24th International Conference Image and Vision Computing New Zealand (IVCNZ 2009)

Undue Influence: Mitigating Range-Intensity Coupling in AMCW 'Flash' Lidar Using Scene Texture

J. P. Godbaz*, M. J. Cree and A. A. Dorrington
University of Waikato
Hamilton, New Zealand
*jpg7@waikato.ac.nz

***Abstract*—We present a new algorithm for mitigating range-intensity coupling caused by scattered light in full-field amplitude modulated continuous wave lidar systems using scene texture. Full-field Lidar works using the time-of-flight principle to measure the range to thousands of points in a scene simultaneously. Mixed pixel are erroneous range measurements caused by pixels integrating light from more than one object at a time. Conventional optics suffer from internal reflections and light scattering which can result in every pixel being mixed with scattered light. This causes erroneous range measurements and range-intensity coupling. By measuring how range changes with intensity over local regions it is possible to determine the phase and intensity of the scattered light without the complex calibration inherent in deconvolution based restoration. The new method is shown to produce a substantial improvement in range image quality. An additional range from texture method is demonstrated which is resistant to scattered light. Variations of the algorithms are tested with and without segmentation – the variant without segmentation is faster, but causes erroneous ranges around the edges of objects which are not present in the segmented algorithm.**

I. Introduction

Full-field amplitude modulated continuous wave (AMCW) lidar systems use the time-of-flight (TOF) principle to measure the range to a large number of points in a scene simultaneously. Whereas a normal camera captures an image where each pixel has an RGB colour value, a range image camera produces 2D matrices of radial distance values. These measurements can be converted to 3D Cartesian coordinates and used in applications like process control and robot navigation. A major unsolved problem in AMCW imaging is the mixed pixel problem, occurring when a pixel integrates range measurements from two or more different objects within a single pixel, causing erroneous range values. In order for full-field AMCW lidar to be a high quality replacement for traditional techniques such as line scanning triangulation systems and photogrammetry, mitigation techniques need to be developed.

A. Previous Work

Mixed pixels are most common around the edges of objects. Previous methods have included median filtering and normal angle thresholding techniques to enable removal of mixed points [1]–[4]. A more recent method developed by Larkins [5] uses clustering and parametric surface fitting to project mixed pixel back onto their parent surfaces; others use beat waveform harmonics to identify the component returns [6]–[7]. More insideous, however, are mixed pixels caused by scattered light. Fig 1(a) and 1(b) show an initial scene that suffers from only limited light scattering. If a much brighter object is moved into the scene (see figs. 1(c) and 1(d)), not only does the bright object now appear in the scene, but the ranges to the darker objects in the scene are changed due to scattered light. In particular, objects at the same distance from the camera with different brightnesses appear at different ranges – this is known as range-intensity coupling. Deconvolution has been applied to this problem [8], but this is unable to handle light scattered from outside the field-of-view (FOV) and requires complex calibration and processing to fully compensate for the spatially variant light scattering.

B. Overview of New Mitigation Algorithm

We present a new method for mitigating range-intensity coupling due to scattered light in a full-field AMCW lidar system. The method does not require prior calibration of the ranging system and can migitate scattered light that is uncorrelated with the scene. The raw range data values corresponding to a particular object can be modelled as a linear function of reflectivity over small regions. By segmenting out each object and determining this linear function, it is possible to identify the true range to an object irrespective of any scattered light. By combining the linear functions corresponding to regions of several objects at different ranges, it is possible to create local estimates of light scattering. We show that subtracting these local estimates produces a substantial improvement in image quality.

The authors recommend printing in colour or viewing on a computer screen to ensure the clarity of the images in this paper.

II. Background

A. Conventions/Definitions

Arg( ) is the complex argument function and returns the phase of a complex number, $\Re(x)$ is the real part of a variable x and $\Im(x)$ is the imaginary part. The symbol $\star$ represents a 2D convolution, $\star_{sv}$ is a 2D spatially variant convolution, $\bullet$ is the Hadamard product (elementwise multiplication) and x^n, where x is a vector, is the Hadamard product of n xs.

Square brackets, [ ], indicate a discretely valued parameter and parentheses, ( ), indicate a continuously valued parameter.

B. AMCW Range-Imagers

AMCW range-imagers illuminate a scene with modulated light. Light takes longer to travel to objects farther away from the camera; this results in a phase shift in the modulated signal that is proportional to range. The range-imager measures this phase shift by correlating the illumination signal with a reference signal at the same or similar frequency. In the case of the University of Waikato Range-Imager [10] the light is gain mixed with the reference signal using a modulated image intensifier. By using high frequency modulation signals that differ only very slightly, a technique known as heterodyning, the high frequency phase measurement problem is reduced to a low frequency phase measurement problem. For an illumination modulation frequency, f_1, and reference signal frequency, f_0, the output beat waveform is at a frequency of $|f_1 - f_0|$. The phase offset of the beat waveform is the same as the phase offset of the returned light, thus proportional to the range of an object from the camera. The phase of the beat waveform is measured by Fourier transforming a sequence of images over time by an off the shelf charge coupled device (CCD) camera which is optically coupled to the rear of the image intensifier. While as few as three frames are sufficient to allow range to be measured, a larger number of frames prevents aliasing of harmonics above the Nyquist frequency and reduces the impact of noise.

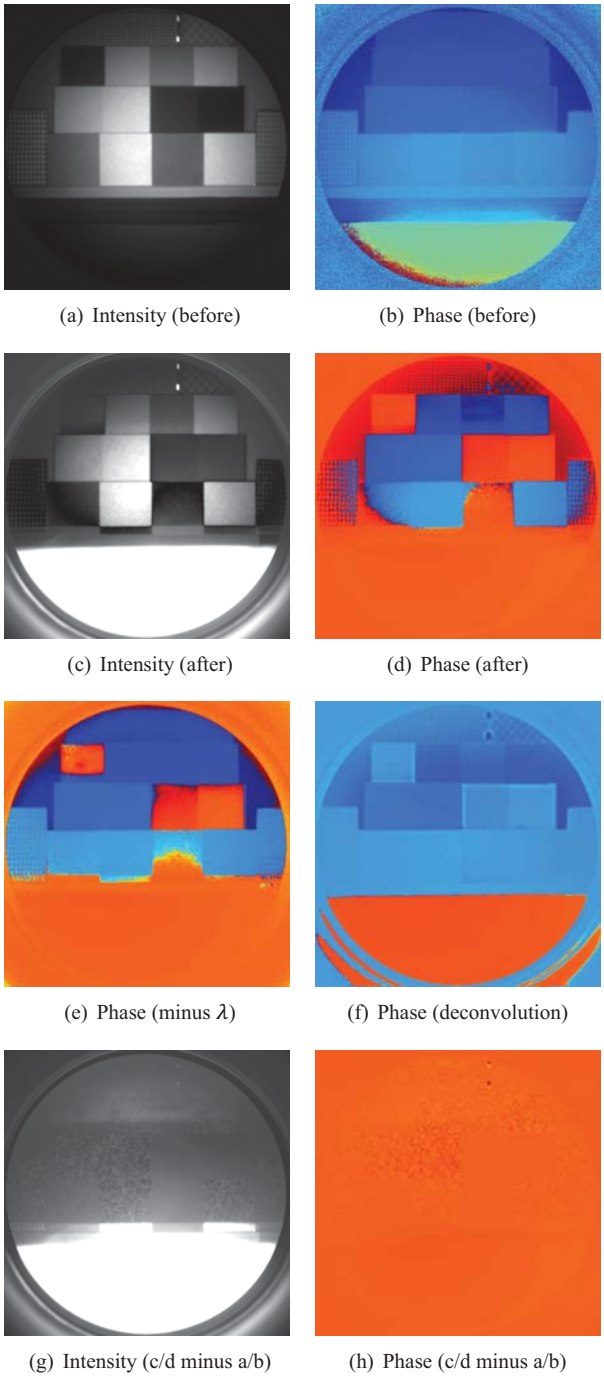
The results presented in this paper were generated at frame rates of either 16 or 32 frames-per-second, two phase cycles per acquisiton, 0.5 Hz beat frequency. The laser modulation frequency was 40 MHz, which results in a range ambiguity distance of 3.75 metres. Having sets of optics both before and after the image intensifier makes the system particularly susceptable to scattered light; in particular, most of the light scattering in this system occurs post-intensifier. The image intensifier uses the photoelectric effect to generate photoelectrons which are then used to illuminate a phosphor screen. Because the light from this phosphor screen is not collimated, it reflects off the walls of the optical cavity. Depending on scene layout anywhere between 2% (typical) and 90% (extreme) of the total light measured at any one pixel is scattered from somewhere else in the scene. This is illustrated well by fig. 2, which shows that in addition to specific specular reflections within the lens system, there is a diffuse scattered background.

C. Modelling Range Images

We henceforth distinguish between component returns and actual range measurements. A model of signal return intensity versus range, $f(r)$, can be considered to be a sparse spike train, composed of N individual component returns

$$f(r) = \sum_{i=0}^{N-1} a_i \delta(r - r_i), \qquad (1)$$

where δ is the Dirac delta function, and a_i and r_i, are the amplitude and range of the ith component return, respectively. A well



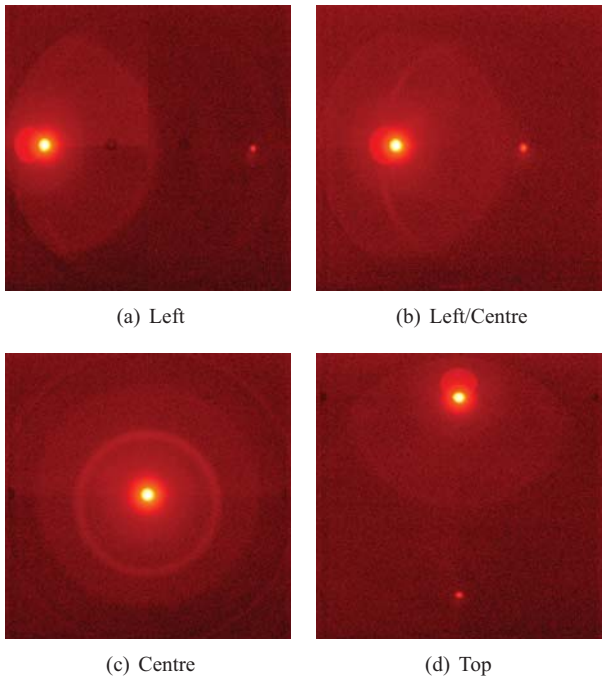
(a) Intensity (before) (b) Phase (before)

(c) Intensity (after) (d) Phase (after)

(e) Phase (minus λ) (f) Phase (deconvolution)

(g) Intensity (c/d minus a/b) (h) Phase (c/d minus a/b)

Figure 1: Real range images of a scene before and after the addition of a bright box which scatters light across the entire image and changes the range to objects in the scene. Dark red corresponds to a phase of 0 and dark blue to a phase of 2π – however mixed pixels result in erroneous ranges. 1(e) and 1(f) show the results of two different restoration methods, while 1(g) and 1(h) show the intensity and phase of the scattered light – determined by subtracting the first range image from the second.



(a) Left (b) Left/Centre

(c) Centre (d) Top

Figure 2: PSF sampled at different pixel locations. The PSF is highly anisoplanatic, varying in a rotationally symmetric manner. Due to the high dynamic range, the logarithm of the intensity is plotted here, black/red is dark and white bright. There is a reflection that 'follows' the light source around, always appearing on the opposite side of the principal point.

calibrated AMCW capture is equivalent to sampling $F(u)$, the Fourier transform of f, at a spatial frequency of $-c/2f_1$, where c is the speed of light. Thus a range measurement ϡ is formed by

$$\text{ϡ} = F\left(-\frac{c}{2f_1}\right) = \sum_{i=0}^{N-1} \text{ϙ}_i = \sum_{i=0}^{N-1} a_i \mathrm{e}^{\theta_i}, \tag{2}$$

where a single component return is represented as a phasor $\text{ϙ}_i \in \mathbb{C}$. Each component return is composed of an active intensity a_i and a phase $\theta_i = 4\pi r_i f_1/c$. If we make the assumption that in the ideal, unblurred, unscattered image each pixel only contains one component return, then we can model the light scattering process as a spatially variant convolution. In actuality this is a naive assumption because light can be scattered onto the scene from bright objects outside the FOV, and mixed pixels occur frequently around the edges of objects where each pixel integrates light from two or more objects. For the ideal image, $\text{Ϙ} \in \mathbb{C}^m$, where m is the number of pixels, the recorded data are formed by

$$\text{ϡ} = \text{Ϙ} \star_{sv} h + \epsilon, \tag{3}$$

where h is the point spread function (PSF) and ϵ is noise.

D. Range-Intensity Coupling

If we have a large object in the scene that has a textured/patterned surface and also have significant light scattering, then we can model the measured phase of the object as a function of the surface brightness,

$$\theta = \arg(\text{ϡ}(\rho)) = \arg(\rho\vec{\text{ϙ}} + \lambda), = \arg(\text{ϙ} + \lambda) \tag{4}$$

where $\rho \in \mathbb{R}^+$ is the brightness, $\lambda \in \mathbb{C}$ is the scattered light, which we assume to be near constant in the region of the object, and $\vec{\text{ϙ}} \in \mathbb{C}$ is a unit phasor corresponding to the underlying range.

There are two approaches to determining ϙ, one is to assume that λ is solely a function of ϙ and apply deconvolution techniques to determine ϙ directly; the other is to determine λ separately and then subtract it from ϡ.

E. Weighted Least Squares Linear Fitting with Windowing

The new algorithm detailed in this paper is based around least squares linear fitting across an entire image in the Fourier domain. For a model $\alpha + \beta x$, this works by finding the values of α and β such that

$$(\alpha_i, \beta_i) = \arg\min_{(\alpha_i,\beta_i)} (y - \alpha_i - \beta_i x)^2 \bullet \omega \bullet \eta[i] \tag{5}$$

for each pixel i, where x and y are vectors corresponding to the independent and dependent variables, ω is data weighing and η is a windowing function. In closed form using convolutions the solution is

$$\Lambda(z) = (z \bullet \omega) \star \eta, \tag{6}$$

$$\Lambda_\emptyset = \omega \star \eta, \tag{7}$$

$$\Upsilon_\alpha(x, y, \omega, \eta) = \frac{\Lambda(x^2) \bullet \Lambda(y) - \Lambda(x \bullet y) \bullet \Lambda(x)}{\Lambda_\emptyset \bullet \Lambda(x^2) - \Lambda(x)^2}, \tag{8}$$

$$\Upsilon_\beta(x, y, \omega, \eta) = \frac{\Lambda_\emptyset \bullet \Lambda(x \bullet y) - \Lambda(y) \bullet \Lambda(x)}{\Lambda_\emptyset \bullet \Lambda(x^2) - \Lambda(x)^2}, \tag{9}$$

where $\Lambda(z)$ is a weighted convolution of a data element with the windowing function and $\Lambda_\emptyset$ is a convolution of the weights with the windowing function. Least squares fitting is notated by two functions: $\Upsilon_\alpha(x, y, \omega, \eta)$ which returns the value of α and $\Upsilon_\beta(x, y, \omega, \eta)$ returns the value of β.

F. Segmentation of Range-Images

Range images can be segmented by identifying contiguous regions larger than a certain size, where all pixels are connected to each other via at least one path where the Δ phase between adjacent connected pixels is less than a certain threshold. In this paper we use a deterministic region-growing method, chosen primarily for ease of implementation. The method assigns each pixel an initial unique region code and then progressively merges regions together. At each iteration, the connected neighbours of each pixel are scanned and the current pixel is set to the minimal region code. Over time any region changes propagate over the image, eventually converging. Any regions of fewer than 100 pixels are discarded.

III. Algorithms

A. Phase From Texture

We can rewrite eqn. 4 as

$$\Im(\text{ϡ}) = \tan(\theta + \pi N)(\rho - \Re(\lambda)) + \Im(\lambda), \tag{10}$$

$$= \alpha + \beta\Re(\text{ϡ}), \tag{11}$$

where $N \in \mathbb{Z}$ is initially unknown. If we assume λ to be constant over a region of the scene, then we can fit a linear model to $\Im(\lambda)$ versus $\Re(\lambda)$ using a least squares approach. In reality, the scattering PSF has a significant local component (see fig. 2), but an assumption of large scale smoothness is necessary to allow estimation of λ. After segmenting the range image into separate contiguous surfaces as described in section II-F, we apply eqn. 8 and 9 to find

$$\alpha = \sum_{i=0}^{m-1} \Upsilon_\alpha(\Re(\lambda), \Im(\lambda), \omega \bullet \zeta[i], \eta) \bullet \zeta[i], \quad (12)$$

$$\beta = \sum_{i=0}^{m-1} \Upsilon_\beta(\Re(\lambda), \Im(\lambda), \omega \bullet \zeta[i], \eta) \bullet \zeta[i], \quad (13)$$

where m is the number of range-image segments, $\zeta[i]$ is a mask corresponding to the ith range-image segment, η is a Gaussian windowing function with $\sigma = 4$ pixels and ω are data weights. In more advanced experiments the data weights could be calculated based on known noise statistics, but for the purposes of this paper, we assign values of 1 to regions within the image intensifier field of view, and 0 to regions outside.

The total integrated intensity for a pixel, is given by

$$\chi_i = \int f_i(r)dr, \quad (14)$$

which consists of a combination of active intensity and any background light. We can use information about the total integrated intensity to disambiguate the possible values of θ (remove N from eqn. 10). This ambiguity occurs because it is impossible to determine from a line's slope which direction it points in. Since the highest integrated intensity values tend to be correlated with greatest possible values of ρ, we can use the covariance of χ with $\Re(\lambda)$ to disambiguate θ for each pixel in the image by

$$\kappa = \mathrm{cov}(\chi, \Re(\lambda), \eta), \quad (15)$$

$$\hat{\theta} = \arctan(\beta \bullet \kappa, \kappa), \quad (16)$$

where $\mathrm{cov}(x, y, \eta)$ is the covariance operator over the data vectors x and y with the same windowing function (η) as above. The estimates of θ returned by this method are henceforth referred to as range/phase from texture.

B. Estimating Light Scattering From Object Texture

In order to find λ, we combine samples of α and β at multiple different ranges. For example, in fig. 1, there are arbitrarily placed patterned regions at the left, top and right of the two scenes. Although relatively dark in the intensity images, the patterned regions are clearly visible in 1(d), where the scattering has resulted in range-intensity coupling. We now utilise these patterned regions to identify the nature of the scattered light. In fig. 3 $\Im(\lambda)$ is plotted versus $\Re(\lambda)$ for two regions at different ranges. Since regions at different ranges result in linearly independent relationships, λ is the phasor corresponding to the intersection of the fits. In this case the intersection of the two fits is at $\lambda = 1084 + 3091j$. The patterned regions are widely

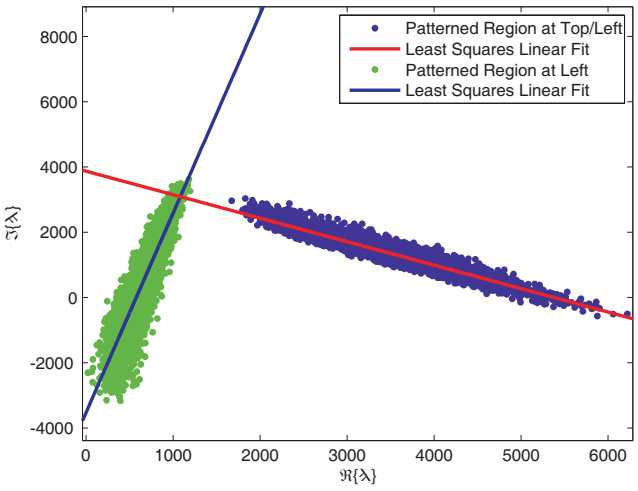


Figure 3: This figure shows the linear relationship between $\Re(\lambda)$ and $\Im(\lambda)$. Combining linear fits from two textured regions at different ranges allows λ to be estimated. In this case, the estimate of λ is based on inconsistant data – the intersection of the two fits requires $\rho < 0$ for some data points, which is physically impossible.

spaced, which is adequate in cases where the scattering is largely constant across the entire image, but this does not always hold due to the local contributions of the scattering PSF. Fig. 3 illustrates the limitations of estimates based on widely spaced samples because $\lambda = 1084 + 3091j$ implies that $\rho < 0$ for some data points, which is physically impossible. In order to avoid the limitations of global estimates, we attempt to generate local estimates of λ over a windowed region. Under the assumption of sufficient patterned sufaces across the entire field of view, then the problem of finding $\hat{\lambda}$ can be reduced to a least squares linear fit, viz

$$\hat{\lambda} = \Upsilon_\beta(\beta, \alpha, \omega, \eta_l) + j\Upsilon_\alpha(\beta, \alpha, \omega, \eta_l), \quad (17)$$

where η_l is a Gaussian windowing function, $\sigma \geq 150$ pixels for our experiments, depending on the amount of useful information in the scene. An estimate of the unscattered data, $\hat{\Theta}$, can now be calculated from λ by subtracting $\hat{\lambda}$.

C. Deconvolution

An alternative approach to removing scattered light is to apply deconvolution techniques. The PSF for our ranger system is shown in fig. 2 – note that the PSF is highly spatially variant, albeit the variance is circularly symmetric in behaviour. There are very noticible lens reflections and halos. In order to achieve a high signal-to-noise ratio (SNR), the camera was deliberately defocussed – the PSF is not only pixel location variant, but also based on range and focal parameters. Attempting to deconvolve a range-image directly using the defocussed PSF resulted in severe ringing effects. A PSF model was generated by least squares fitting the sum of three concentric, centred weighted pillbox PSFs and one Gaussian to the centre PSF sample. This required operator intervention in order to find good solutions.

Subsequently the central pillbox and Gaussian were replaced by a point source to remove the defocus and reduce ringing. Rather than the method used by [8], we implemented a spatial derivative regularised, isoplanatic deconvolution similar to that used in the positive real domain by [9], that is

$$\hat{\Omega} = \arg\min_{x}(||\beth - x \star h||_2^2 + \tau||lx||_2^2), \qquad (18)$$

where l is a Laplacian filter, τ a regularisation constant and $\hat{\Omega}$ the recovered image. This method makes the assumption that the spatial derivatives are Gaussian distributed. The solution was implemented as direct inversion of the normal equations using 2D Fourier transforms. This requires accurate calibration of the PSF, and cannot handle light scattered from outside the image.

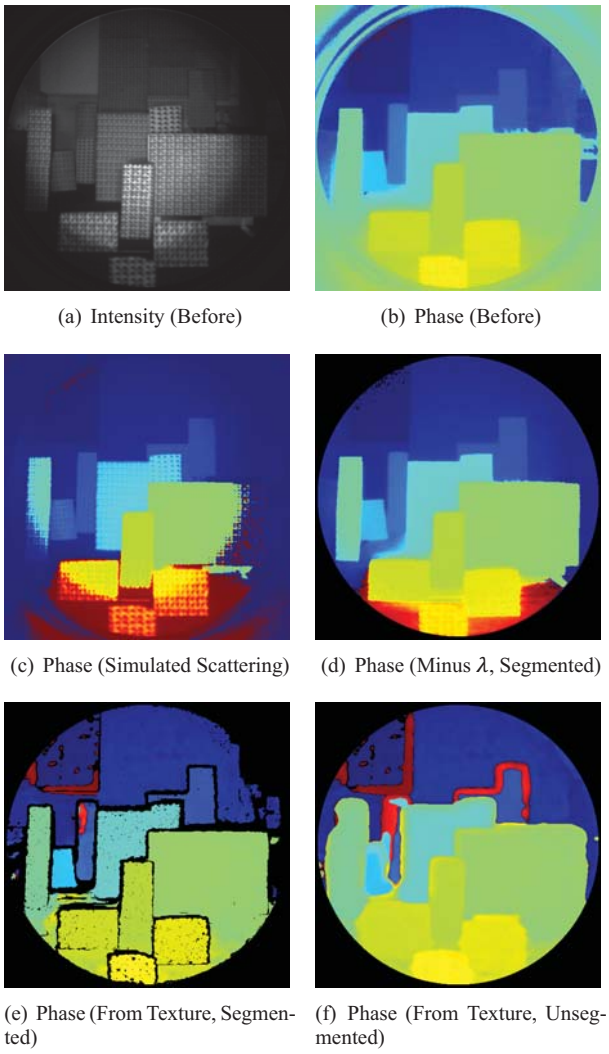
IV. Methodology and Results

A. Light Scattering Behaviour in a Typical Scene

Fig. 1 illustrates the effect of light scattering on a typical scene. Figs. 1(a) and 1(b) are a test scene consisting of a staircase structure with square patches of different reflectivities. Note that in 1(b) there is a very slight range-intensity coupling. Figs. 1(c) and 1(d) show the scene after the addition of a very bright box in the foreground. The greatest changes in range occurred in regions where the reflectivity of the staircase was very low, such as the middle step on the far right. Using the textured regions on the left, and top/left a global estimate for λ was calculated (as described in section III-B). These textured regions are quite widely spaced, meaning that only large scale scattering can be resolved. Subtraction of λ gives fig. 1(e), which shows significantly reduced range-intensity coupling, although the image has not been fully restored. This can be attributed to misestimation of λ. Fig. 1(f) shows the results of applying deconvolution to the image. The use of Fourier convolution and assumption of isoplanaticity has resulted in underestimated phase across the top of the image. The results are quite different to those produced by subtracting λ. By subtracting the original, largely unscattered image from the scattered image it is possible to estimate the amount of light scattered within the system. This is shown in figs. 1(g) and 1(h). The brightest regions of the background have much higher noise levels than the darker regions because the variance of a Poisson distribution is linearly proportional to intensity. Even when the intensity has been reduced by subtraction the noise still persists. The region at the top of the image exactly opposite the scattering object is slightly brighter than the rest of the scattering, which may be contributed to by the reflection seen in fig. 2 that is always on the opposite side of the principal point from the illumination source.

B. Results of New Algorithms

Two different patterned scenes were imaged. In order to provide rough ground truth, these scenes were set up in such a manner as to minimise the amount of naturally scattered light. Simulated scattered light (uncorrelated with the scene) was then added to produce corrupted range data as input to the algorithm for testing. However, since these were based on real data, there is a small amount of natural scattering variance across the image.



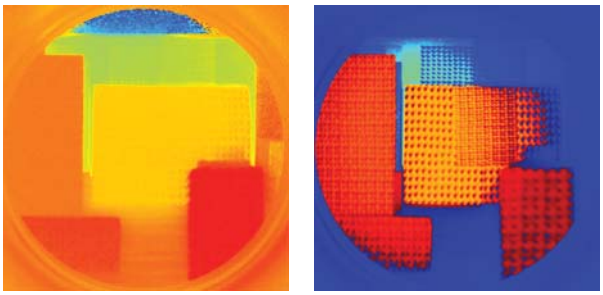
(a) Intensity (Before) (b) Phase (Before)

(c) Phase (Simulated Scattering) (d) Phase (Minus λ, Segmented)

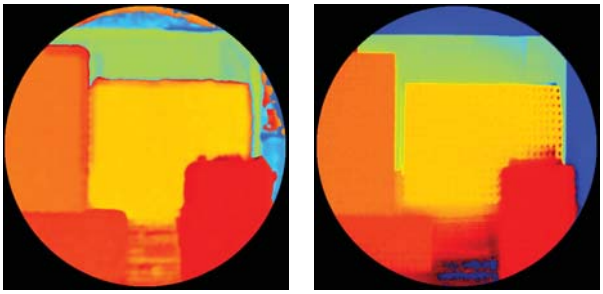
(e) Phase (From Texture, Segmented) (f) Phase (From Texture, Unsegmented)

Figure 4: The textured scene restoration process. 4(a) and 4(b) are the intensity and phase of the original highly textured scene. Red represented objects close to the camera and blue objects far away. Adding simulated scattered light gives 4(c), in which the patterned surfaces clearly show range-intensity coupling. Estimating λ using the segmented method and subtracting it from 4(c) gives 4(d). This is a substantial improvement, but darker regions are restored less well. Applying the phase from texture algorithm in the segmented and unsegmented versions results in 4(e) and 4(f) respectively.

Fig. 4 was processed using the range-image segmentation algorithm from section II-F. Because the segmentation relies on phase differences between adjacent pixels, the corrupted range data results in the scene being incorrectly segmented. In order to accurately segment the scene a rough estimate of λ is required – this is provided by running the scene through the algorithm once without any segmentation. An initial estimate of λ was then calculated and subtracted from $\beth$ in order to produce a correct range image segmentation model. The results of the segmented phase from texture algorithm are shown in fig. 4(e), this shows that the phase has been correctly recovered everywhere except



(a) Phase (Before) (b) Phase (Simulated Scattering)



(c) Phase (From Texture) (d) Phase (Minus λ)

Figure 5: Restoration of a textured scene. In this particular case, we have not segmented the image before processing it. This makes the algorithm a lot faster, but means that the phase from texture information is blurred around the edges of objects.

for a small box on the left. The resultant phase measurements are far more useful for most practical applications than the initial fig. 4(c) phases. Estimating λ via the segmented algorithm and subtracting from λ gives fig. 4(d), which is correct in all the patterned regions, however small parts of the table, which are unpatterned, have retained the perturbed phases from fig. 4(c).

Fig. 5 was processed in a similar fashion, but without any image segmentation. Fig. 5(b) shows severe range-intensity coupling, which is substantially reduced in the output image fig. 5(d) – however, the improvement is not as good as fig. 4(d). This is not entirely due to the lack of segmentation, but is contributed to by the scene layout and much smaller number of objects. As with fig. 4(f), the unsegmented phase from texture image (fig. 5(c)) contains erroneous values around edges of objects due to the size of the Gaussian window function η. Reducing σ could reduce these edge effects, but at the expense of increasing the impact of noise on the final results.

V. Conclusion

We have presented a new method for mitigating range-intensity coupling in full-field AMCW lidar images. The method is robust to light scattered from outside the FOV and does not require complex hardware calibration like previous deconvolution based methods. Phase information can also be directly extracted from patterned surfaces and the intensity and range of the scattered light estimated independently of the output range measurements. The method results in a substantial improvement in range data quality and can be implemented in either a fast unsegmented or slower, higher quality variant. Future development work includes analysis to determine optimal scene layout to facilitate accurate removal of scattered light and alteration of this method to enable its application to the object edge mixed pixel problem.

Acknowledgment

John Godbaz acknowledges the support of a TEC Top Achiever Doctoral Scholarship. This research is supported by the University of Waikato Strategic Fund.

References

[1] M. Hebert and E. Krotkov, "3D measurements from imaging laser radars: How good are they?" *Image and Vision Computing*, vol. 10, pp. 170–178, 1992.

[2] M. D. Adams, "Amplitude modulated optical range data analysis in mobile robotics," in *Proc. Robotics and Automation, 1993*. IEEE, 1993, pp. 8–13 vol.2.

[3] J. Tuley, N. Vandapel, and M. Hebert, "Analysis and removal of artifacts in 3-D ladar data, Carnegie Mellon Research Report, pp. 2203 – 2210.

[4] P. Tang, D. Huber, and B. Akinci, "A comparative analysis of depth-discontinuity and mixed-pixel detection algorithms, in *IEEE Int. Conf. on Robotics and Automation*, Los Alamitos, CA, USA, 2007, pp. 29–38.

[5] R. L. Larkins, M. J. Cree, A. A. Dorrington, and J. P. Godbaz, "Surface projection for mixed pixel correction," in *Image and Vision Computing New Zealand (IVCNZ'09)*, Wellington New Zealand, 2009.

[6] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, "Mixed pixel return separation for a full-field ranger," in *Image and Vision Computing New Zealand (IVCNZ'08)*, Christchurch, New Zealand, 2008.

[7] J. P. Godbaz, "Multiple return separation for a full-field ranger via continuous waveform modelling," in *Image Processing: Machine Vision Applications II, Proc. SPIE vol. 7251*, San Jose, California, 2009.

[8] J. Mure-Dubois and H. Hugli, "Optimized scattering compensation for time-of-flight camera," Proc. SPIE vol. 6762. Boston, MA, USA: SPIE, 2007.

[9] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *Proc. SIGGRAPH 07*, 2007.

[10] A. A. Dorrington, M. J. Cree, A. D. Payne, R. M. Conroy, and D. A. Carnegie, "Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera," *Meas. Sci. and Tech.*, vol. 18, no. 9, pp. 2809–2816, 2007.