruction in cross-range and depth.

The final step in conventional reconstruction limits reconstruction quality in two key ways: lost spatial and temporal resolution due to the missing information in the frequency domain, and noise due to speckle.

In contrast, our algorithm replaces the conventional final 3D DFT and averaging step with an iterative model-based strategy which employs a physics-based forward model in conjunction with neural network denoisers to ensure the reconstructed image is high-resolution and free of speckle and noise.

B. Forward Model

Assuming the target's depth is small compared to the range of the lidar system and that the fractional bandwidth of the chirp is small [7], the complex field in the pupil plane, $\tilde{y}_\ell : \mathbb{R}^3 \rightarrow \mathbb{C}$, for $\ell = 1, \dots, L$, can be determined by the Fresnel propagation integral [42] of the complex reflectance, $\tilde{g}_\ell : \mathbb{R}^3 \rightarrow \mathbb{C}$ as

$$\begin{aligned} \tilde{y}_\ell(\xi, \nu, t) &= \frac{e^{2\pi i d/\lambda(t)}}{i\lambda d} \tilde{a}(\xi, \nu, t) \exp \left\{ \frac{2\pi i}{2\lambda(t)d} (\xi^2 + \nu^2) \right\} \\ &\quad \iint \int_{-\infty}^{\infty} \left\{ \tilde{g}(x, y, z) \exp \left\{ \frac{2\pi i}{2\lambda d} (x^2 + y^2) \right\} \right\} \\ &\quad \exp \left\{ -\frac{2\pi i}{\lambda d} \left(\xi x + \nu y + \frac{1}{\lambda(t)} z \right) \right\} dx dy dz, \end{aligned}$$

where (ξ, ν) are the coordinates in the pupil plane, (x, y, z) are the coordinates of the target, d_z is the distance from the object plane to the pupil plane, $\tilde{a}(\xi, \nu, t)$ is the aperture, and $\lambda(t)$ is the instantaneous wavelength of the chirp at time t . By a small angle approximation, the quadratic phase term outside the integral can be ignored, and since we only consider the magnitude of the complex field, the quadratic phase term inside the integral can be absorbed into \tilde{g}_ℓ as in [18]. This results in a Fourier relationship,

$$\tilde{y}_\ell(\xi, \nu, t) = \frac{e^{2\pi i d/\lambda(t)}}{i\lambda d} \tilde{a}(\xi, \nu, t) \cdot \mathcal{F}[\tilde{g}_\ell] \left(\frac{\xi}{\lambda d}, \frac{\nu}{\lambda d}, \frac{1}{\lambda(t)} \right), \quad (1)$$

where \mathcal{F} is the continuous 3D Fourier transform.

The measured and processed data, $y_\ell \in \mathbb{C}^n$ where $n = q^3 N_x N_y N_t$, is a normalized, discrete sampling of \tilde{y}_ℓ in lexicographical order. A discrete version of (1) can be used to relate y_ℓ to the complex reflectance image, $g_\ell \in \mathbb{C}^n$, by

$$y_\ell = A g_\ell + \eta_\ell, \quad (2)$$

where A is a linear operator that models the imaging system, and $\eta_\ell \sim \text{CN}(0, \sigma_\eta^2 I)$ is circularly-symmetric complex Gaussian white noise, which is the dominant noise source [43]. The operator A can be written as

$$A = \mathcal{D}(a)F,$$

where \mathcal{F} is the orthonormal 3D DFT, $\mathcal{D}(\cdot)$ represents a diagonal matrix with entries given by its argument, and the vector a is a binary vector that encodes the aperture of the imaging system after zero-padding.

Our ultimate goal will be to recover $r \in \mathbb{R}^n$, the real-valued speckle-free reflectivity at each 3D voxel. While the data y_ℓ directly relates to g_ℓ by (2), images reconstructed from estimates of g_ℓ are degraded by speckle. To address this, we need a forward model that relates r to g_ℓ .

We will assume that each look produces a sample of g_ℓ that is conditionally independent given r . Using a standard model of fully developed speckle [13], [44], we can model the conditional distribution of g_ℓ given r as being a circularly-symmetric complex Gaussian random variable with conditional distribution

$$g_\ell | r \sim \text{CN}(0, \mathcal{D}(r)), \quad (3)$$

where $\mathcal{D}(r)$ is a diagonal covariance with entries r . We can then directly model the data y_ℓ in terms of r by composing the forward model (2) with the speckle model of (3) to yield

$$y_\ell | r \sim \text{CN}(0, A \mathcal{D}(r) A^H + \sigma_\eta^2 I), \quad (4)$$

where y_ℓ for $\ell = 1, \dots, L$ are assumed conditionally independent given r .

III. CLAMP

In this section, we present the Coherent Lidar Aperture Modeled Plug-and-Play (CLAMP) algorithm for reconstructing 3D images from multi-look coherent lidar measurements.

A. MACE Formulation

CLAMP is built on the MACE framework. For a more detailed discussion on MACE, we refer the reader to [41].

Our goal is to reconstruct r from a collection of L coherent lidar looks, $\{y_\ell\}_{\ell=1}^L$. In order to do this, MACE balances multiple agents, such as data-fitting operators or denoisers, to form a single reconstruction [41]. In our case, we will use L data-fitting agents, F_ℓ , and a single denoising agent, H . Each agent operates on an image, w_j , and returns a new image that is improved according to the directive of the agent.

The first L agents are forward agents, F_ℓ , that enforce fidelity to the measurements, y_ℓ , for $\ell = 1, \dots, L$. Specifically, F_ℓ is a proximal map given by

$$F_\ell(w_\ell) = \arg\min_r \left\{ f_\ell(r) + \frac{1}{2\sigma^2} \|r - w_\ell\|^2 \right\}.$$

where w_ℓ is a candidate reconstruction, and $f_\ell(r) = -\log p(y_\ell|r)$ is the negative log likelihood associated with the ℓ^{th} look. These operators have an intuitive interpretation of taking a candidate reconstruction as input and returning an image that better fits the data from look ℓ . The final agent is a deep neural network image denoiser that enforces a prior on the 3D image by reducing noise and speckle. The consensus equilibrium of these $L+1$ agents yields an image that is both consistent with all measurements and regularized in all three dimensions. The implementation of these agents is described in Section III-B and III-C.

To formulate the MACE solution, we stack the images into a single state vector, $\mathbf{w} = [w_1, \dots, w_{L+1}]$, and stack the agents into a single operator, \mathbf{F} , given by

$$\mathbf{F}(\mathbf{w}) = [F_1(w_1), \dots, F_L(w_L), H(w_{L+1})].$$

In order to produce a single reconstruction that is a compromise between the agents, we average the components of \mathbf{w} to form a new image, $\bar{\mathbf{w}} = \frac{1}{L+1} \sum_{j=1}^{L+1} v_j w_j$ where v_j sums to one. In this work, we will use $v_j = 1/2L$ for all $j = 1, \dots, L$ and $v_{L+1} = 1/2$, which gives equal weight to the data-fitting agents and the prior agent. Finally, the MACE solution is determined by the solution to the equilibrium equation

$$\mathbf{F}(\mathbf{w}^*) = \mathbf{G}(\mathbf{w}^*), \quad (5)$$

where $\mathbf{G}(\mathbf{w}) = [\bar{\mathbf{w}}, \dots, \bar{\mathbf{w}}]$ is a weighted average of the components of \mathbf{w} . Intuitively, \mathbf{w}^* is the image that is the consensus solution of the agents in \mathbf{F} . The final reconstruction, r^* , can be obtained from any single component of \mathbf{w}^* , however, we typically use $r^* = \bar{\mathbf{w}}^*$ for numerical robustness.

The MACE equation of (5) is commonly solved as a fixed point of the operator $\mathbf{T} = (2\mathbf{G} - \mathbf{I})(2\mathbf{F} - \mathbf{I})$ by the iterations

$$\mathbf{w} \leftarrow (1 - \rho)\mathbf{w} + \rho\mathbf{T}\mathbf{w}, \quad (6)$$

where $\rho \in (0, 1)$. This approach (summarized in Algorithm 1) is guaranteed to converge to a solution of (5) when \mathbf{T} is non-expansive.

Algorithm 1 MACE

- 1: **Input:** Initialize $\mathbf{w} \in \mathbb{R}^{n(L+3)}$, $\rho \in (0, 1)$
 - 2: **while** not converged **do**
 - 3: $\mathbf{r} \leftarrow \mathbf{F}(\mathbf{w})$
 - 4: $\mathbf{x} \leftarrow 2\mathbf{r} - \mathbf{w}$
 - 5: $\mathbf{w} \leftarrow \mathbf{w} + 2\rho(\mathbf{G}(\mathbf{x}) - \mathbf{r})$
 - 6: **end while**
 - 7: **Output:** $r^* = \bar{\mathbf{w}}$
-

B. EM Surrogate Forward Agents

Unfortunately, direct application of the MACE iterations in (6) is computationally infeasible. The proximal map of the forward model in (4) quickly becomes intractable for even moderately large images due to the non-linear relationship between r and y_ℓ introduced by speckle interference.

In order to overcome this problem, we will take the approach first proposed by Pellizzari et al. [17]–[21] and use

the EM algorithm to compute surrogate functions to the exact negative log likelihood functions. These surrogate functions are given by

$$\hat{f}_\ell(r; r'_\ell) = \mathbb{E}_{g|y_\ell, r'_\ell} [-\log p(y_\ell, g|r)], \quad (7)$$

where the expectation is taken over the conditional distribution of the latent complex reflectance g given y_ℓ and r'_ℓ , an approximation of r . Assuming (2) and (3), and using $y_\ell|g, r = y_\ell|g$, [17] shows that this conditional distribution of g given y_ℓ, r'_ℓ is Gaussian with the form

$$p(g|y_\ell, r'_\ell) = \frac{1}{Z} \exp \left\{ -(g - \mu_\ell)^H \tilde{C}_\ell^{-1} (g - \mu_\ell) \right\},$$

where Z is a normalizing constant and μ_ℓ and \tilde{C}_ℓ are the mean and covariance given by

$$\mu_\ell = \operatorname{argmin}_{g \in \mathbb{C}^n} \left\{ \frac{1}{2\sigma_w^2} \|y_\ell - Ag\|^2 + \frac{1}{2} g^H \mathcal{D} \left(\frac{1}{r'_\ell} \right) g \right\}, \quad (8)$$

$$\tilde{C}_\ell = \left[\frac{1}{\sigma_w^2} A^H A + \mathcal{D} \left(\frac{1}{r'_\ell} \right) \right]^{-1}. \quad (9)$$

Computing the expectation in (7) then gives the surrogate function as

$$\hat{f}_\ell(r; \mu_\ell, \tilde{C}_\ell) = \sum_{j=1}^n \left\{ \log(r_j) + \frac{|\mu_{\ell,j}|^2 + \tilde{c}_{\ell,j}}{r_j} \right\}, \quad (10)$$

where j is the voxel index, $\mu_{\ell,j}$ is the j th element of μ_ℓ , and $\tilde{c}_{\ell,j}$ is the j th diagonal element of \tilde{C}_ℓ .

However, using this surrogate function is also intractable for large images since computing the covariance, \tilde{C}_ℓ , requires a large matrix inversion, while computing the mean, μ_ℓ , of (8) directly requires the solution of a large quadratic optimization problem. We address each of these issues separately.

First, to improve the tractability of computing \tilde{C}_ℓ , we make the approximation that $A^H A = \alpha I$, where $\alpha = \|a\|_1/n$ has the interpretation of the fraction of light passing through the aperture. In this case, \tilde{C}_ℓ is replaced by diagonal matrix C_ℓ with diagonal entries $c_{\ell,j}$, given by

$$c_{\ell,j} = \frac{\sigma_w^2 r'_{\ell,j}}{\alpha r_{\ell,j} + \sigma_w^2}.$$

Alternatively, if a more accurate approximation is needed, one could use the method used in [34] at the cost of more computation. Importantly, we only make this orthogonality assumption in approximating \tilde{C}_ℓ , not throughout the entire forward model.

Next, to evaluate (8) efficiently, we solve the minimization problem iteratively. Rather than fully solving it at each CLAMP iteration, we compute only a single gradient descent step. As CLAMP progresses, these incremental steps gradually minimize (8) over multiple iterations.

To increase stability and avoid division by zero in the objective in (8), we add the noise floor σ_w^2/α to r'_ℓ . This yields the regularized objective

$$h(g; y_\ell, r'_\ell) = \frac{1}{2\sigma_w^2} \|y_\ell - Ag\|^2 + \frac{1}{2} g^H \mathcal{D} \left(\frac{1}{r'_\ell + \sigma_w^2/\alpha} \right) g.$$

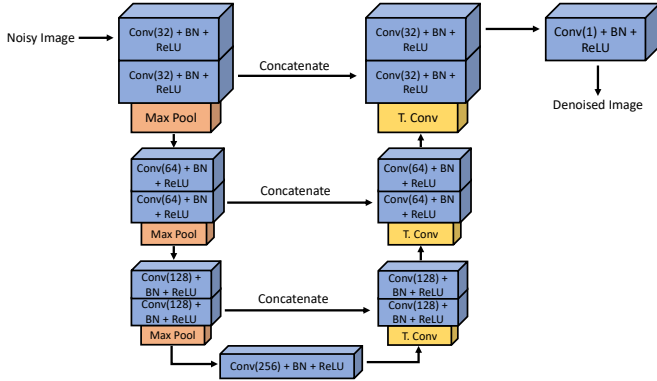


Fig. 2: The architecture of the 3D UNet used as a prior agent. The number of output channels for the convolutional layers is shown in parentheses.

The gradient step is then given by

$$\begin{aligned} d &\leftarrow -\nabla h(\mu_\ell; y_\ell, r'_\ell) \\ \mu_\ell &\leftarrow \mu_\ell + \gamma_* d, \end{aligned} \quad (11)$$

where γ_* is the optimal step-size and is given by

$$\gamma_* = \frac{(Ad)^H (y_\ell - A\mu_\ell) - \sigma_w^2 d^H \mathcal{D}(1/r'_\ell) \mu_\ell}{d^H (A^H Ad + \sigma_w^2 \mathcal{D}(1/r'_\ell) d)}.$$

Finally, using these forms for $\mu_{\ell,j}$ and $c_{\ell,j}$, the EM surrogate of (7) can be calculated [17] to be

$$\hat{f}_\ell(r; \mu_\ell, c_\ell) = \sum_{j=1}^n \left\{ \log r_j + \frac{|\mu_{\ell,j}|^2 + c_{\ell,j}}{r_j} \right\}. \quad (12)$$

From this, the CLAMP forward agent is the proximal map

$$\hat{F}_\ell(v; \mu_\ell, c_\ell) = \operatorname{argmin}_r \left\{ \hat{f}_\ell(r; \mu_\ell, c_\ell) + \frac{1}{2\sigma^2} \|r - v\|^2 \right\}, \quad (13)$$

which can be computed in closed form by solving for a root of a cubic equation for each voxel in the image. However, in practice, to further simplify computation and speed up implementation, we employ an additional quadratic surrogate of \hat{f}_ℓ . The details of which are given in Appendix A.

C. 3D Deep Prior Agent

In order to regularize the image, we use a 3D U-Net image denoiser similar to the one used in [16] for 3D radar imaging. This model was chosen because it was shown to be effective in jointly exploiting correlations in cross-range and range dimensions.

The architecture of the 3D U-Net is shown in Figure 2. It consists of three encoding blocks, each with two convolutional layers, followed by three decoding blocks, each with two convolutional layers. Each convolutional block is followed by a batch normalization (BN) layer and a ReLU activation function and a max pooling to reduce the dimensions of feature tensors by a factor of two, whereas the decoding stages employ a transposed convolution to upsample the feature tensor by a factor of two. A final 3D convolutional block is used to produce the denoised image.

The network was trained to denoise 3D images from ShapeNet [45] with 10% additive white Gaussian noise. The training set consisted of 800 models rendered on a $256 \times 256 \times 256$ grid, and an additional 200 models were used for validation. The images were patched into $32 \times 32 \times 32$ patches in order to fit into the GPU memory. We used the Adam optimizer with an exponentially decaying learning rate and the MSE loss function. The initial learning rate was set to 10^{-2} and decayed by a factor of 0.975 every 300 training batches. The batch size was set to 24 and the network was trained for 100 epochs. The network weights were saved when the validation loss was lowest during training.

Algorithm 2 CLAMP

```

1: Input:  $y_\ell \in \mathbb{C}^n$  for  $\ell = 1, \dots, L$ 
2: Initialize:  $\mu_\ell = A^H y_\ell$ ,  $r_\ell = \frac{1}{L} \sum_{\ell=1}^L |A^H y_\ell|^2$ ,  $w_\ell = r_\ell$ ,
   for  $\ell = 1, \dots, L$ 
3: while not converged do
4:   for  $\ell = 1, \dots, L$  do
5:      $\forall j, c_{\ell,j} \leftarrow \frac{\sigma_w^2 r_{\ell,j}}{\alpha r_{\ell,j} + \sigma_w^2}$  ▷ Update  $C_\ell$ 
6:      $d \leftarrow -\nabla h(\mu_\ell; y_\ell, r_\ell)$  ▷ Update  $\mu_\ell$ 
7:      $\gamma \leftarrow \operatorname{argmin}_\gamma \{h(\mu_\ell - \gamma d; y_\ell, r_\ell)\}$ 
8:      $\mu_\ell \leftarrow \mu_\ell + \gamma d$ 
9:      $r_\ell \leftarrow \hat{F}_\ell(w_\ell; \mu_\ell, c_\ell)$  ▷ Update  $r_\ell$  by (13)
10:   end for
11:    $r_{L+1} \leftarrow \mathbf{H}(w_{L+1})$  ▷ Apply prior agent
12:    $\mathbf{r} \leftarrow [r_1, \dots, r_{L+3}]$  ▷ Do MACE iteration
13:    $\mathbf{x} \leftarrow 2\mathbf{r} - \mathbf{w}$ 
14:    $\mathbf{w} \leftarrow \mathbf{w} + 2\rho(\mathbf{G}(\mathbf{x}) - \mathbf{r})$ 
15: end while
16: Output:  $r^* = \bar{\mathbf{w}}$ 

```

D. Summary of CLAMP

A summary of CLAMP is given in Algorithm 2. CLAMP begins with an initialization of the image. In steps 4–9, the each forward agent is approximated by the EM surrogate, and the image is updated using the proximal map of the surrogate function to enforce fidelity to the measurements, y_ℓ . In step 11, the deep prior agent are applied to the image to regularize the image in all three dimensions. Finally, in steps 12–14, a consensus among the agents is enforced, resulting in an image that is both consistent with the measurements and regularized in all three dimensions.

A key benefit of CLAMP is its computational efficiency compared to a direct application of MACE. CLAMP achieves this efficiency by using EM surrogates, given in (10), in replacement of the exact negative log likelihood functions in the forward agents. However, these surrogate approximations alone are not enough to make the algorithm computationally feasible. The computational complexity of MACE with EM surrogates is dominated by the computation of the diagonal of the covariance matrix, \hat{C}_ℓ , in (9). Solving for this using direct methods has a complexity of $O(n^3)$ [46], which is generally infeasible. Similarly, solving the optimization problem in (8) requires the same $O(n^3)$ operation to compute μ_ℓ exactly.

In contrast, CLAMP uses the diagonal approximation of the covariance matrix, C_ℓ , and an iterative gradient step to compute the mean, μ_ℓ . These steps, steps 5–8 in Algorithm 2, are dominated by the computation of the gradient in (11) for each of the L looks. Since this gradient can be computed efficiently using the FFT, the total complexity of each iteration of the algorithm is $O(n \log(n))$ [47]. This makes CLAMP computationally feasible for large images. To give insight into the amount of computation required, the experiments in Section V were run on a 3D image of size 128^3 on a single NVIDIA Quadro RTX 8000 GPU, and each iteration took less than 0.5 seconds.

IV. MAJORIZED MACE THEORY

A crucial aspect of CLAMP is the use of a surrogate function through the EM algorithm. In this section, we present a generalized theory of majorization-minimization within the MACE framework with guaranteed convergence to an exact MACE solution as defined in (5).

Although our theory is limited to the case where all agents are proximal maps, it applies in practice to many widely used denoisers. Many PnP algorithms (including CLAMP) use denoising agents that are not necessarily proximal maps. However, typical neural network denoisers are trained to be minimum mean squared error (MMSE) denoisers, which were shown to be proximal maps in [48]. This equivalence was also demonstrated experimentally in [49], in which the authors show that the convergence of a PnP algorithm with an exact MMSE denoiser agrees remarkably well with the convergence of the same algorithm with a neural network denoiser. This highlights the practical relevance of our theory.

In our setting, we modify the operator F_i by approximating its objective function, f_i , with a surrogate function, \hat{f}_i , whose proximal map, \hat{F}_i , can be computed more efficiently. We make a few standard assumptions on the surrogate functions, given in Definition 1, which are adapted from those made in [38], [50], [51]. Commonly used surrogates, such as the Jensen surrogate used in the EM algorithm or quadratic approximations of twice differentiable functions meet these requirements.

Definition 1 (Surrogate). *Let $\Theta \subset \mathbb{R}^n$ be convex, and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. A function $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{R}$ is a **surrogate** of f near $\xi \in \Theta$ when the following conditions hold:*

- **Majorization**: we have $\hat{f}(x) \geq f(x)$ for all $x \in \mathbb{R}^n$.
- **Smoothness**: the approximation error $e \triangleq \hat{f} - f$ is differentiable, and its gradient is L -Lipschitz continuous. Moreover, we have that $e(\xi) = 0$ and $\nabla e(\xi) = 0$,
- **Strong convexity**: \hat{f} is p -strongly convex with $p \geq L$.

When the point ξ is relevant, we write the surrogate as $\hat{f}(x; \xi)$, and we denote the set of such surrogates as $\mathcal{S}_{L,p}(f, \xi)$.

We do not require the objective function, or its surrogate, to be differentiable, nor do we require the surrogate function to be continuous as a function of ξ . These are common assumptions in the literature, but are not necessary for our theory.

Algorithm 3 Majorized-MACE

```

1: Input: Initialize  $\mathbf{w}^{(0)}, \mathbf{r}^{(0)} \in \mathbb{R}^{n(L+3)}, \rho \in (0, 1)$ 
2:  $k = 0$ 
3: while not converged do
4:   for  $i = 1, \dots, N$  do
5:     Compute surrogate  $\hat{f}_i^{(k)} \in \mathcal{S}_{L_i, p_i}(f_i, r_i^{(k)})$ 
6:   end for
7:    $\mathbf{r}^{(k+1)} = \hat{\mathbf{F}}^{(k)}(\mathbf{w}^{(k)}; \mathbf{r}^{(k)})$ 
8:    $\mathbf{x} = 2\mathbf{r}^{(k+1)} - \mathbf{w}^{(k)}$ 
9:    $\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + 2\rho(\mathbf{G}(\mathbf{x}) - \mathbf{r}^{(k+1)})$ 
10:   $k = k + 1$ 
11: end while

```

Majorization-minimization schemes work by alternately minimizing the surrogate function and then updating the surrogate itself, which gives updates of the form

$$r^{(k+1)} \leftarrow \underset{r}{\operatorname{argmin}} \left\{ \hat{f}(r; r^{(k)}) \right\}.$$

The conditions in Definition 1 ensure that the original objective is decreasing, $f(r^{(k+1)}) \leq f(r^{(k)})$, and that the iterates $r^{(k)}$ converge to a minimizer of f .

Algorithm 3 applies these principles within the MACE framework. Assuming we have N agents, on the k -th iteration, we compute a surrogate \hat{f}_i of the objective function f_i at $r_i^{(k)}$ for $i = 1, \dots, N$. We define the concatenated operators

$$\hat{\mathbf{F}}^{(k)}(\mathbf{w}; \mathbf{r}^{(k)}) = [\hat{F}_1^{(k)}(w_1; r_1^{(k)}), \dots, \hat{F}_N^{(k)}(w_N; r_N^{(k)})]$$

and

$$\hat{\mathbf{T}}^{(k)} = (2\mathbf{G} - \mathbf{I}) \left(2\hat{\mathbf{F}}^{(k)}(\cdot; \mathbf{r}^{(k)}) - \mathbf{I} \right), \quad (14)$$

where $\hat{F}_i^{(k)}$ is the proximal map of $\hat{f}_i^{(k)}$. Akin to the fixed point iterations in (6), we analyze the fixed point of the system

$$\begin{aligned} \mathbf{w}^{(k+1)} &= \rho \hat{\mathbf{T}}^{(k)} \mathbf{w}^{(k)} + (1 - \rho) \mathbf{w}^{(k)}, \\ \mathbf{r}^{(k+1)} &= \hat{\mathbf{F}}^{(k)}(\mathbf{w}^{(k)}; \mathbf{r}^{(k)}). \end{aligned} \quad (15)$$

Our main result is Theorem 1, which shows that Algorithm 3 converges to an exact MACE solution. Part (i), states that any fixed point $(\mathbf{w}^*, \mathbf{r}^*)$ of (15) is an exact MACE solution as in (5). Part (ii) states that convergence to a fixed point is guaranteed if the surrogate is strongly convex. Note that if $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)}$, then $\hat{\mathbf{F}}^{(k+1)} = \hat{\mathbf{F}}^{(k)}$ and $\hat{\mathbf{T}}^{(k+1)} = \hat{\mathbf{T}}^{(k)}$, so the functions are also fixed at such a fixed point.

Theorem 1 (Majorized-MACE Solution). *Let $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, N$, be convex, and $\hat{f}_i^{(k)} \in \mathcal{S}_{L_i, p_i}(f_i, r_i^{(k)})$. Let F_i and $\hat{F}_i^{(k)}$ be the proximal maps of f_i and $\hat{f}_i^{(k)}$, respectively, and $\hat{\mathbf{T}}^{(k)}$ be the operator defined in (14). Then:*

- (i) Any fixed point of the augmented system (15) is a solution to the original MACE equation given in (5).
- (ii) The iterates defined by (15) converge to a fixed point, and hence a MACE solution, if one exists.

Proof. Proof is in Appendix B. □

V. RESULTS

We now demonstrate the effectiveness of CLAMP at reconstructing high-resolution images from multi-look coherent lidar data. We perform a series of experiments on both synthetic and experimental data to compare the performance of CLAMP with the standard speckle average method, as well as using sparsity-based regularization in the form of $\ell_{2,1}$ -regularization (ℓ_2 -regularization in cross-range and ℓ_1 -regularization in range), and isotropic total variation. These methods are implemented as described in Algorithm 2 with their corresponding regularizing agents in place of the proposed deep prior agent.

Further, we perform multiple experiments with different zero-padding factors, q , to investigate the effectiveness of CLAMP at different sampling rates. Specifically, we use $q = 1, 1.5$, and 2 . This results in images of various sizes, with the number of voxels in the image being q^3 times the number of measurements.

We also perform an ablation study of the aperture model in CLAMP to demonstrate its importance in achieving high-resolution reconstructions. In this case, we use the algorithm as described in Algorithm 2, but with $\mathbf{a} = \mathbf{1}$, effectively removing the aperture model from the reconstruction process. The resulting algorithm differs only in the μ -update (steps 6-8 of Algorithm 2), which can be simplified to be a closed-form solution of (8).

A. Methods

For each reconstruction presented in this section, we ran the corresponding algorithm for 250 iterations with $\rho = 0.5$. Each algorithm was initialized with $r = \frac{1}{L} \sum_{\ell=1}^L |A^H y_\ell|^2$ and setting \mathbf{w}, \mathbf{r} , to be stacked copies of r . Each μ_ℓ was initialized as $\frac{1}{\alpha} A^H y_i$. The forward model proximal parameter, σ^2 , was empirically chosen in a range of $[0.001, 1.0]$ to produce the best image quality.

To measure convergence of CLAMP, we define the convergence error based on the MACE equation (5) as

$$\text{Convergence Error} = \frac{\|\hat{\mathbf{F}}(\mathbf{w}) - \mathbf{G}(\mathbf{w})\|}{\|\mathbf{G}(\mathbf{w})\|}. \quad (16)$$

Additionally, to show convergence of the μ_ℓ updates, we define the relative residual averaged across looks as

$$\mu\text{-Residual} = \frac{1}{L} \sum_{\ell=1}^L \frac{\left\| \tilde{C}_\ell^{-1} \mu_\ell - \frac{1}{\sigma_w^2} A^H y_\ell \right\|}{\left\| \frac{1}{\sigma_w^2} A^H y_\ell \right\|}, \quad (17)$$

where the numerator comes from the exact solution of (8). Plots of CLAMP's convergence behavior are shown in Section V-E.

In the synthetic data experiments, we compare the reconstructed images to a ground truth image. We use peak-signal-to-noise ratio (PSNR), computed as

$$\text{PSNR}(\beta^* \hat{r}, r) = 10 \log_{10} \left(\frac{n}{\|\beta^* \hat{r} - r\|^2} \right),$$

to compare the reconstructions, \hat{r} , at $q = 2$ to the ground truth image, r . This was only computed for the $q = 2$

reconstructions, for that is when r and \hat{r} are the same size ($128 \times 128 \times 128$). The factor β^* is a multiplicative factor that accounts for any scaling differences between the reconstructed images and the ground truth image, and is computed as $\beta^* = \arg\max_{\beta} \{\text{PSNR}(\beta \hat{r}, r)\} = \hat{r}^T r / \|\hat{r}\|^2$. This will allow for a fair comparison of the reconstructions — any differences in PSNR will be due to the reconstruction method and not to a scaling factor.

In sparse 3D imaging, metrics on full 3D images can be misleading if they are dominated by small volumetric differences that otherwise have little impact on the image quality. To address this, we employ two point-cloud metrics to quantify the accuracy of the reconstruction of the surface reflectivity. We convert a 3D image, r , to a point cloud, $P \subseteq \mathbb{R}^4$, such that each point $p = (\mathbf{s}_p, r_p) \in P$ corresponds to a voxel in r with a value greater than a given threshold. In this work, we use the natural threshold, the noise floor, σ_w^2/α . The first component, $\mathbf{s}_p \in \mathbb{R}^3$, of p represents the spatial coordinates of the point, and r_p represents the reflectivity of that point.

We quantify the spatial accuracy of the reconstructions by computing the Euclidean distance between points in the reconstruction and points in the ground truth point cloud. For each reconstruction, \hat{r} , we transform the image into a point cloud $P = \{(\mathbf{s}_p, \hat{r}_p)\}$. For each point in P , we compute the Euclidean distance to its closest point in the ground truth point cloud, $Q = \{(\mathbf{t}_q, r_q)\}$. Let $\tilde{n}(p) = \arg\min_{q \in Q} \|\mathbf{s}_p - \mathbf{t}_q\|$ be the point in Q nearest to p , which, in practice, we determine by a k -d tree search [52]. The average Euclidean distance is then given by

$$\text{Euclidean Distance} = \frac{1}{|P|} \sum_{p \in P} \|\mathbf{s}_p - \mathbf{t}_{\tilde{n}(p)}\|.$$

To avoid corrupting this metric with outliers (points in the reconstruction with no nearby neighbor in the ground truth), we remove any point with a distance to the ground truth more than 3 times the Rayleigh criterion for resolution (approximately 1.5cm). Additionally, we report the false positive rate as the proportion of removed points relative to the total number of reconstructed points.

Using this method, we can also compute the normalized root mean squared error (NRMSE) of a reconstruction as

$$\text{NRMSE}(\beta^* \hat{r}, r) = \sqrt{\frac{\sum_{p \in P} (\beta^* \hat{r}_p - r_{\tilde{n}(p)})^2}{\sum_{p \in P} |r_{\tilde{n}(p)}|^2}},$$

where β^* is computed as $\beta^* = \arg\min_{\beta} \{\text{NRMSE}(\beta \hat{r}, r)\}$.

B. Synthetic Data Generation

In the synthetic data experiment, we generated data by simulating the multi-look coherent lidar imaging process described in Section II-B.

Each data simulation included nine looks at a 3D model of a train, shown in Figure 3a. Since 3D volumes can be difficult to visualize, we also show a pair of 2D depth and reflectivity images. The color bars shown in this figure are representative and can be applied to all images in this paper. The reflectivity

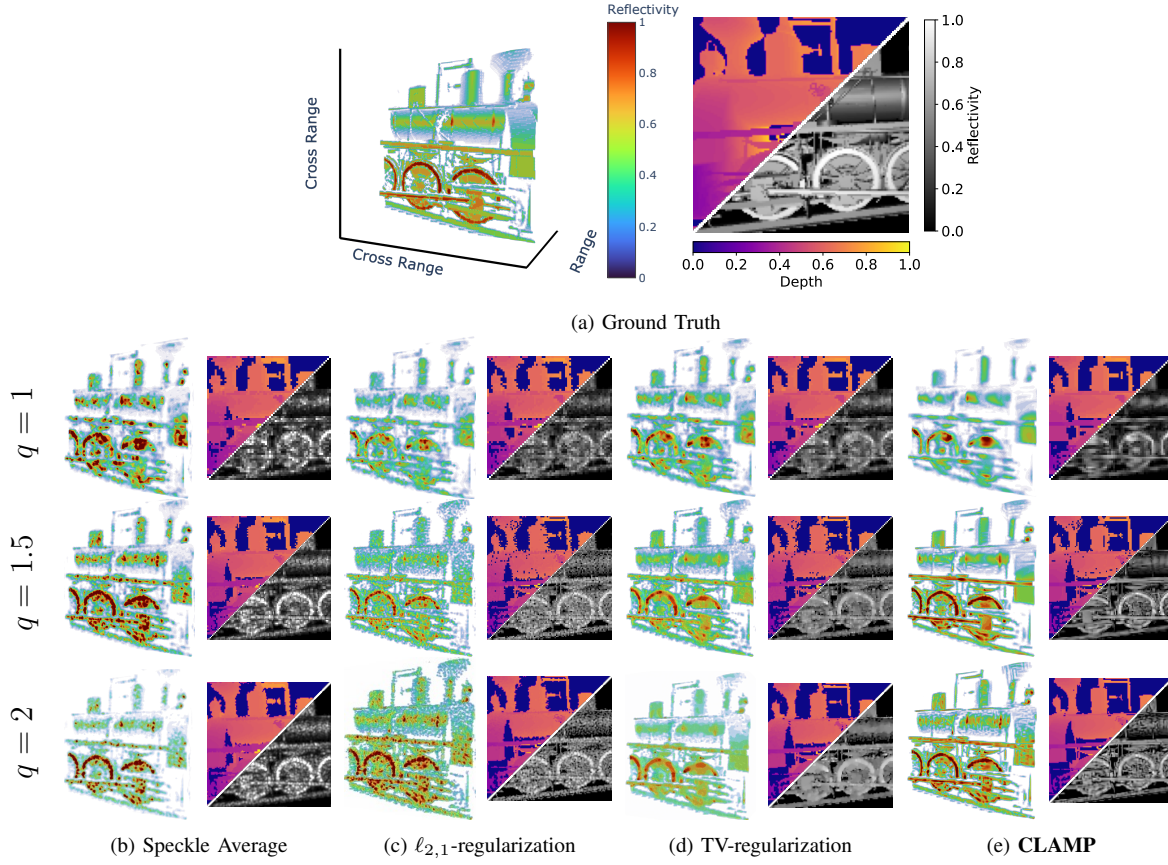


Fig. 3: Sample reconstructions obtained from synthetic data using different zero-padding factors, $q = 1, 1.5$ and 2 ; (a) the ground truth image of the train, (b)-(e) the reconstructions using the Speckle Average, $\ell_{2,1}$ -regularization, TV-regularization, and CLAMP, respectively. Each 3D image is shown with its 2D depth and reflectivity image. The CLAMP reconstructions show the highest resolution and best agreement with the ground truth.

image (shown in the bottom-right half) represents the maximum value along the depth dimension, while the depth (shown in the top-left half) is given by the location of the maximum. These images often reveal more detail than the 3D volume image, which can be useful for visualizing the reconstruction quality. However, due to the nature using the maximum value, undue noise can be introduced into the 2D images, particularly in the depth image.

The train was assumed to be 1 meter in size in each dimension. Assuming the train to be a Lambertian surface, the reflectivity of each point on the target was computed as proportional to the cosine of the angle between the surface normal and the viewing direction. The reflectivity was then normalized so that the brightest point in the image had unit reflectivity. A speckle realization was generated by multiplying the reflectivity of each voxel in the image by a complex Gaussian random variable with unit variance. The speckle realization was then propagated to the hologram plane by applying the forward model with parameters listed in the ‘Simulation’ column of Table I. The aperture was modeled as a centered, circular aperture with a diameter equal to 50% of the hologram grid length. Finally, the data is corrupted by adding complex white Gaussian noise with variance $\sigma_\eta^2 = 10^{-3}$.

C. Experimental Data Measurement

In addition to synthetic data, we evaluate CLAMP on data collected at the Air Force Research Laboratory. The experiment consisted of two targets, a toy car and a hexagonal nut, which were painted with matte white paint so that the surfaces had a uniform Lambertian reflectance. In order to measure multiple, statistically independent speckle realizations, the targets were placed on a high-precision rotation stage and rotated slightly after each measurement.

The targets were illuminated by a linear-frequency modulated waveform with central wavelength of 1550 nm (193.4 THz) and chirp rate of 117.9 THz/s. The laser was split so that 95% of the power was transmitted to the target and the remaining 5% was used as a reference beam for holographic imaging. The reference beam was placed off-axis in the plane of the exit pupil and pointed at the center of a focal plane array, where it interfered with a focused image of the target. The interference pattern was recorded on a 640×512 pixel InGaAs photodetector with $20 \mu\text{m}$ pixel pitch at 8-bit resolution. The camera’s region of interest was narrowed to fit the image of the target, thereby allowing faster frame rates. The toy car was measured by a 128×128 array of pixels at a frame rate of 17.60 kHz, and the hexagonal nut was measured by an 80×80 array at a frame rate of 28.92 kHz. The noise variance was estimated to be $\sigma_\eta^2 = 0.0012$ for the toy car

TABLE I: Parameters used in our multi-look coherent lidar simulation and experimental measurements of a hexagonal nut and toy car.

Item	Simulation	Hex. nut	Toy car	Units
Distance	52.9	2.64	2.64	m
Central wavelength	1550	1550	1550	nm
Chirp rate	—	117.9	117.9	THz/s
Chirp duration	—	2.0	2.0	ms
Frame rate	—	28.9	17.6	kHz
Frames, N_t	64	46	28	samples
Frequency step size	0.15	6.7	4.2	GHz
Aperture diameter	6.4	6.4	6.4	mm
Focal length	—	18.2	18.2	cm
Pixel pitch	—	20.0	20.0	μm
Hologram grid size	(128, 128)	(80, 80)	(128, 128)	pixels
Noise variance, σ_η^2	0.001	0.0025	0.0012	—
SNR	28.9	24.9	28.1	dB

and $\sigma_\eta^2 = 0.0025$ for the hexagonal nut. This estimation was done by computing the mean power within the pupil region of the hologram and dividing by the mean power outside of the pupil region. This is similar to the noise variance used in the synthetic data experiment, which allows for a comparison of the reconstruction quality between the two experiments.

Further details of the experimental setup, system hardware, and calibration methods can be found in [7, Chapter 4]. We summarize important experimental parameters in Table I.

D. Synthetic Data Results

In Figure 3, we compare the reconstructions obtained from the synthetic data using the speckle average method, $\ell_{2,1}$ -regularization, TV-regularization, and CLAMP using $q = 1, 1.5$, and 2. As shown in the figure, incorporation of sparsity-based regularization, such as $\ell_{2,1}$ and TV, can reduce speckle noise and improve the resolution of the reconstruction. However, in some cases, sparsity priors can also exacerbate speckle noise, as seen in the $\ell_{2,1}$ -regularization reconstructions. In contrast, CLAMP, with the use of a deep prior, is able to reconstruct a sharper and more detailed surface than the other methods. This is most evident in the undercarriage of the train, where the wheels and the tracks are more clearly defined in the CLAMP reconstruction with $q = 2$.

In Figure 4, we show the results of an ablation study of the aperture model in CLAMP. We show reconstructions of the synthetic train data from two nearly identical CLAMP algorithms with $q = 1, 1.5$, and 2. The first column of reconstructions, labeled “No aperture model,” is the result of running CLAMP using $\mathbf{a} = \mathbf{1}$, effectively ignoring the aperture model. The second column, labeled “With aperture model,” is the result of running the CLAMP algorithm as described in Algorithm 2. When $q = 1$, the reconstructions are nearly identical, as one should expect. However, at $q = 2$, the aperture model in CLAMP results in a sharper image and improves the ability to resolve high frequency components of the image.

In Table II, we show the PSNR, surface reflectivity NRMSE, Euclidean distance error and false positive rate for each reconstruction. As seen from the table, the CLAMP reconstruction is more similar to the ground truth than the Speckle Average, $\ell_{2,1}$ -regularization, and TV-regularization reconstructions, and with PSNR exceeding 30dB. The full CLAMP reconstruction

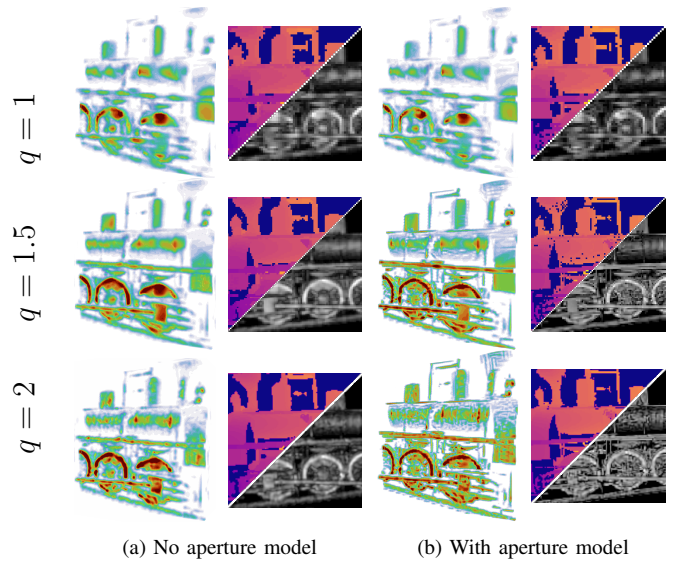


Fig. 4: Ablation study of the aperture model in CLAMP. Each 3D image is shown with its 2D depth and reflectivity image. When the zero-padding factor $q = 1$, the reconstructions are nearly identical. However, when $q = 1.5$ or $q = 2$, the aperture model in CLAMP improves the resolution of the reconstruction.

also tend to have the lowest surface reflectivity NRMSE and Euclidean distance error. This is likely since the other reconstructions are blurred and exhibit points further from the ground truth with lower reflectivity. While the sparsity-based regularization methods also improve this metric, not to the extent of the proposed method. The full CLAMP reconstruction shows improvement over the CLAMP reconstruction without the aperture model, particularly at higher zero-padding factors, further suggesting that the aperture model is crucial for quality, high-resolution reconstructions. Finally, the advantage of regularization is further supported by the false positive rate, as Speckle Average reconstruction exhibits a significantly higher rate compared to the regularized methods.

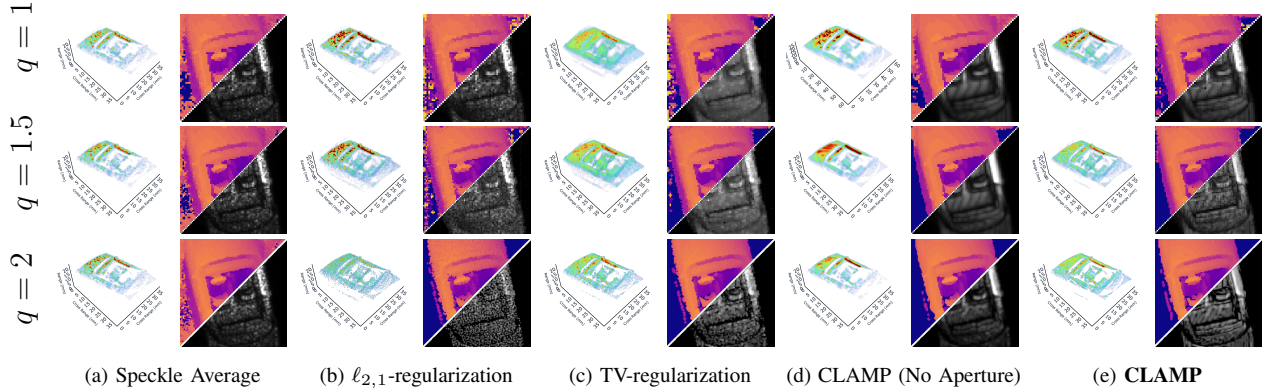
E. Experimental Data Results

Figure 5 shows reconstructions from nine looks at a toy car at zero-padding factors, $q = 1, 1.5$ and 2. Due to the advanced image prior, the CLAMP reconstructions show improved speckle reduction and sharpness over the Speckle Average, $\ell_{2,1}$, and TV-regularized reconstructions. The $\ell_{2,1}$ -regularized and TV-regularized reconstructions show some improvement over the Speckle Average, but the $\ell_{2,1}$ -regularized suffers from speckle noise, and the TV-regularized reconstruction is overly smoothed, with some of the high reflectivity surfaces spreading across the surface. The full CLAMP reconstruction again shows sharper and more detailed features over the CLAMP reconstructions without the aperture model.

In Figure 6, we show reconstructions of a hexagonal nut from nine looks for $q = 1, 1.5$, and 2. In each case, the CLAMP reconstructions show large improvements in reducing speckle noise. Particularly at $q = 2$, the full CLAMP method produces sharp image in which the threads of the nut are clearly visible. This is further illustrated in Figure 7,

TABLE II: PSNRs, surface reflectivity NRMSEs, and Euclidean distances (meters) between reconstructions and ground truth in our synthetic data experiment.

	Zero-padding Factor, q	Speckle Average	$\ell_{2,1}$ Regularization	TV Regularization	CLAMP (no aperture model)	CLAMP (with aperture model)
PSNR (dB)	2	26.97	27.29	28.40	28.85	30.65
Surface Reflectivity NRMSE	1	0.867	0.752	0.733	0.699	0.701
	1.5	0.847	0.805	0.770	0.621	0.427
	2	0.796	0.742	0.699	0.662	0.410
Euclidean Distance (m)	1	0.017	0.011	0.012	0.011	0.011
	1.5	0.018	0.031	0.010	0.031	0.008
	2	0.018	0.010	0.013	0.013	0.009
False Positive Rate	1	0.33	0.054	0.039	0.029	0.036
	1.5	0.48	0.031	0.006	0.021	0.008
	2	0.56	0.018	0.004	0.016	0.004

Fig. 5: Reconstructions of the toy car from experimental data with zero-padding factors $q = 1, 1.5$ and 2 . Each 3D image is shown with its 2D depth and reflectivity image. The CLAMP reconstructions show improved speckle-reduction over the Speckle Average, $\ell_{2,1}$ -, and TV-regularized reconstructions.

which shows the reflectivity values along the line segments shown in Figure 6. The full CLAMP reconstruction shows the consistently sharp and narrow peaks that are expected from a machine-threaded nut, while the other methods show broader and less consistent, well-defined peaks.

In Figure 8, we show the convergence error of CLAMP, defined in (16), and a normalized μ -residual averaged across looks, defined in (17), at each iteration of CLAMP for each dataset with $q = 2$. All instances of CLAMP converge to an error of less than 10^{-3} . Similarly, the μ -residual converges, indicating our iterative approach to solving (8) is converging to the exact solution.

VI. CONCLUSION

In this paper, we introduced CLAMP, a majorized PnP algorithm for multi-look coherent lidar image reconstruction. CLAMP is built on the MACE framework and consists of three main components: (1) an accurate surrogate physics-based model of coherent lidar, (2) an efficient method to invert the blurring effects of the aperture and (3) a deep 3D image prior model. Together, these components enable CLAMP to produce high-resolution images with reduced speckle.

To demonstrate its effectiveness, we applied CLAMP to synthetic and measured coherent lidar data. Our results show that CLAMP produces high-quality 3D images with reduced speckle and noise compared to speckle-averaging, while maintaining fidelity to multi-look data. Furthermore, ablating the

aperture model in CLAMP highlights the importance of our method in improving resolution.

Finally, we formalized the use of surrogate optimization in the MACE framework and proved convergence to the exact consensus equilibrium solution for a general class of surrogate functions. Our formalization provides a theoretical foundation for the use of majorization-minimization in consensus optimization problems, is general and applicable to many other imaging problems, and includes many existing image reconstruction algorithms as special cases.

APPENDIX A QUADRATIC SURROGATE

To further reduce the computational cost of the proximal operator in (13), we follow [53] and construct a quadratic first-order surrogate, \hat{Q}_ℓ , of \hat{f}_ℓ , which is then a first-order surrogate of the original function f_ℓ . For the remainder of this section, we drop the ℓ subscript for simplicity, but note that the following surrogate is formed for each look $\ell = 1, \dots, L$.

To construct the surrogate, we use the approach in [54, Section IV], which has two primary components: determining the interval of validity for the surrogate, which depends on r' and v , and determining the surrogate itself. We first obtain a convex surrogate, Q , of \hat{f} by a first-order approximation at r' of the logarithm in (12):

$$Q(r; r') = \sum_{j=1}^n \frac{r_j}{r'_j} + \frac{1}{r_j} (c_j + |\mu_j|^2).$$

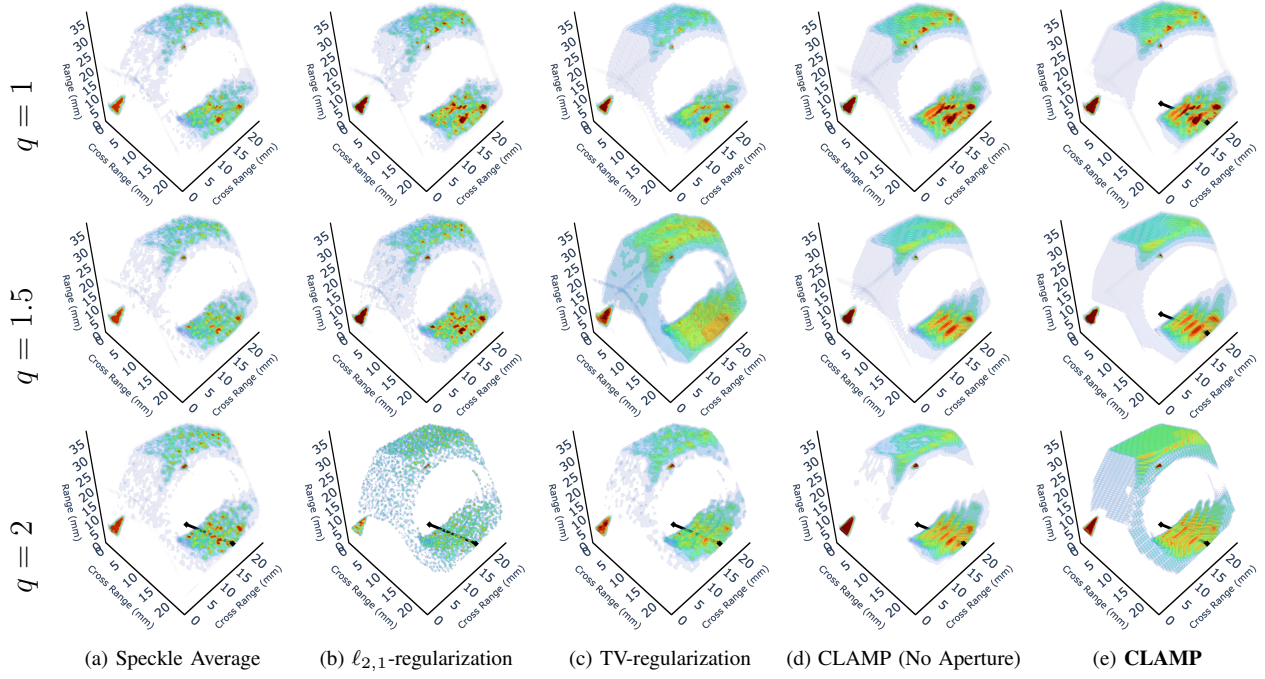


Fig. 6: Reconstructions of the hexagonal nut from experimental data with zero-padding factors, $q = 1, 1.5$ and 2 . Several reconstructions show a line segment through the threads of the nut which are plotted in Figure 7. The CLAMP reconstructions show the best speckle reduction, resolution, and sharpest edges.

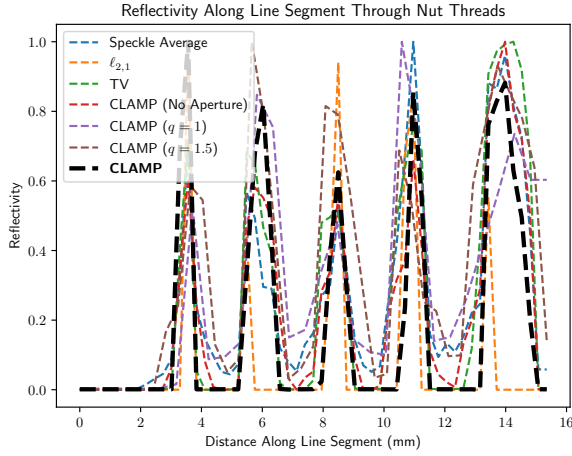


Fig. 7: A line segment from the hexagonal nut reconstructions shown in Figure 6. The CLAMP reconstruction (black) shows the sharpest and most well-defined peaks.

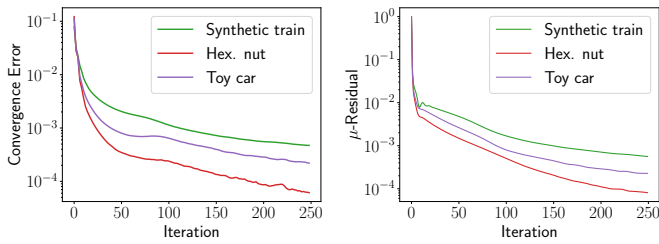


Fig. 8: Convergence error and μ -residual of CLAMP with aperture modeling for synthetic and experimental data. CLAMP converged to an error of less than 10^{-3} in all cases.

Treating each j separately, it suffices to form a surrogate the function

$$\rho_j(r_j) = \frac{r_j}{r'_j} + \frac{c_j + |\mu_j|^2}{r_j} + \frac{1}{2\sigma^2}(r_j - v_j)^2.$$

We use (11) and (12) of [54] to fit a quadratic surrogate of ρ_j by a linear interpolation of its derivatives given by

$$\hat{\rho}_j(r_j) = \frac{a_j}{2}r_j^2 + b_jr_j.$$

Summing over j yields the surrogate \hat{Q} for Q .

The surrogate $\hat{\rho}_j$ is constructed by a linear interpolation of the derivative ρ'_j at two points: r'_j and ξ_j . The point ξ_j is chosen so that it is in the direction of the minimum of ρ_j from r'_j . When $\rho'_j(r'_j) > 0$, we will need $r_j < r'_j$ to minimize ρ , while if $\rho'_j(r'_j) < 0$ we will need $r_j > r'_j$. Hence we select a scaling factor $\beta > 1$ and choose $\xi_j = r'_j/\beta$ in the first case and $\xi_j = \beta r'_j$ in the second. This defines an interval of validity of the surrogate between r'_j and ξ_j , which we denote \mathcal{I}_j .

The reflectivity update in (13) is then replaced by the proximal map of \hat{Q} at v , which is

$$r' \leftarrow \operatorname{argmin}_{r_j \in \mathcal{I}_j, \forall j} \left\{ \hat{Q}(r; r') + \frac{1}{2\sigma^2} \|r - v\|^2 \right\}.$$

Since the objective is separable, this simplifies to

$$r'_j \leftarrow \frac{v_j - \sigma^2 b_j}{1 + \sigma^2 a_j}$$

but clipped to lie in the interval \mathcal{I}_j . Finally, we empirically choose $\beta = 1 + \frac{2}{\log(k)}$ where k is the CLAMP iteration number, which ensures that β decays to 1 more slowly than the convergence of r' .

APPENDIX B PROOF OF THEOREM 1

In this section, we prove the convergence of majorized MACE to a fixed point of the original MACE equation when the surrogate functions meet the conditions of Definition 1.

We begin by stating a few lemmas before proving the main results. First, a crucial, perhaps intuitive, fact is that the objective function and its surrogate have the same subgradients at the point of approximation. This is stated in Lemma 1 and is a result of the smoothness assumption in Definition 1.

Lemma 1. *Let $\hat{f} \in \mathcal{S}_{L,p}(f, \xi)$ (see Definition 1). Then $\partial f(\xi) = \partial \hat{f}(\xi)$.*

Proof. Proof is given in the supplemental material. We note that this result holds even if \hat{f} is not strongly convex. \square

Another critical fact is given in Lemma 2, which is adapted from [36, Lemma 2].

Lemma 2. *Let $\hat{f} \in \mathcal{S}_{L,p}(f, \xi)$ (see Definition 1), and let $u \in \partial \hat{f}(x)$ be any subgradient of \hat{f} at x . Then for any y ,*

$$\begin{aligned} f(x) - f(y) - \langle u, x - y \rangle \\ \leq \frac{1}{2} (L \|y - \xi\|^2 - p \|x - y\|^2). \end{aligned}$$

Proof. Proof is given in the supplemental material. \square

We are now ready to prove part (i) of Theorem 1.

Proof of Theorem 1 part (i). Let $(\mathbf{w}^*, \mathbf{r}^*)$ be a fixed point of (15). Since $\mathbf{r}^* = \hat{\mathbf{F}}^* \mathbf{w}^*$, where $\hat{\mathbf{F}}^*$ is the stacked proximal operator of surrogates $\hat{f}_i(x; \mathbf{r}_i^*)$, we have

$$-\frac{1}{\sigma^2} (r_i^* - w_i^*) \in \partial \hat{f}_i(r_i^*; \mathbf{r}^*),$$

and by Lemma 1,

$$-\frac{1}{\sigma^2} (r_i^* - w_i^*) \in \partial f_i(r_i^*),$$

for $i = 1, \dots, N$.

Thus, $\mathbf{r}^* = \mathbf{F} \mathbf{w}^*$. Since $\mathbf{w}^* = (2\mathbf{G} - \mathbf{I})(2\hat{\mathbf{F}}^* - \mathbf{I}) \mathbf{w}^*$, we get $\mathbf{w}^* = (2\mathbf{G} - \mathbf{I})(2\mathbf{F} - \mathbf{I}) \mathbf{w}^*$, proving \mathbf{w}^* is a MACE solution. \square

In order to prove Theorem 1 part (ii), we prove convergence of an algorithm that is equivalent to Algorithm 3. First, it is noted in [41] and elsewhere that using Mann iterations to solve the MACE equations, as written in (6), is equivalent up to a change of variables to a form of ADMM [35].

We present the equivalency and prove convergence for $\rho = 1/2$, though with extra bookkeeping, and using the appropriate variant of ADMM, one can generalize this result for $\rho \in (0, 1)$.

To begin, consider one iteration of Algorithm 3, written as in (15), using $\rho = 1/2$,

$$\begin{aligned} \mathbf{w}^{(k+1)} &= \frac{1}{2} (2\mathbf{G} - \mathbf{I})(2\hat{\mathbf{F}}^{(k)} - \mathbf{I}) \mathbf{w}^{(k)} + \frac{1}{2} \mathbf{w}^{(k)}, \\ \mathbf{r}^{(k+1)} &= \hat{\mathbf{F}}^{(k)} \mathbf{w}^{(k)}, \end{aligned}$$

where we recall that $\hat{\mathbf{F}}^{(k)} \mathbf{w}^{(k)} = \hat{\mathbf{F}}^{(k)}(\mathbf{w}^{(k)}; \mathbf{r}^{(k)})$. Since \mathbf{G} is linear and $\mathbf{G}\mathbf{G} = \mathbf{G}$, some algebra shows that applying \mathbf{G} to

both sides of the first equation yields $\mathbf{G}\mathbf{w}^{(k+1)} = \mathbf{G}\mathbf{r}^{(k+1)}$, and hence $\mathbf{G}\mathbf{w}^{(k+1)} = \mathbf{G}\hat{\mathbf{F}}^{(k)} \mathbf{w}^{(k)}$ from the second equation.

Using this to re-express the first equation, recalling that $\bar{\mathbf{r}}$ is one averaged image in $\mathbf{G}\mathbf{r}$, and noting that \mathbf{w} and \mathbf{r} may be updated in either order, the i -th component (r_i, w_i) satisfies

$$r_i^{(k+1)} = \hat{F}_i^{(k)} w_i^{(k)}, \quad (18)$$

$$\begin{aligned} w_i^{(k+1)} &= w_i^{(k)} + (\bar{\mathbf{r}}^{(k+1)} - r_i^{(k+1)}) \\ &\quad + (\bar{\mathbf{r}}^{(k+1)} - \bar{\mathbf{r}}^{(k)}). \end{aligned} \quad (19)$$

Now, we introduce the variables $u_i^{(k)} = \bar{\mathbf{r}}^{(k)} - w_i^{(k)}$ and $z^{(k+1)} = \bar{\mathbf{r}}^{(k+1)} + \bar{\mathbf{u}}^{(k)}$. Since $\bar{\mathbf{w}}^{(k+1)} = \bar{\mathbf{r}}^{(k+1)}$, we have $\bar{\mathbf{u}}^{(k+1)} = 0$ for $k \geq 1$, and so for $k \geq 3$, the updates (18)–(19) can be written equivalently as

$$r_i^{(k+1)} = \hat{F}_i^{(k)} (z^{(k)} - u_i^{(k)}), \quad (20)$$

$$z^{(k+1)} = \frac{1}{N} \sum_{i=1}^N (r_i^{(k+1)} + u_i^{(k)}), \quad (21)$$

$$u_i^{(k+1)} = u_i^{(k)} + (r_i^{(k+1)} - z^{(k+1)}), \quad (22)$$

where (22) follows from (19) using $w_i^{(k)} = \bar{\mathbf{r}}^{(k)} - u_i^{(k)}$ and simplifying. The updates (20)–(22) are a form of consensus ADMM [35, Chapter 7] to solve the problem,

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{r}) = \sum_{i=1}^N f_i(r_i) \\ \text{subject to} \quad & r_i = z, \end{aligned} \quad (23)$$

except the updates of r_i in (20) are given by proximal maps of surrogate functions instead of proximal maps of f_i .

In Theorem 2, we prove that despite this difference, the iterations (20)–(22), converge to a Karush-Kuhn-Tucker (KKT) point of (23), which is sufficient for optimality. A KKT point can be found from the Lagrangian of (23),

$$\mathcal{L}(\mathbf{r}, z, \mathbf{u}) = f(\mathbf{r}) + \frac{1}{\sigma^2} \sum_{i=1}^N \langle u_i, r_i - z \rangle, \quad (24)$$

(24) where \mathbf{u} is the dual variable, and $\langle \cdot, \cdot \rangle$ is the inner-product. Specifically, the point $(\mathbf{r}^*, z^*, \mathbf{u}^*)$ is a KKT point if and only if

$$\begin{aligned} -\frac{1}{\sigma^2} u_i^* &\in \partial f_i(r_i^*), \\ r_i^* &= z^*, \end{aligned}$$

for $i = 1, \dots, N$.

Theorem 2. *Assume the hypotheses of Theorem 1, and that (23) has a KKT point. Then the sequence $\{(\mathbf{r}^{(k)}, z^{(k)}, \mathbf{u}^{(k)})\}$ generated by the iterations (20)–(22) converges to a KKT point of (23).*

To prove Theorem 2, we first use Proposition 1 to characterize the subgradients of the surrogate functions. Then in Lemma 3, we state the vital property that the distance between the iterates and a KKT point decreases at each iteration.

Proposition 1. Assume the hypotheses of Theorem 1 and let $\{(\mathbf{r}^{(k)}, z^{(k)}, \mathbf{u}^{(k)})\}$ be generated by (20)-(22). Then

$$-\frac{1}{\sigma^2} \left(r_i^{(k+1)} - z^{(k)} + u_i^{(k)} \right) \in \partial \hat{f}_i^{(k)}(r_i^{(k+1)}), \quad (25)$$

for $i = 1, \dots, N$.

Proof. This follows from the definition of the proximal map in (20). \square

To simplify notation, we define

$$u_i^{(k+\frac{1}{2})} \triangleq u_i^{(k)} + (r_i^{(k+1)} - z^{(k)}) \quad (26)$$

which can be thought of as an intermediate update of the dual variable after the update of \mathbf{r} in (20) but before the update of z in (21). We can then rewrite the subgradients in (25) as

$$-\frac{1}{\sigma^2} u_i^{(k+\frac{1}{2})} \in \partial \hat{f}_i^{(k)}(r_i^{(k+1)}).$$

Lemma 3. Assume the hypotheses of Theorem 1, and that (23) has a KKT point, $(\mathbf{r}^*, z^*, \mathbf{u}^*)$. Then the sequence $\{(\mathbf{r}^{(k)}, z^{(k)}, \mathbf{u}^{(k)})\}$ generated by the iterations (20)-(22) satisfies

$$E^{(k+1)}(\mathbf{r}^*, z^*, \mathbf{u}^*) \leq E^{(k)}(\mathbf{r}^*, z^*, \mathbf{u}^*),$$

where

$$E^{(k)}(\mathbf{r}^*, z^*, \mathbf{u}^*) = \sum_{i=1}^N \left(L_i \left\| r_i^{(k)} - r_i^* \right\|^2 + \frac{1}{\sigma^2} \left\| z^{(k)} - z^* \right\|^2 + \frac{1}{\sigma^2} \left\| u_i^{(k)} - u_i^* \right\|^2 \right).$$

Proof. Proof is given in the supplemental material. \square

We are now ready to state and prove convergence of the iterations (20)-(22).

Proof of Theorem 2. Let $(\mathbf{r}^*, z^*, \mathbf{u}^*)$ be a KKT point of (23). By Lemma 3, we have that $\{(\mathbf{r}^{(k)}, z^{(k)}, \mathbf{u}^{(k)})\}$ is bounded. Let $\{(\mathbf{r}^{(k_j)}, z^{(k_j)}, \mathbf{u}^{(k_j)})\}$ be a convergent subsequence with limit point $(\mathbf{r}^\dagger, z^\dagger, \mathbf{u}^\dagger)$. We will show that this limit point is also a KKT point of (23).

From (43)-(44) in the proof of Lemma 3, the iterates satisfy

$$\frac{1}{2\sigma^2} \sum_{i=1}^N \left\| z^{(k+1)} - z^{(k)} \right\|^2 + \left\| u_i^{(k+1)} - u_i^{(k)} \right\|^2 \quad (28)$$

$$\leq \frac{1}{2} \left(E^{(k)}(\mathbf{r}^*, z^*, \mathbf{u}^*) - E^{(k+1)}(\mathbf{r}^*, z^*, \mathbf{u}^*) \right). \quad (29)$$

Using Lemma 3, summing both sides of (28)-(29) over all k , and discarding the telescoping terms, we get $\|z^{(k+1)} - z^{(k)}\| \rightarrow 0$ and $\|u_i^{(k+1)} - u_i^{(k)}\| \rightarrow 0$ as $k \rightarrow \infty$. Since $z^{(k_j)} \rightarrow z^\dagger$ as $j \rightarrow \infty$, we can set $k = k_j$ and take $j \rightarrow \infty$ to get $z^{(k_j+1)} \rightarrow z^\dagger$. Likewise, $u_i^{(k_j)} \rightarrow u_i^\dagger$.

Rearranging (22), we have

$$r_i^{(k_j+1)} = u_i^{(k_j+1)} - u_i^{(k_j)} + z^{(k_j+1)},$$

so $r_i^{(k_j+1)}$ converges to z^\dagger , which we denote also as r_i^\dagger . Lastly, $u_i^{(k_j+\frac{1}{2})} \rightarrow u_i^\dagger$ can be seen by taking the limit of (26) along the subsequence $k = k_j$.

Since $-\frac{1}{\sigma^2} u_i^{(k_j+\frac{1}{2})} \in \partial \hat{f}_i^{(k_j)}(r_i^{(k_j+1)})$, we have by Lemma 2 that for all \mathbf{r} ,

$$f(\mathbf{r}^{(k_j+1)}) - f(\mathbf{r}) + \frac{1}{\sigma^2} \sum_{i=1}^N \left\langle u_i^{(k_j+\frac{1}{2})}, r_i^{(k_j+1)} - r_i \right\rangle \leq 0.$$

Since f is convex and finite on \mathbb{R}^n and $r_i^{(k_j+1)} \rightarrow r_i^\dagger$, we have that $f(\mathbf{r}^{(k_j+1)}) \rightarrow f(\mathbf{r}^\dagger)$, and so taking the limit of the above inequality along the subsequence gives

$$f(\mathbf{r}^\dagger) - f(\mathbf{r}) + \frac{1}{\sigma^2} \sum_{i=1}^N \left\langle u_i^\dagger, r_i^\dagger - r_i \right\rangle \leq 0,$$

which implies that $-\frac{1}{\sigma^2} u_i^\dagger \in \partial f_i(\mathbf{r}^\dagger)$ for $i = 1, \dots, N$. This completes the proof that $(\mathbf{r}^\dagger, z^\dagger, \mathbf{u}^\dagger)$ is a KKT point.

Finally, we show that $(\mathbf{r}^\dagger, z^\dagger, \mathbf{u}^\dagger)$ is the only limit point of the sequence $\{(\mathbf{r}^{(k)}, z^{(k)}, \mathbf{u}^{(k)})\}$. First, we replace $(\mathbf{r}^*, z^*, \mathbf{u}^*)$ with $(\mathbf{r}^\dagger, z^\dagger, \mathbf{u}^\dagger)$ in Lemma 3. Then for any $k > k_j$, we have

$$E^{(k)}(\mathbf{r}^\dagger, z^\dagger, \mathbf{u}^\dagger) \leq E^{(k_j)}(\mathbf{r}^\dagger, z^\dagger, \mathbf{u}^\dagger).$$

Since the right hand side converges to 0 as $j \rightarrow \infty$, the definition of $E^{(k)}$ implies that $(\mathbf{r}^{(k)}, z^{(k)}, \mathbf{u}^{(k)}) \rightarrow (\mathbf{r}^\dagger, z^\dagger, \mathbf{u}^\dagger)$ as $k \rightarrow \infty$. \square

To end, we prove Theorem 1 part (ii).

Proof of Theorem 1 part (ii). As noted after equation (19), by introducing the variables $u_i^{(k)} = \bar{\mathbf{r}}^{(k)} - w_i^{(k)}$, and $z^{(k+1)} = \bar{\mathbf{r}}^{(k+1)} + \bar{\mathbf{u}}^{(k)}$, the iterates of (15) are equivalent to the iterates (20)-(22). By Theorem 2, $(\mathbf{r}^{(k)}, z^{(k)}, \mathbf{u}^{(k)})$ converges to a KKT point, $(\mathbf{r}^*, z^*, \mathbf{u}^*)$, of (23). Resubstituting the change of variables, we get $(\mathbf{w}^*, \mathbf{r}^*)$ is a fixed point of the system (15), which by Theorem 1 part (i) is a solution to the exact MACE formulation. \square

REFERENCES

- [1] M. F. Spencer, "Spatial Heterodyne," in *Encyclopedia of Modern Optics*, vol. 4, Elsevier, 2018.
- [2] N. H. Farhat, "Holography, wave-length diversity and inverse scattering," *AIP Conference Proceedings*, vol. 65, no. 1, p. 627-642, 1980.
- [3] J. C. Marron and T. J. Schulz, "Three-dimensional, fine-resolution imaging using laser frequency diversity," *Optics Letters*, vol. 17, no. 4, p. 285, 1992.
- [4] J. C. Marron and K. S. Schroeder, "Three-dimensional lensless imaging using laser frequency diversity," *Applied Optics*, vol. 31, no. 2, p. 255, 1992.
- [5] J. W. Stafford, B. D. Duncan, and D. J. Rabb, "Phase gradient algorithm method for three-dimensional holographic lidar imaging," *Applied Optics*, vol. 55, no. 17, p. 4611, 2016.
- [6] M. T. Banet and J. R. Fienup, "Effects of speckle decorrelation on motion-compensated, multi-wavelength, 3D digital holography," in *Unconventional Imaging and Adaptive Optics 2022* (J. J. Dolne and M. F. Spencer, eds.), vol. 12239, p. 122390C, International Society for Optics and Photonics, SPIE, 2022.
- [7] W. Farriss, *Iterative Phase Estimation Algorithms in Interferometric Systems*. PhD thesis, University of Rochester, 2021.
- [8] W. E. Farriss, J. R. Fienup, J. W. Stafford, and N. J. Miller, "Sharpness-based correction methods in holographic aperture lidar (HAL)," in *Unconventional and Indirect Imaging, Image Reconstruction, and Wavefront Sensing 2018*, vol. 10772, pp. 178-185, SPIE, 2018.
- [9] M. T. Banet and J. R. Fienup, "Image sharpening on 3D intensity data in deep turbulence with scintillated illumination," in *Unconventional Imaging and Adaptive Optics 2021*, vol. 11836, pp. 116-124, SPIE, 2021.

- [10] M. T. Banet, J. R. Fienup, J. D. Schmidt, and M. F. Spencer, "3D multi-plane sharpness metric maximization with variable corrective phase screens," *Applied Optics*, vol. 60, no. 25, p. G243, 2021.
- [11] g.-i. family=OKTEM, given=FIGEN, "Sparsity-based three-dimensional image reconstruction for near-field MIMO radar imaging," vol. 27, no. 5, pp. 3282–3295.
- [12] Y. Wang, Z. He, X. Zhan, Q. Zeng, and Y. Hu, "A 3-d sparse sar imaging method based on plug-and-play," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.
- [13] J. W. Goodman, *Speckle Phenomena in Optics: Theory and Applications, Second Edition*. SPIE, 2020.
- [14] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.
- [15] D. Munson and J. Sanz, "Image reconstruction from frequency-offset Fourier data," *Proceedings of the IEEE*, vol. 72, pp. 661–669, June 1984.
- [16] O. Oral and F. S. Oktem, "Plug-and-play regularization on magnitude with deep priors for 3d near-field mimo imaging," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 762–773, 2024.
- [17] C. J. Pellizzari, M. F. Spencer, and C. A. Bouman, "Phase-error estimation and image reconstruction from digital-holography data using a Bayesian framework," *Journal of the Optical Society of America A*, vol. 34, no. 9, p. 1659, 2017.
- [18] C. J. Pellizzari, R. Trahan, H. Zhou, S. Williams, S. E. Williams, B. Nemati, M. Shao, and C. A. Bouman, "Synthetic Aperture LADAR: A Model-Based Approach," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 901–916, 2017.
- [19] C. J. Pellizzari, M. F. Spencer, and C. A. Bouman, "Coherent Plug-and-Play: Digital Holographic Imaging Through Atmospheric Turbulence Using Model-Based Iterative Reconstruction and Convolutional Neural Networks," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1607–1621, 2020.
- [20] C. J. Pellizzari, M. F. Spencer, and C. A. Bouman, "Imaging through distributed-volume aberrations using single-shot digital holography," *Journal of the Optical Society of America A*, vol. 36, no. 2, p. A20, 2019.
- [21] C. J. Pellizzari, R. Trahan, H. Zhou, S. Williams, S. E. Williams, B. Nemati, M. Shao, and C. A. Bouman, "Optically coherent image formation and denoising using a plug and play inversion framework," *Applied Optics*, vol. 56, no. 16, p. 4735, 2017.
- [22] L. E. Baum and T. Petrie, "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554 – 1563, 1966.
- [23] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 698 – 704, 1970.
- [24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [25] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-Play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948, IEEE, 2013.
- [26] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-Play Priors for Bright Field Electron Tomography and Sparse Interpolation," *IEEE Transactions on Computational Imaging*, pp. 1–1, 2016.
- [27] C. A. Bouman, *Foundations of Computational Imaging: A Model-Based Approach*. Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, 2022.
- [28] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 85–97, 2023.
- [29] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Robust bayesian target detection algorithm for depth imaging from sparse single-photon data," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 456–467, 2016.
- [30] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 445–459, 2017.
- [31] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.
- [32] A. Halimi, R. Tobin, A. McCarthy, S. McLaughlin, and G. S. Buller, "Restoration of multilayered single-photon 3d lidar images," in *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 708–712, 2017.
- [33] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tournet, and S. McLaughlin, "Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nature Communications*, vol. 10, p. 4984, Nov. 2019.
- [34] D. Tucker and L. C. Potter, "Speckle Suppression in Multi-Channel Coherent Imaging: A Tractable Bayesian Approach," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1429–1439, 2020.
- [35] S. Boyd, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [36] C. Lu, J. Feng, S. Yan, and Z. Lin, "A unified alternating direction method of multipliers by majorization minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 527–541, 2018.
- [37] J. Mairal, "Incremental majorization-minimization optimization with application to large-scale machine learning," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 829–855, 2015.
- [38] J. Mairal, "Optimization with first-order surrogate functions," in *Proceedings of the 30th International Conference on Machine Learning (S. Dasgupta and D. McAllester, eds.)*, vol. 28 of *Proceedings of Machine Learning Research*, (Atlanta, Georgia, USA), pp. 783–791, PMLR, 17–19 Jun 2013.
- [39] D. R. H. Kenneth Lange and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 1–20, 2000.
- [40] T. Allen, D. Rabb, G. T. Buzzard, and C. A. Bouman, "Multi-Agent Consensus Equilibrium for Range Compressed Holographic Surface Reconstruction," in *21st Coherent Laser Radar Conference*, 2022.
- [41] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, "Plug-and-Play Unplugged: Optimization-Free Reconstruction Using Consensus Equilibrium," *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2001–2020, 2018.
- [42] J. Goodman, *Introduction to Fourier Optics*. McGraw-Hill, 2nd ed., 1996.
- [43] D. E. Thornton, C. J. Radosevich, S. Horst, and M. F. Spencer, "Achieving the shot-noise limit using experimental multi-shot digital holography data," *Opt. Express*, vol. 29, pp. 9599–9617, Mar 2021.
- [44] J. W. Goodman, "Some fundamental properties of speckle*," *Journal of the Optical Society of America*, vol. 66, no. 11, p. 1145, 1976.
- [45] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An Information-Rich 3D Model Repository," Tech. Rep. arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.
- [46] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, third ed., 1996.
- [47] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of computation*, vol. 19, no. 90, pp. 297–301, 1965.
- [48] R. Gribonval, "Should Penalized Least Squares Regression be Interpreted as Maximum A Posteriori Estimation?," vol. 59, no. 5, pp. 2405–2410, 2011.
- [49] X. Xu, Y. Sun, J. Liu, B. Wohlberg, and U. S. Kamilov, "Provable Convergence of Plug-and-Play Priors With MMSE Denoisers," vol. 27, pp. 1280–1284, 2020.
- [50] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2411–2422, 2007.
- [51] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.
- [52] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software*, vol. 3, p. 209–226, Sept. 1977.
- [53] V. Sridhar, S. J. Kisner, S. P. Midkiff, and C. A. Bouman, "Fast algorithms for model-based imaging through turbulence," in *Artificial Intelligence and Machine Learning in Defense Applications II (J. Dijk, ed.)*, vol. 11543, p. 1154304, International Society for Optics and Photonics, SPIE, 2020.
- [54] Z. Yu, J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "Accelerated Line Search for Coordinate Descent Optimization," in *2006 IEEE Nuclear Science Symposium Conference Record*, vol. 5, pp. 2841–2844, 2006.