6-10-2016

Phase Gradient Algorithm Method for 3-D Holographic Ladar Imaging

Jason W. Stafford  
University of Dayton

Bradley D. Duncan  
University of Dayton, bduncan1@udayton.edu

David J. Rabb  
Air Force Research Laboratory

Follow this and additional works at: http://ecommons.udayton.edu/eop_fac_pub

Part of the Electromagnetics and Photonics Commons, Optics Commons, and the Other Physics Commons

eCommons Citation  
Stafford, Jason W.; Duncan, Bradley D.; and Rabb, David J., "Phase Gradient Algorithm Method for 3-D Holographic Ladar Imaging" (2016). Electro-Optics and Photonics Faculty Publications. Paper 41.  
http://ecommons.udayton.edu/eop_fac_pub/41

This Article is brought to you for free and open access by the Department of Electro-Optics and Photonics at eCommons. It has been accepted for inclusion in Electro-Optics and Photonics Faculty Publications by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.
Phase gradient algorithm method for 3-D holographic ladar imaging

JASON W. STAFFORD\textsuperscript{1,2,*}, BRADLEY D. DUNCAN\textsuperscript{1}, DAVID J. RABB\textsuperscript{2}

\textsuperscript{1}Ladar Technologies Branch, 3109 Hobson Way, Building 622, AFRL/RYMM, Wright-Patterson AFB, OH 45433-7700
\textsuperscript{2}Electro-Optics Program, University of Dayton, 300 College Park, Dayton, OH 45469-0245
\*Corresponding author: staffojw@udayton.edu

Received X

3-D holographic ladar uses digital holography with frequency diversity to add the ability to resolve targets in range. A key challenge is that since individual frequency samples are not recorded simultaneously, differential phase aberrations may exist between them making it difficult to achieve range compression. We describe steps specific to this modality so that phase gradient algorithms (PGA) can be applied to 3-D holographic ladar data for phase corrections across multiple temporal frequency samples. Substantial improvement of range compression is demonstrated with a laboratory experiment where our modified PGA technique is applied. Additionally, the PGA estimator is demonstrated to be efficient for this application and the maximum entropy saturation behavior of the estimator is analytically described.

\textbf{OCIS codes:} (090.1995) Digital holography; (120.0280) Remote sensing and sensors; (280.3640) Lidar; (280.4788) Optical sensing and sensors; (280.6730) Synthetic aperture radar, (100.6890) Three-dimensional image processing.

http://

1. INTRODUCTION

Digital holography is a well-known method for recovering the complex electromagnetic field backscattered from a potentially distant target. The reconstructed field is, however, in effect only 2-D because phase information (i.e., depth) wraps modulo 2π on the order of the optical wavelength. A significant advancement in the field was the development of 3-D reconstruction methods by the introduction of frequency diversity during hologram recording [1]. With this technique, a series of 2-D complex images of differing temporal frequencies are range compressed by performing an inverse discrete Fourier transform (IDFT) over the temporal frequency spectrum. This has been demonstrated with up to 64 discrete frequencies in both phase stepping interferometry and spatial heterodyne (or, off-axis recording) configurations [2-4]. In a scenario where this technique is implemented at extended ranges, which we call 3-D holographic ladar, important issues arise that have not been fully addressed in the literature to date.

A key challenge is that when individual frequency samples are not recorded simultaneously, differential phase aberrations may exist between them. Over long ranges, the cumulative effect of atmospheric phase disturbances makes these aberrations even more probable. In addition, other key sources of phase aberrations arise from performance limitations of the illumination source such as imperfect linear frequency steps and laser decoherence over the time required to collect multiple frequency samples. These challenges make it difficult to combine the frequency samples to achieve a range compressed 3-D image. An early solution to this problem was shown in [3], where prominent scatters were used to estimate local phase corrections. This method often works well but requires a prominent point, which is not always available. Another approach recently demonstrated in [4], is to monitor aberrations in situ using a monitor (or "pilot") laser at a stationary frequency to calculate the phase errors. Whatever the method, to realize maximum performance, the phase errors must be estimated and corrected before the IDFT is performed.

In synthetic aperture ladar (SAL), which uses a temporal interferometric technique analogous to the spatial one of digital holography, many sophisticated algorithms have been developed to address the problem of aberration correction over the synthetic aperture [5,6]. Of particular interest here, is the phase gradient algorithm (PGA) [7]. The PGA method efficiently provides phase corrections between neighboring cross-range (or azimuth) bins, ultimately allowing an aberration corrected synthetic aperture to be formed. In fact, PGA (or digital shearing) has been previously applied to 2-D digital holography to correct for phase errors across the image [8]. In this paper, we seek corrections for the range, or temporal frequency bins, so that they can be coherently combined. To accomplish this, the concepts of the PGA method could prove useful for estimating the range phase errors present in 3-D holographic ladar.

The research presented here will detail the processing steps necessary for proper phasing of discrete temporal frequency samples in 3-D holographic ladar. In Sec. 2, we briefly describe range resolution and range ambiguity using properties of the DFT, before stepping through the
processing chain in detail in Sec. 3. The different detection geometries of SAL and 3-D holographic ladar lead to differences in data handling and in the application of some of the steps of the PGA method, both of which are discussed in this section.

In Sec. 4, experimental data is presented that clearly demonstrates the benefit of applying our PGA method to the temporal frequency bins of 3-D holographic ladar data. Finally, in Sec. 5, a simple model with a canonical target is simulated to show the theoretical phase variance of the PGA estimator. In particular, we demonstrate saturation at low signal-to-noise ratio (SNR) due to the modulo 2\pi nature of the phase. While this is usually disregarded in the literature, it is important for 3-D holographic ladar since it is advantageous to minimize pulse energy with the expectation of compression gain in the final image.

2. 3-D HOLOGRAPHIC LADAR: RANGE RESOLUTION AND AMBIGUITY

For this paper we assume that we begin with a digital hologram image dataset, collected over some temporal frequency range, with equal spacing \(\Delta\omega\) between the discrete frequencies. Furthermore, the complex image data is assumed to be arranged in increasing temporal frequency order and focused, or sharpened, in both cross-range dimensions. Sharpening eliminates cross-range phase aberrations; what remains is a piston-like phase uncertainty between neighboring frequency plane images. There are many excellent references for digital holography processing, and the method used to obtain the data is immaterial if the above conditions are met [6,7].

Figure 1 shows the general principle of 3-D holographic ladar. By using the relationship between temporal frequency and time, along with a simple coordinate transform, we can create a set of complex valued digital hologram images of sequential temporal frequency, which can be compressed via an IDFT to yield a full 3-D data set.

![3-D Holographic Ladar](image)

Note in Fig. 1 the convention, used throughout this paper, of explicitly treating only temporal frequency \(n\Delta\omega\) and range \(k\Delta z\) as discrete variables. We acknowledge the discrete nature of the sampling sensor elements in the cross-range dimensions but assume them to be of consistent size and response, coherent with respect to each other and of sufficient sampling period to be nearly continuous. Also, for a particular frame, all cross-range samples are simultaneously collected. Temporal frequency, however, is assumed to have an unknown and random phase relationship among the samples. Furthermore, the frequency separation is user selectable.

Now, consider a signal \(S(n\Delta\omega)\) from a single cross-range location \((x_0,y_0)\) of Fig. 1. The relationship between \(S(n\Delta\omega)\) and its time domain signal \(s_0(k\Delta t)\) can be expressed in discrete form by use of the IDFT according to

\[
s_0(k\Delta t) = \frac{1}{N} \sum_{n=-\infty}^{\infty} S_0(n\Delta\omega) \exp\left[ j \frac{2\pi}{N} nk \right],
\]

where \(k\Delta t\) is the discrete temporal variable, \(\Delta\omega\) is the frequency sampling interval and the sum is over \(N\) total frequency samples. Now the temporal spacing of samples \(\Delta t\) is dependent on the total frequency range or bandwidth according to [8]

\[
\Delta t = \frac{2\pi}{N\Delta\omega},
\]

and is related to range by

\[
\Delta z = \frac{c}{2} \Delta t,
\]

where \(c\) is the speed of light and the factor of \(\frac{1}{2}\) is due to the round trip travel of the pulse from the transceiver to the target and back. Eq. (3) is the coordinate transform that, when applied to Eq. (1), produces the explicit relationship between the left and right datasets of Figure 1. Furthermore, inserting Eq. (2) into (3) gives the smallest achievable differential range measurement (range resolution), given by
\[ \Delta z = \frac{c\pi}{N\Delta\omega}. \]  

Note that \( N\Delta\omega/2\pi \) is the total pulse bandwidth \( B \), allowing us to write

\[ \Delta z = \frac{c}{2B}. \]

which is identical to the form commonly found in SAR and SAL literature [9]. Here, Eq. (5) is an exact expression of range resolution arising from the properties of the DFT.

Another interesting property of Eq. (1) is that the IDFT of a signal regularly sampled in frequency results in a periodic function in time. Presuming the sampling locations are relatively closely spaced, the IDFT yields a signal whose ambiguity period \( T_{amb} \) is related to the frequency sample spacing by

\[ T_{amb} = \frac{2\pi}{\Delta\omega}. \]

Any time samples separated by \( T_{amb} \) will have the same value and will therefore be ambiguous. Again, relating time to range, we can find from Eq. (6)

\[ z_{amb} = \frac{c\pi}{\Delta\omega}. \]

where \( z_{amb} \) is the range ambiguity or the unambiguous range of the dataset. This sets a practical limit on the un-aliased scene depth to be imaged for a given frequency separation and is again directly due to properties of the DFT.

So in theory, a set of digital hologram images collected at regularly spaced temporal frequencies can be range compressed with an IDFT to produce a 3-D image. Due to the properties of the DFT, the separation of the frequencies has important impacts on the scene depth, while the total bandwidth affects range resolution. Of course, in the process of collecting actual data, noise with both amplitude and phase is often introduced and simply applying an IDFT may no longer correctly compress the signal. The process to estimate and correct this noise is the subject of the next section.

3. APPLICATION OF PHASE GRADIENT AUTOFOCUS ALGORITHMS

As detailed above, the steps to combine and transform a set of multi-frequency digital hologram images into a 3-D image are straightforward; however, differential phase aberrations across the frequency bins can lead to poor compression. These errors can be induced by temporal atmospheric effects as well as phase wander of the laser illumination source, or platform jitter. To correct the errors, we can exploit existing algorithms designed for SAL. Perhaps the best known of these is the phase gradient auto-focus algorithm. The central assumption in the PGA method is that the field returning from a point target centered in the scene will manifest as a plane wave at the pupil of the sensor; in other words, it will possess a phase which is flat across the entire synthetic aperture. The algorithm requires organizing the complex image data so that the brightest point targets are registered to the central cross-range bin and then windowed so that weaker targets’ contributions are minimized. It is designed to force any aberrations to be revealed in the synthetic aperture plane as deviations from the flat phase just described. Then, a phase correction can be estimated and refined as the algorithm is repeated. We seek to apply the PGA technique to 3-D holographic ladar data, not for correcting cross-range phase aberrations but for phase errors across multiple temporal frequency images. Figure 2 shows the steps required to accomplish this.

![Figure 2: Processing steps](image-url)
3.1 Assemble complex image plane data cube

We begin with a data set arranged as shown in Fig. 1, where a set of cross-range complex images of equal frequency separation are assembled into a complex data cube. Within this dataset, a mean phase aberration between neighboring temporal frequencies exists that is assumed to be spatially invariant over the \((x, y)\) plane. This assumption requires that the temporal frequency steps are small enough that the speckle doesn’t evolve appreciably between neighboring samples. Since all of the cross-range samples of each hologram are simultaneously recorded, each contains a copy of this common mode aberration along with any pixel specific noise inherent to the detection process. The PGA method, although designed for other applications, is well suited for data with these characteristics.

3.2 Apply a cross-range target support mask

This is an optional step. Since each image already contains a 2-D projection of the target whose cross-range extent is often readily delimited, a spatial support filter can be applied by masking around the desired cross-range target information. Note that this will affect information in the range phase dimension. The support filter simply removes known cross-range clutter data from the algorithm.

3.3 Range compress (1-D IDFT)

Step 3 requires that the data cube be range compressed, pixel by pixel, via 1-D IDFT over \(n\Delta\omega\). The data is now entirely in 3-D complex image space.

3.4 Centershift brightest range values

Now, the brightest value in range of each cross-range pixel, assumed to be a high confidence measurement of a point target, is located and centershifted in range to the central range bin. The result of this step yields a data cube whose intensity appears to be a single bright plane (the central range bin) in a 3-D volume of noise and clutter. Centershifting will eliminate phase corresponding to the depth dimension of the target’s 3-D structure. The remaining phase profile of each pixel will then contain registered common mode aberration content along with independent realizations of noise and clutter. With the target depth information removed, we can accurately estimate the aberration. With the many independent pixel realizations available, through averaging we can enhance the accuracy of the aberration estimate while simultaneously suppressing the noise.

3.5 Window in the range dimension

After the centershifting step, above, the other weaker pixels are effectively treated as noise or clutter. An optional window can therefore be applied around the central range plane. Just as with SAL data, it is important not to overly truncate the centered response while minimizing the contributions due to the weaker pixels. Windowing is performed to increase signal to clutter ratio (SCR); however, step 5 is designated optional since in holographic remote sensing applications clutter is often minimal along the range dimension. This occurs because, in most cases, the pulse will be completely reflected (neglecting target absorption effects) upon incidence with a target, leading to one strong return in each cross-range pixel. Also recall the assumption that the data is focused in cross-range. If cross-range defocus is present, target content from different range locations can overlap in cross-range, making it appear as if multiple range bins within a pixel are occupied when in fact only one may contain true target information. In this case, the apparent clutter in the data, due to defocus, can be mitigated by applying a range window. If the data is well focused and SNR is adequate, then the benefit of windowing is minimal. As SNR decreases, windowing again becomes more effective as noise begins to contribute more significantly to the point response.

3.6 Decompress in range (1-D DFT)

For SAL, the preceding steps are designed such that any aberrations will now be manifest across the synthetic aperture as a deviation from an expected flat phase. For 3-D holographic ladar, the phase deviations occur instead across \(n\Delta\omega\). The windowed, centershifted data is decompressed in step 6 via a 1-D DFT, in order to allow us to estimate the phase error in subsequent steps.

3.7 Compute the phase gradient estimate

In step 7, a phase gradient vector is estimated. There are a number of estimation methods (also called kernels) in the literature from which to choose. We selected the maximum likelihood estimator (MLE) \(\psi_{ML}\), which is compactly written as [10]

\[
\hat{\psi}_{ML} = \angle \left( \sum_{x} \sum_{y} S(x, y, n\Delta\omega)S^*(x, y, (n + 1)\Delta\omega) \right),
\]

where the outer operator indicates the angle function and the double sum is over all of the detector array pixels. Equation (8) can be understood as the combination of three steps, the first of which is used to compute an N-element phase difference vector \(S_0\) for each pixel, where the first sample is arbitrarily set to zero. For example, at some pixel \((x_0, y_0)\) a complex vector is computed

\[
S_0 = S(x_0, y_0, n\Delta\omega)S^*(x_0, y_0, (n + 1)\Delta\omega),
\]

where the data from each frequency bin is multiplied by its conjugated neighbor. The phase of this vector contains one realization of an estimate of the phase gradient over temporal frequency. Once this is done for all pixels, we have an ensemble of phase gradient estimates. To minimize the
variance of the uncorrelated noise, an average is then calculated by performing a complex sum over all of the pixels. This is important because holographic ladar, like all coherent processes, is susceptible to speckle and Eq. (8) appropriately weights the phase gradient measurement based on the amplitude of the samples.

Note that since the phase is the quantity of interest, the modulus scaling factor can be omitted from the averaging step of Eq. (8). After correcting the phase of the ensemble average, the final result of step 7 is a single N-element phase gradient vector, where again the first element is arbitrarily set to zero. Lastly, we reiterate that the phase gradient estimated here is over temporal frequency bins, unlike SAL where the gradient is over cross-range aperture bins.

3.8 Integrate the phase gradient to recover the phase aberration estimate

Once the phase gradient vector has been calculated, the temporal frequency phase aberration estimate is retrieved by integrating the result of step 7.

3.9 Detrend

Any piston phase offset and linear phase trend in the phase aberration vector is now removed before its conjugate is applied to the complex image plane data. Failure to remove any residual linear phase trend may result in data that is circularly shifted in range. This has no effect on the algorithm accuracy, and if present can be removed once processing is completed. However, an algorithm exit decision may be based on a correction metric that benefits from detrended phase, leading to unnecessary algorithm iterations if the trend is not removed.

3.10 Apply phase correction to complex image plane data

At this point a single N-element vector containing the phase aberration estimate has been created. The conjugate of this estimate is then applied across every pixel of the data cube created in step 1.

3.11 Corrected?

The decision step implicitly includes a correction metric of the user’s choosing. This can be simple visual scrutiny, be based on residual phase error or be image content driven (e.g., morphological analysis or image entropy minimization). In any case, once it has been determined that the algorithm has sufficiently converged, the loop is exited and the final 3-D holographic ladar image is formed through a final IDFT step.

The steps outlined above will allow conventional PGA estimators to be directly applied to 3-D holographic data. Application as described here will enable range compression when phase aberrations are present. In Sec. 4, the results of applying the algorithm to real laboratory data will be discussed.

4. EXPERIMENT AND RESULTS

Figure 3 shows the experimental setup designed to acquire 3-D holographic data on which our PGA method was then applied for range compression. As shown in Figure 3, an image plane spatial heterodyne configuration was used with an f/40 imaging optic (a) placed in the pupil plane at \( z_i = 0.7 \) m from a 320x256 InGaAs detector array (b) having a 30 μm pixel pitch. A tunable laser (c), at center wavelength \( \lambda_0 = 1.55 \) μm, was split (d) into local oscillator (LO) and transmitter (TX) fiber optic paths. The TX path was first directed into an (e) erbium doped fiber amplifier (EDFA) and, along the way, was tapped for wavelength monitoring (f). The EDFA output fiber was terminated at the pupil plane, where the beam then propagated in free space to the target (\( z_o = 1.75 \) m). The LO was also inserted as a bare fiber tip at the pupil plane and directed back to the detector. There it mixed with the image of the target on the focal plane, where the resulting intensity of the interference pattern was then recorded. The LO power was adjusted to ensure shot noise limited performance.
Fig. 3: Experimental setup.

The target (g), a pair of overlapped and reversed pennies, was mounted to a rotation stage. The rotation stage was used to acquire data at multiple aspect angles for purposes of speckle averaging. For this arrangement, small rotations of approximately 3 mrad were sufficient to provide uncorrelated speckle realizations. Complete datasets were collected and processed at each rotation. Then the 3-D intensities were registered and incoherently combined (i.e., added together) to mitigate speckle.

The imaging optic had a clear aperture diameter of $d_c = 11.43$ mm giving a theoretical cross-range resolution of approximately $\frac{\lambda z_o}{d_c} = 237$ μm. The laser was tuned over 4.77 THz in 159 steps of 30 GHz separation. Using Eqs. (5) and (7), this gives a range resolution and range ambiguity of 31.6 μm and 5 mm, respectively. In its fastest (lowest precision) mode, the laser required up to 45 seconds to tune between consecutive frequencies, leading to total data collection times of over 2 hours. The long tuning time guaranteed loss of phase coherency across the multiple holograms, and no attempt was made to track the phase migration.

Fig. 4: Digital hologram image. To create this image, the intensities of 20 independent speckle realizations were averaged.
Figure 4 shows a speckle averaged intensity image from the experimental data. It is the result of incoherently combining multiple data planes from Fig. 2, step 1. This image can be obtained through any digital holography processing technique and contains no range information. Since the penny has a homogenous material surface, any contrast is due to surface normal variations, or shadows from off-axis illumination. For near on-axis illumination, as is used in this experiment, most shadows are eliminated, leading to the low contrast seen in Fig. 4. Finally, note that as described in Fig. 2, step 2, the target boundaries are readily apparent, allowing a mask to be designed for implementation during the PGA algorithm.

Before attempting to correct any phase errors, a 3-D image was formed prior to applying steps 3-12 of the PGA technique of Fig. 2. As described in Sec. 3.1, the individual hologram images were first assembled into an ascending frequency data cube. After applying the mask of step 2, the data was range compressed via a 1-D IDFT. Figure 5 shows the result for a single speckle realization. To form this range image, the range bin containing the maximum intensity was located for each pixel. Range bin number is then displayed as a grayscale value, where lighter pixels represent range values nearer to observer while dark represents farther values. Effectively, Figure 5 is a first surface 2-D projection of the 3-D data. Note that while two disk surfaces do appear, range details on the order of the expected resolution of 31.6 μm are not visible. Also note that the solid black area is the cross-range target support mask of Sec. 3.2.

Next, the data was fully processed using the algorithm of Sec. 3. As already demonstrated in Figs. 4 and 5, a mask was applied to remove the empty cross-range areas (Sec. 3.2). The data was then range compressed (Sec. 3.3) and the brightest targets of each pixel were shifted to the central range bin (Sec. 3.4). Since the target is essentially a continuous opaque surface, most pixels will contain a single range return. Therefore, the range window of Sec. 3.5 is not required to remove weaker range clutter. The centershifted data was then decompressed in range (Sec. 3.6). Using Eq. (8), the phase gradient estimator was computed (Sec. 3.7) and then integrated (Sec. 3.8), yielding a single 159-element vector of the estimated temporal frequency phase aberration. The built-in Matlab™ function detrend.m was used to remove piston and tilt from the phase aberration vector (Sec. 3.9). The vector was then conjugated and applied to every cross-range pixel (Sec. 3.10). To determine if the data was sufficiently corrected (Sec. 3.11), a simple visual examination of the range compressed 3-D data was performed. We observed that excellent results were obtained after only two iterations of the algorithm were executed.

The results are shown in Fig. 6, where again, these are range (not intensity) images. Figure 6 shows a single range compressed speckle realization on the top. Even with speckle effects, high resolution information in both range and cross-range is discernable. Not only does our PGA technique dramatically improve range compression, but it is also clear that some form of phase correction is required to compress the holographic data. In the bottom image of Fig. 6, 20 speckle realizations were incoherently averaged by performing our PGA method on each dataset and then adding the phase-corrected intensities. Note that to enhance the contrast in these grayscale images, the range ambiguity was artificially reduced so that the feature range extent of each coin nearly filled the dynamic range of the color-map; i.e., the relative range values between the two coins’ surfaces are now ambiguous, while the relationship among range values from the same coin surface is preserved. This was done only for display purposes.
The apparent contrast of Fig. 6 is entirely due to differential range values across pixels and is not at all dependent on reflectance differences. This suggests that range compression of holographic ladar is useful for extracting detail from low contrast targets, even before fully exploiting the 3-D nature of the data.

Next, in Fig. 7, the data was rotated to highlight the surface relief and the targets’ separation in range. The image was generated using the `surfl` routine in Matlab™. Range compression reveals fine details that are highlighted at this angle, including the buttresses and tripods flanking the stairs of the Lincoln Memorial, and the carved scroll on the memorial’s cornice.
To confirm the range resolution achieved by applying our PGA technique, the ideal point response (IPR) was estimated, as shown in Fig. 8. To do so, a known flat area of the target was identified, in this case, the region below the memorial steps. This was done to ensure only one range bin of the data would be occupied. Furthermore, to ensure maximum signal, a single pixel from that region containing a strong return, or bright speckle, was then isolated. Finally, the intensity, as a function of range, was extracted from this pixel to estimate the IPR. As is evident in the left plot of Fig. 8, no significant range compression was accomplished with the uncorrected data, while the IPR peak is prominent in the center plot when PGA corrections were applied. The theoretical IPR is shown with the data estimate in the right-most plot where the estimate of the range resolution achieved was approximately 35 μm. This was calculated from the 3 dB full width of the peak and is in good agreement with theory. The effects of the strong sidelobes manifested in the data as relatively bright ghosting artifacts within the ambiguity range. Fortunately, these ghost images were easily identified and removed. To do so, the fully corrected 3-D intensity data was rotated and viewed from the side, where the main target image and the fainter ghost images were readily apparent. A range window was then applied to remove the unwanted images.

5. PGA PERFORMANCE FOR 3-D HOLOGRAPHIC DATA

To determine the theoretical performance of our PGA technique for 3-D holographic ladar, a Monte-Carlo simulation was created. Twenty-eight SNR values were selected and, for each value, a dataset was created and corrected as follows. A uniform amplitude planar target of 128x128 pixels was created and located in the central range bin of a 3-D volume. A Gaussian phase representing a surface roughness on the order of the nominal
wavelength (λ₀ = 1.55 μm) was then applied. The complex field of the target was then propagated by a FFT to the sensor plane at 64 equally spaced frequencies (Δν = Δω/2π = 75.7 GHz). This led to an unambiguous 3-D volume of 2 mm with a range resolution of 31.6 μm. The resolution matches our experiment, while the number of frequencies and their spacing was changed for reasons discussed below. The complex field at the sensor plane was then cropped to a pupil size of 22x22 pixels (again, to be explained shortly), simulating the physical process of a finite aperture sampling a field with a much larger spatial extent. When implemented properly, this lowpass filtering step ensures that realistic speckle statistics are realized [11]. At each frequency, a single random phase value ψ(nΔω) was then added to the entire field, where the p subscript signifies prescribed. This number simulated an unknown phase error for that frequency. In addition, to simulate shot noise, a circularly complex Gaussian random number was added to the field (as described in [11]). The noise power was in turn scaled according the SNR value of the iteration. Our PGA technique, as defined in Sec. 3, was then performed. Since the model contained no clutter and the target was confined to the central range bin, neither centershifting nor windowing was necessary. Therefore, only one iteration of the algorithm was required. The residual phase gradient mean square error (MSE) ̂ of the corrected field was then calculated according to

\[ ̂ = \left( \left( \nabla_p (nΔω) - \nabla_{\text{ph}}(nΔω) \right)^2 \right), \]  

(10)

where \( \nabla_p \) and \( \nabla_{\text{ph}} \) are the gradients of the prescribed phase and the phase correction estimate given by the PGA estimator, respectively. Finally, 50 trials were run over which a new random draw of the speckle, noise and prescribed phase were implemented, after which the MSE was calculated and stored. This process was repeated for each SNR value. The resulting average MSE \( \hat{\epsilon} \), as a function of SNR, is plotted in Fig. 9.

![Fig. 9: The CRLB, mean square phase error, and theoretical phase error for random phasor sum plus a constant phasor.](image)

The phase gradient estimator of Eq. (8) can be shown to be a maximum likelihood estimator [10,12,13]. A MLE is asymptotically efficient, meaning that for sufficiently large data records, its MSE approaches the theoretical minimum, known as the Cramer-Rao lower bound (CRLB) [14]. The CRLB for the gradient between two neighboring phasors of the model described here, in terms of SNR and number of image pixels \( L_{\text{pix}} \), has been shown to be [10,15,16]

\[ \text{CRLB} = \frac{1 + 2\text{SNR}}{2L_{\text{pix}} \cdot \text{SNR}^2}. \]  

(11)

Figure 9 shows the CRLB for \( L_{\text{pix}} = 484 \) corresponding to an image dimension of 22x22 pixels. The number of pixels as well as the number of frequencies was chosen so that CRLB and \( \epsilon \) plots in the central portion of Fig. 9 are virtually identical to those in Fig. 2 of [10]. As is evident in the figure, the mean square residual error \( \epsilon \) approaches the CRLB. Although not shown in Fig. 9, \( \epsilon \) will diverge from the CRLB at higher SNR. This phenomenon is due to speckle saturation and has been well documented [15]. As SNR decreases, the estimator can be seen to diverge a bit from the CRLB, first by producing slightly higher MSE and then by asymptotically approaching 5.17 dB.

As alluded to in Sec. 1, it may be advantageous to operate 3-D holographic ladar systems in the low SNR regime where the algorithm begins to degrade in performance, as shown in the upper left of Fig. 9. To gain greater insight into the algorithm’s behavior there, we propose another method of analyzing the phase statistics of our model. The goal of applying PGA techniques to our data is to estimate some phase that is constant over all pixels of the same frequency which must be computed from a measurement of the vector sum of the desired phasor and a noise phasor. The sum of Eq. (8) over the pixels allows us to invoke the Central Limit Theorem and declare a Gaussian distribution for the noise, independent of the actual underlying distributions. Alternately, we can show that in a shot noise limited scenario (a goal of all well designed holography systems), the Poisson statistics of the shot noise are well approximated by a Gaussian distribution. So, the challenge for the algorithm is to estimate a constant signal phasor in the presence of Gaussian noise. Goodman describes exactly this scenario in [11], calling it “a random phasor sum plus a known phasor.” He gives the marginal probability density function for the estimator phase as
\[ p_\theta(\theta) = \frac{e^{-\frac{1}{2\beta^2}}}{2\pi} + \frac{1}{2\pi\beta} \left( e^{-\frac{\sin^2 \theta}{2\beta^2}} \right) \frac{1 + \text{erf} \left( \frac{\cos \theta}{\sqrt{2\beta}} \right)}{2} \cos \theta, \]  

(12)

where \( \beta \) is the estimator's SNR and the error function is defined as

\[ \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \]  

(13)

To better define \( \beta \), recall that Eq. (11) defines the minimum MSE and, for unbiased estimators, the MSE and variance of the estimator are equal. This suggests that the inverse of the CRLB is a normalized SNR for \( \psi_{ML} \). Furthermore, for a given detection SNR, it is the maximum achievable estimator SNR. Accordingly, we define

\[ \beta = \frac{1}{\text{CRLB}}. \]  

(14)

Finally, using Eq (12), the phase variance can be calculated according to

\[ \sigma_\theta^2 = \int_\pi^\pi \theta^2 p_\theta(\theta) d\theta. \]  

(15)

Equation 15 is also plotted in Fig. 9 and is seen to accurately predict the behavior of \( \hat{\epsilon} \) in the low SNR regime. The saturation of the average MSE occurs as the angle of the resultant phasor becomes dominated by the contributions of the noise phasor. The phase distribution converges to a uniform distribution over \((-\pi, \pi]\) which has a variance of 5.17dB. Since the phase is modulo 2\( \pi \), this represents a state of maximum entropy. We therefore call this behavior the maximum entropy saturation of the estimator. The plots of \( \hat{\epsilon} \) and Eq. (15) can both be seen asymptotically approaching this value in Fig. 9. This suggests that operation at detection SNR values lower than -10dB, for the modeled parameters, is intractable. That is, the range phase aberration cannot be computed. Equation (11) does suggest that a larger detector array, i.e., increasing \( L_{\text{pix}} \) will lower the CRLB thereby improving estimation performance. To realize any further improvements, an estimator other than the one described here must be found.

For clarity, we have described the algorithm and its performance using a conceptually simple form of the phase gradient estimator. As mentioned earlier, we have chosen a phase gradient MLE based only upon consideration of nearest neighbor phase differences. However, a MLE for the phase differences between temporal frequency images, of arbitrary separation, can be found by considering the covariance matrix of the data resulting from the range decompression step (step 6) of Fig. 2 [10]. Given \( N \) frequency images, the \( N \times N \), covariance matrix is computed and its principle (i.e., largest) eigenvalue is determined. The phase of the corresponding eigenvector is the estimate of the temporal frequency image phase differences we seek. In practice, implementation of these steps is straightforward; we simply use the Matlab™ functions cov.m and eigs.m to calculate the covariance matrix and eigenvectors, respectively. The above steps are then inserted into the algorithm in lieu of steps 7-8 of Fig. 2 and offer the advantage of now allowing more than just adjacent phasor pairs to be considered when calculating the phase aberration correction. For this estimator, the CRLB is generalized to

\[ \text{CRLB} = \frac{1 + M \times \text{SNR}}{M \times L_{\text{pix}} \times \text{SNR}^2}, \]  

(16)

where \((M-1)\Delta\omega\) is the maximum temporal frequency image separation considered in the calculation of the covariance matrix. This is implemented by calculating the full covariance matrix and then setting all values of the diagonals exceeding \((M-1)\Delta\omega\) to zero. Note that \( M = 2 \) in Eq (11) since only the covariance (phase difference) of nearest neighbor samples was considered.

To evaluate the performance of this estimator in our algorithm, the Monte-Carlo simulation was again performed using the eigenvector phase difference estimation method; all other parameters were unchanged. As shown in Fig. 10, and predicted by Eq. (16), when the parameter \( M \) is increased, the CRLB is decreased. As with Eq (11), Eq. (16) shows that the CRLB can be decreased by increasing \( L_{\text{pix}} \). 3-D holographic datasets are often much larger than those of 2-D SAR making large \( L_{\text{pix}} \) and \( M \) values easy to achieve. As shown in Fig. 10, the eigenvector method is much more robust in low SNR environments. Finally, we have observed that increasing \( M \) and \( L_{\text{pix}} \) reduces the number of iterations required; however, the computation load per iteration is often increased.
6. CONCLUSION

Range compressed holographic ladar is sensitive to phase aberrations distributed over temporal frequency. To address this, we have described a novel application of the PGA technique to range compressed holographic ladar. Holographic images formed before range compression provide the opportunity to apply a cross-range target support mask. With reasonable detection SNRs, it is not usually necessary to window the strongest range returns because of the limited number of scatterers in that dimension. While the PGA method is an iterative process, typically only a few iterations are required for range compressed holographic ladar.

In a laboratory experiment, we demonstrated a significant improvement in range compression when the modified PGA steps described were applied to the data. In fact, for this configuration, where discrete temporal frequencies with unknown phase offsets are used, it was determined that a phase correction algorithm was essential for range compression. To help quantify the performance, the IPRs for the uncorrected and PGA corrected data were estimated and improvement of the IPR is clearly shown for the latter case.

With a numerical model and a canonical target, the estimator was demonstrated to be asymptotically efficient for this application. Additionally, the maximum entropy saturation behavior of the estimator was analytically described. This has implications for 3-D holographic ladar operation at low SNR and a possible solution was proposed with the use of a more sophisticated and generalized form of the estimator.

As clearly demonstrated in Fig. 9, the modified PGA method using the MLE of Eq. (8) yields optimal results for phase aberration correction along the temporal frequency bins of 3-D holographic ladar. The maximum entropy saturation of the estimator at low SNR can also be accounted for analytically. For improved performance in low SNR environments a more efficient phase estimation kernel such as the eigenvector method can be utilized, as demonstrated in Fig. 10.

For range compressed holographic ladar, phase correction of aberrations across temporal frequency is critical. Whether due to the discrete stepped frequency waveform imposed on the laser or to atmospheric effects, the stochastic nature of the aberration lends itself to estimation methods already developed for SAR. We have presented a modified version of the PGA method, tuned specifically for the data format of holographic ladar, which leverages some of its unique aspects and rapidly converges to an accurate solution that enables excellent range compression performance.

This effort was supported in part by the U.S. Air Force, through contract number FA8650-12-D-1377, and by the University Of Dayton Office Of Graduate, Professional & Continuing Education. The views expressed in this article are those of the authors and do not reflect the official policy of the Air Force, Department of Defense or the U.S. Government.

References