Separating true range measurements from multi-path and scattering interference in commercial range cameras.

A.A. Dorrington[*], J.P. Godbaz, M.J. Cree, A.D. Payne, L.V. Streeter
University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

ABSTRACT

Time-of-flight range cameras acquire a three-dimensional image of a scene simultaneously for all pixels from a single viewing location. Attempts to use range cameras for metrology applications have been hampered by the multi-path problem, which causes range distortions when stray light interferes with the range measurement in a given pixel. Correcting multi-path distortions by post-processing the three-dimensional measurement data has been investigated, but enjoys limited success because the interference is highly scene dependent. An alternative approach based on separating the strongest and weaker sources of light returned to each pixel, prior to range decoding, is more successful, but has only been demonstrated on custom built range cameras, and has not been suitable for general metrology applications. In this paper we demonstrate an algorithm applied to both the Mesa Imaging SR-4000 and Canesta Inc. XZ-422 Demonstrator unmodified off-the-shelf range cameras. Additional raw images are acquired and processed using an optimization approach, rather than relying on the processing provided by the manufacturer, to determine the individual component returns in each pixel. Substantial improvements in accuracy are observed, especially in the darker regions of the scene.

Keywords: Time-of-flight, range imaging, range camera, multi-path, mixed pixel, scattering, separation, accuracy

1. INTRODUCTION

Full field time-of-flight range cameras, such as the MESA Imaging SR-4000 or PMD Technologies CamCube, are a relatively new technology that is gaining popularity for many applications. They are attractive because they are easy to use and can produce a three-dimensional (3D) point cloud and intensity image simultaneously for an entire scene at video frame rates from a single viewing location. However, this technology still suffers a number of limitations, including limited spatial resolution, limited distance measurement accuracy and precision, and motion artifacts.

The acquisition of 3D information at video rates provides significant advantages in many machine vision applications where absolute size or the detection of motion along the optical axis is important. In such applications, the precision and accuracy requirements are not particularly demanding, hence range cameras are often suitable if the low spatial resolution and motion artifacts can be overcome.

Several attempts have been made to use range cameras for metrology applications[1]. Typically, these measurements are of a static scene, meaning motion artifacts are irrelevant and precision can be improved with averaging. Accuracy errors, however, are a significant problem. One common problem described in the literature is the multi-path effect, which causes range distortions when stray light from other parts of the scene interferes with the range measurement in a given pixel[2]. Multi-path interference can originate from intra-camera scattering within the lens system, and from extra-camera scattering within the scene. The effect is significantly more prominent on darker objects because the multi-path interfering light is much brighter relative to the light returned from the object of interest.

Numerous attempts have been made to characterize multi-path interference in an effort to develop correction algorithms[3-9]. However, success with this approach is limited because the nature of the scattering and interference is highly scene dependent and many of the algorithms to date do not consider the physical cause of the distortions. Alternative approaches for correcting multi-path distortions based on separating the individual sources of light returned to each pixel, called Mixed Pixel Restoration, have been published[10-12]. Although these approaches are successful in counteracting multi-path distortions, they have so far only been demonstrated on custom built, laboratory based, uncalibrated, range cameras, hence have not been suitable for general metrology applications.

---

[*] a.dorrington@waikato.ac.nz; phone +64 7 858-5062; fax +64 7 838-4835; eng.waikato.ac.nz

In this paper, we first review the basic operating principles of range cameras and the mechanisms behind common sources of error. We then review the Mixed Pixel Restoration algorithm and discuss how it can be applied to off-the-shelf range camera hardware. Finally, we introduce the hardware used in this case and demonstrate a version of the Mixed Pixel Restoration algorithm applied to unmodified off-the-shelf commercial range cameras, and conclude with some comments of limitations and future work.

2. RANGE CAMERA OPERATION AND ERROR MECHANISMS

Amplitude modulated continuous wave (AMCW) full-field time-of-flight range cameras operate by measuring the propagation time for light to travel from the camera, to objects in the scene, and back to the camera again[7,13]. The entire scene is illuminated with a diverging light source, typically from a bank of light emitting diodes or laser diodes, that is intensity modulated in the region of 10 MHz to 100 MHz. Some of this illumination light is scattered back to the camera and is imaged onto the sensor. The propagation time for the light to travel to the object and return to the camera imparts a phase shift on the intensity modulation envelope. This phase shift is measured and distance for every pixel calculated, providing a 3D point cloud after perspective projection. The amplitude of the modulation envelope is also measured to simultaneously provide an active brightness image, providing a grayscale value for each pixel of the sensor, hence each point in the point cloud.

A specialized image sensor incorporating a gain modulation feature is necessary to measure the phase shift and amplitude of the modulation envelope. The sensor is modulated, or shuttered, at the same frequency as the illumination modulation during the image integration time, effectively correlating the collected light with the sensor's reference signal. A series of images, typically four, are acquired with predetermined phase offsets between the illumination and sensor modulation signals in order to calculate the amplitude and phase independently of background lighting or object reflectivity[7]. Distance is calculated from the phase value as a fraction of the wavelength of the modulation frequency, accounting for the light's return trip out to the object and back to the camera, as

$$d = \frac{\theta\lambda}{4\pi} \ , \tag{1}$$

where d is the distance from the camera's perspective center to the object, θ is the modulation envelope phase, and λ is the wavelength of the modulation signal.

Distance measurement accuracy is determined by the accuracy of the modulation envelope phase measurement, which is influenced by a number of factors both internal and external to the camera. Internal sources of error typically include phase measurement linearity errors, usually due to modulation signal harmonic aliasing[13], and offset drift caused by temperature effects or electronic signal generator drift.

In this paper, however, we are concerned with the external influences, which generally arise from interference effects, when multi-path causes light returned to a particular image sensor pixel to be contaminated with stray light in the scene. Figure 1 illustrates various potential sources of multi-path, which can originate from:

1. diffuse scattering or specular reflections within the scene, where illumination bounces off one object onto another;

2. scattering within the lens system of the camera, where a portion of the light returned from one object falls on pixels imaging a different object; and

3. object edges, where two object are imaged by a signal pixel (not illustrated).

The amount of interfering light from multi-path effects may seem small and insignificant compared to the direct return light, but only small amounts of interference can cause significant deviation to the phase value[14]. For a scene that contains a range of high contrast objects or highly specular surfaces, the interfering multi-path light can be significantly than the direct light returned from darker objects.
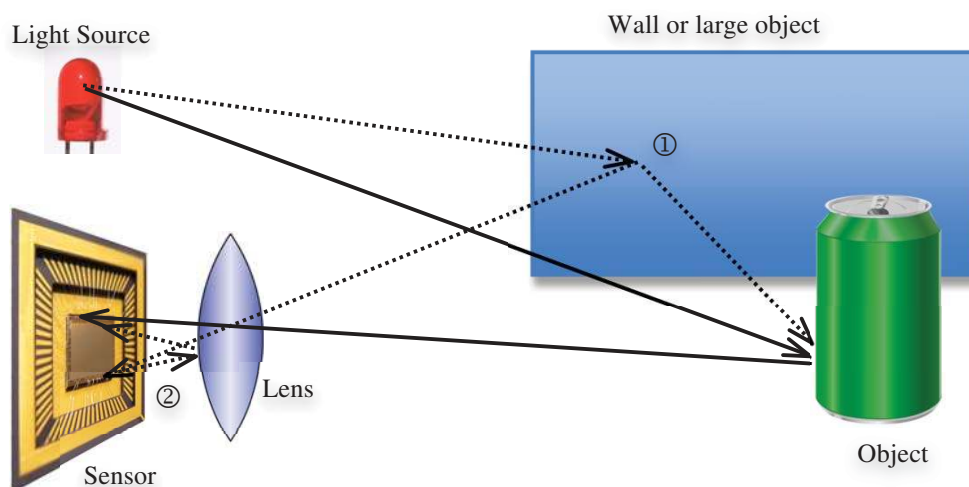
Figure 1 – Illustration of light propagation in a range imaging camera. Solid lines show direct part illumination and detection, broken lines show interfering multi-path. Sources of multi-path include ① scattering within the scene and② scatting within the lens system. (Images from Wikimedia Commons).

3. RESOLVING MULTI-PATH RETURNS

In the multi-path situation, the correlation waveform is a combination of individual waveforms for each component of the multiple returns. To find the true distance value, the individual components of these multiple returns must be separated. One approach demonstrated utilizes the harmonics present in the correlation signal[10-11]. However, these approaches involve measuring the correlation waveform shape in some detail, requiring the acquisition of a significant number of phase steps for each range image measurement. This is not possible with the majority of off-the-shelf range cameras (which are typically configured to acquire four phase steps per range measurement). Correlation waveform based approaches are also highly limited by temporal changes in the harmonic content of the correlation waveform due to temperature and other hardware instabilities.

An ideal AMCW capture is equivalent to sampling a particular bin of the spatial Fourier transform of component signal returns within pixel versus range; the harmonics of the correlation waveform in a practical system implicitly encode the same information at additional spatial frequencies, but with limited SNR. At a fundamental level correlation waveform harmonic/shape based algorithms can be considered to be a type of sparse spike deconvolution problem where the source data is a highly limited set of spatial Fourier transform bins. Whereas correlation waveform based methods implicitly sample a plurality of spatial Fourier transform bins, encoding them in the correlation waveform harmonics, the (patent pending) approach we take in this paper is to explicitly sample two different spatial frequencies of the component returns[12]. Although the assumption of only two component returns may lead to in accuracies in the resolved return components, it still provides a significant improvement versus the unprocessed alternative. This approach is suitable for implementation on off-the-shelf cameras because it requires only standard range image acquisitions that the cameras generate in normal operation, at two different modulation frequencies.

The most natural representation for an AMCW range measurements is in the complex domain. We notate the phase and amplitude values of the modulation envelope for a given pixel as

$$m_l = a_l \cdot e^{j\varphi_l}, \tag{2}$$

where a_l and φ_l are the measured amplitude and phase respectively for a series of l acquisitions at a particular modulation frequency.

In the case of a mixed pixel, each of these measured phasors is a combination of multiple component returns given by

$$m_l = \sum_i \alpha_i \cdot e^{j\vartheta_i (f_l/f_0)},$$ (3)

where a_i and ϑ_i are the amplitude and phase values respectively for the ith component return in a particular pixel. For convenience, we notate the phase measured at a particular frequency f_l as if it were measured at frequency f_0, to ensure all phase values are related to distance in the same way regardless of modulation frequency. Note that the amplitude of a particular return component indicates the amount of light in that return, hence it should, in principle, remain constant for all modulation frequencies; while the phase value changes in proportion to the modulation wavelength, because the propagation time remains constant (see equation 1).

Performing multiple measurements of the same scene at a variety of modulation frequencies provides a set of simultaneous equations. For this paper we have used a frequency ratio of 2:1. If we use the lowest frequency to notate the phase of the component returns and assume two component returns, we get

$$m_1 = a_1 \cdot e^{j\varphi_1} = \alpha_1 \cdot e^{j\vartheta_1} + \alpha_2 \cdot e^{j\vartheta_2}$$
$$m_2 = a_2 \cdot e^{j\varphi_2} = \alpha_1 \cdot e^{2j\vartheta_1} + \alpha_2 \cdot e^{2j\vartheta_2}$$ (4)

where $l \in \{1,2\}$ are the low and high frequency measurements respectively, and $i \in \{1,2\}$ are the primary and secondary component returns. This can be considered to be equivalent to a reduction of the earlier correlation waveform deconvolution method to the simplest useful case of the fundamental frequency with an additional second harmonic. Thus far we have been unable to find an analytic inverse for the non-overdetermined case; as a result, for this paper we pursue a numerical optimization based approach. An alternate, real-time implementable algorithm as well as analytic inverses in the overdetermined case are presented in Godbaz et. al.[12]

For each pixel we use a least squares approach to solve for the component returns. This is given by

$$\underset{(\alpha_1, \alpha_2, \vartheta_1, \vartheta_2)}{\arg\min} |m_1 - \hat{m}_1|^2 + |m_2 - \hat{m}_2|^2$$ (5)

where $\hat{m}_l$ is a forward estimate of measurement l using equation 4.

Once an appropriate set of component returns estimates are found, an estimate of the original unperturbed component return is generally desired; unfortunately, given two arbitrary returns this is an ambiguous situation. Simple approaches include finding the brightest return or the closest distance, but these can sometimes lead to errors. More sophisticated approaches such as smoothness constraints could be used, but so far have not been investigated – possible mechanisms including solving a Markov Random Field.

Although the algorithm is capable of operating on phase and amplitude values calculated by the camera or its accompanying software, better results are often obtained by operating the camera in a "raw" mode where the individual phase step images are acquired, and phase and amplitude values are calculated externally to the camera. This averts deleterious impacts on the algorithm due to any attempts at phase non-linearity compensation.

4. CALIBRATION AND PERSPECTIVE PROJECTION

Previous publications have demonstrated how multi-returns can be resolved in range cameras, but stop short of showing how this can improve the accuracy of real-world measurements[10-12]. In order to demonstrate this, the range images must be projected to Cartesian coordinates, which involves considering the radial lens calibration. Most off-the-shelf cameras are shipped calibrated, and are capable of generating real-world point cloud outputs. However, the raw uncalibrated data is more suitable for the mixed pixel algorithm, and the manufacture's calibration is lost in this mode of image acquisition.

In order to retrospectively apply the factory lens calibration, the calibration values are determined from an initial calibrated image acquired from the camera. A 3D point cloud is acquired and these data are then normalized by radial distance, providing a bundle of unit pointing vectors, with

$$\hat{\boldsymbol{c}}_{i,j} = \frac{\boldsymbol{x}_{i,j}}{\left|\boldsymbol{x}_{i,j}\right|}, \tag{5}$$

where $\hat{\boldsymbol{c}}_{i,j}$ is the calibration pointing vector for pixel in row i, column j; $\boldsymbol{x}_{i,j}$ is the vector (from the origin) pointing to the measured 3D point produced by the camera for the same pixel; and $\left|\boldsymbol{x}_{i,j}\right|$ is the Euclidean distance from the origin to that measured point (assuming the origin of the camera's coordinate system is at the perspective center).

The range cameras used in this paper have a fixed lens, with a fixed focal setting, meaning the radial lens calibration should remain constant. Hence, once the calibration pointing vector bundle is generated for a specific camera, a perspective projection can be performed by simply multiplying the pointing vector by the radial distance for each pixel, as

$$\boldsymbol{x}'_{i,j} = \hat{\boldsymbol{c}}_{i,j} d_{i,j}, \tag{6}$$

where $\boldsymbol{x}'_{i,j}$ is the vector for the projected 3D point for each pixel; and $d_{i,j}$ is the multi-return resolved radial distance for each pixel.

Amplitude calibration is required because the cameras are non-ideal and report different amplitude values for the same amount of returned light at different modulation frequencies. Ordinarily, camera manufacturers are not concerned with this effect, as the amplitude measurement is simply a grayscale image of the scene, and absolute values are not important. However, when attempting to resolve multiple returns, accurate amplitude values are vitally important. Therefore, the amplitude response of the camera is calibrated by normalizing the acquired amplitude data to a particular region of the scene that has been identified as minimally affected by multi-path interference. Calibration of the phase offset is also important when resolving multi-path interference. Electronic and optical delays in the cameras impart a phase offset in the measurements, which changes with modulation frequency. Some cameras utilize a feedback mechanism to compensate for this offset, but this often not be implemented in the raw capture mode.

5. EQUIPMENT AND DATA ACQUISITION

The multi-return processing algorithm was tested on two different camera platforms currently available off-the-shelf. These cameras, shown in figure 2, are the SR-4000 from Mesa Imaging (Zurich, Switzerland) and the XZ-422 Demonstrator, utilizing the Jaguar image sensor, from Canesta (Sunnyvale, California, USA). The camera specifications are summarized in table 1.

For a particular scene, data was collected in both the camera's normal Cartesian coordinate calibrated mode, and in the raw mode needed for multi-path processing. In each mode, 100 measurements were acquired and averaged to improve noise performance and precision. Total measurement time to acquire both sets of data is of the order of three to four minutes for each camera. Integration time was adjusted and set manually for each scene to around 20% below obvious saturation to ensure the results from Cartesian coordinate and raw modes are comparable. Care was taken to disable the illumination system for the camera not in use, ensuring the cameras did not interfere with each other.

Table 1. Camera configuration summary

Camera Model	SR-4000	XZ-422
Manufacturer	Mesa Imaging	Canesta
Image resolution	176×144	160×120
Illumination type	IR LED	IR LED
Modulation frequency 1	15 MHz	22 MHz
Modulation frequency 2	30 MHz	44 MHz

Figure 2. Range imaging cameras mounted together on a tripod. SR-4000 on top and XZ-422 on the bottom.

6. METHADOLOGY AND RESULTS

Initially a very simple scene designed to promote multi-path was set up. This consisted of a dark background wall and bright foreground object approximately 200 mm by 200 mm in size. The cameras' position was fixed at 3.78 m from the dark wall. The foreground object was initially placed 250 mm in front of the wall and moved forward in increments of 250 mm, acquiring data at each location to a maximum distance of 2.50 m from the back wall. Figure 3 shows plots of one row of pixels, in the middle of the image, across the foreground object for the Cartesian coordinate data from both cameras. Note that the ambiguity distance for the XZ-422 camera operating in this mode is 3.41 m, so a correction has been applied by adding one ambiguity distance (projected from the perspective center) to all pixels with radial distances less than 1.1 m.

In figure 3 the planar surface of the back wall should be represented by a straight line at approximately 3.8 m. However, close to the foreground object, errors are evident in that the background appears to deviate from a planar surface as light scattered from the bright foreground object interferes with the light received from the back wall. This is most noticeable for images from the XZ-422. These errors can be corrected with the mixed pixel separation algorithm, as the light from the back wall is resolved separately from the interference. Figure 4 shows these corrected measurements after the acquired data is processed with the multi-path separation algorithm. Note the errors close to the object boundaries have been significantly reduced, but in the case of the SR-4000, additional noise or uncertainty is apparent.

For clarity, figure 5 shows two cases showing profound errors (foreground to background separation of 2.5 m and 2.0 m) plotted directly comparing corrected and uncorrected acquisitions from the XZ-422 camera. Note the elimination of more than 500 mm of error in the 2.5 m separation case, and how the back wall is shown as considerably more planar in the corrected data.
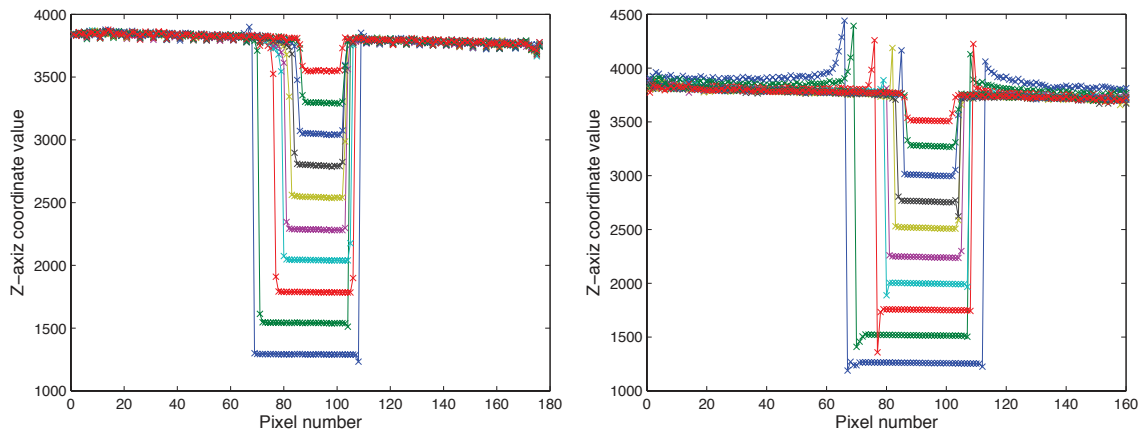
Figure 3 – Z-axis cross section of Cartesian coordinate data of a dark back wall with a bright foreground object from the SR-4000 (left) and XZ-422 (right), averaged over 100 acquisitions.
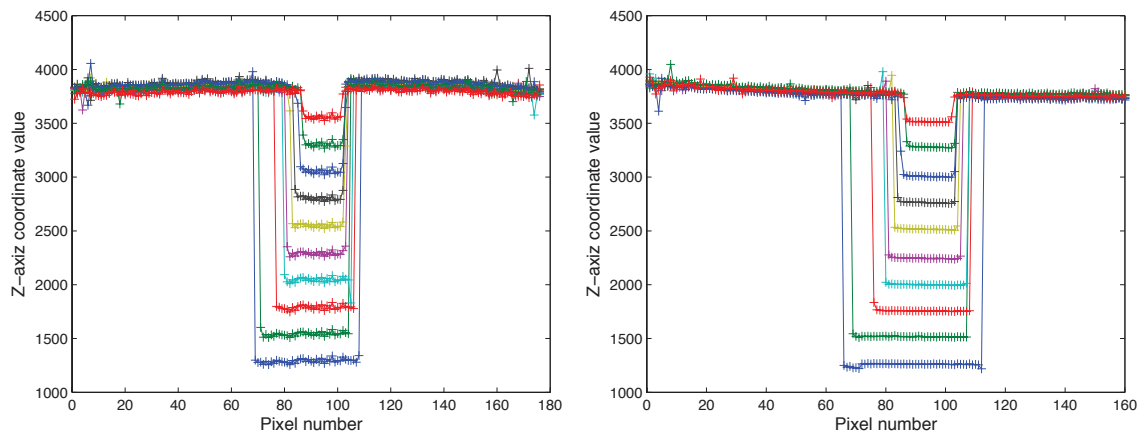

Figure 4 – Z-axis cross section of multi-path resolved data of a dark back wall with a bright foreground object from the SR-4000 (left) and XZ-422 (right), averaged over 100 acquisitions.
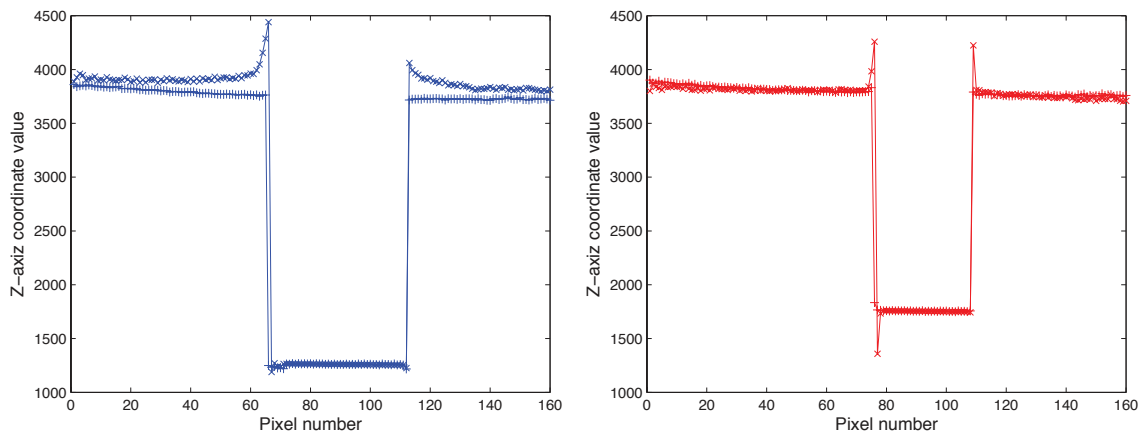

Figure 5 – Z-axis cross section comparison of corrected (+) and uncorrected (×) acquisitions from the XZ-422 camera with separation between foreground and background objects of 2.5 m (left) and 2.0 m (right), averaged over 100 acquisitions.
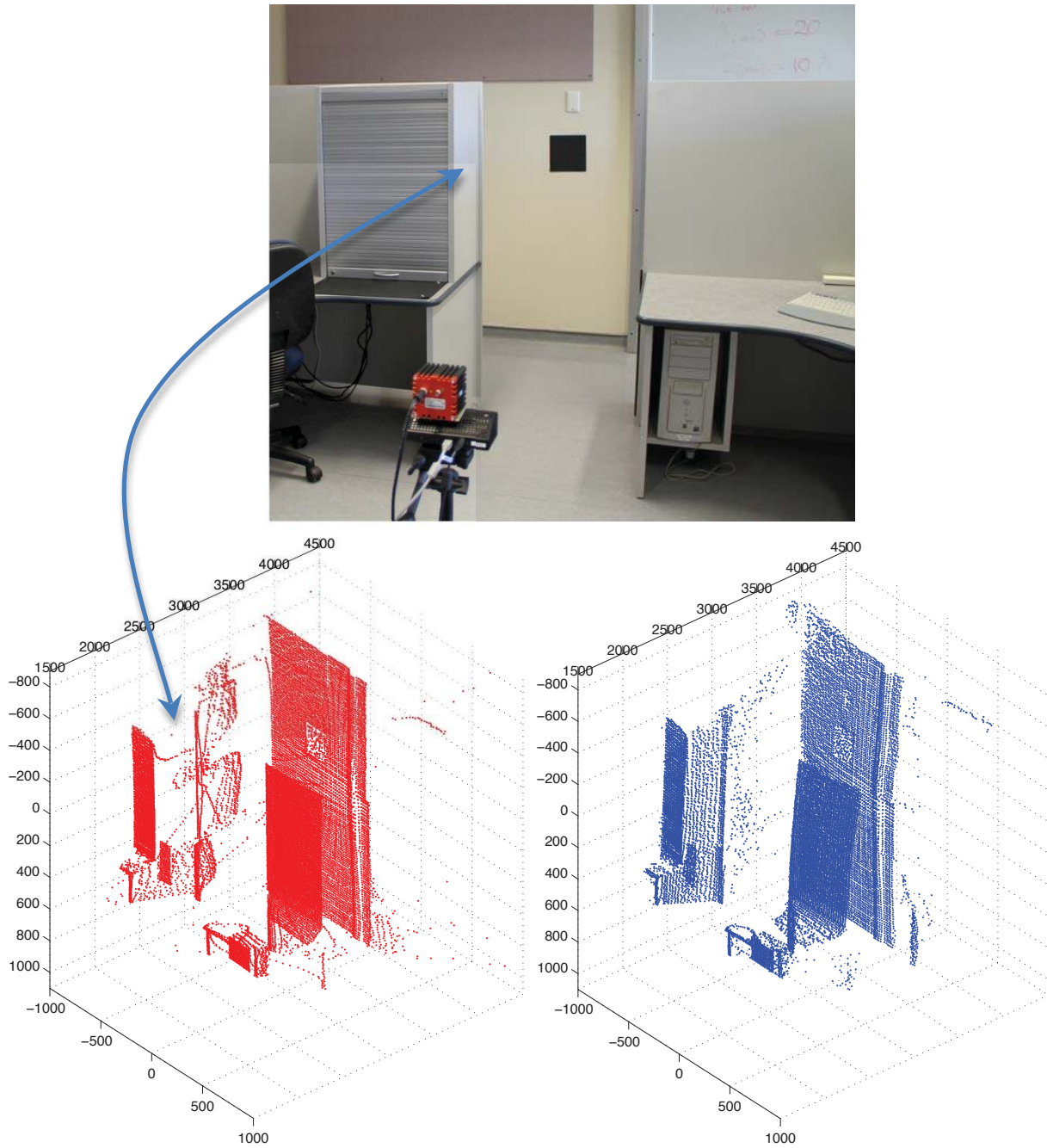
Figure 6 – Photograph of the range imaging cameras acquiring data in a real-world scene (top), and 3D point clouds of the uncorrected data from that scene (bottom left) and the corrected data (bottom right). Note the surface indicated by the arrows is incorrectly represented in the uncorrected data, but more correctly represented in the corrected data.

Figure 6 shows a more realistic scene of an electronics teaching lab. Take particular note of the surface indicated by the arrows. It can be seen in the photograph that this surface has some specular qualities. This, combined with its oblique angle to the camera means it is particularly susceptible to multi-path interference. It can be clearly seen in the uncorrected measurement that the data points for this surface have been measured incorrectly, appearing much closer to the back wall that they should be. In the corrected data, the points are qualitatively in the correct position. Without ground truth data, it's impossible to quantify the error. For all previous data sets in this paper, the brightest of the resolved multi-path returns have been taken as the primary and correct return. However, in this scene, light from the back wall reflected off the surface of interest was, in some regions, brighter then the return from the surface itself. Hence, selecting the brightest return gave an incorrect result, leading to the need to choose the closest return, with some sensibility constraints, as the correct result.

7. CONCLUSION, LIMITATIONS, AND FUTURE WORK

The results presented in this paper show how the mixed pixel and multi-path separation algorithm can be applied to range imaging measurements projected into Cartesian coordinates. Significant improvements in accuracies can be obtained in particular situations where the influence multi-path interference normally causes distortions. In one example, errors greater than 500 mm have been reduced by more than one order of magnitude.

However, it has been noted while acquiring data that not all scenes benefit from applying the return separation algorithm. In some cases it simply adds noise and reduces precision, while in other cases it produces incorrect results and reduces accuracy. One possible reason for this is the ad-hoc way in which a calibration is applied to the data during processing. As discussed in section 4, the multiple return separation algorithm requires a consistent and known relationship between both amplitude and phase response of the cameras at both modulation frequencies. This type of calibration was not available for these cameras, so it was performed by comparing a manually selected regions of the image judged to be minimally affected by multi-path interference. If these regions did contain multi-path, then the calibration would be compromised leading to poor performance. Another source of potential error is distance linearity measurement inaccuracies common in uncalibrated range cameras. It is also possible that the optimization algorithm encountered local minima, producing an incorrect answer.

Future work could involve repeating this study after a full radiometric and linearity calibration of the cameras operating in the raw mode at both operating frequencies. Assuming the cameras are suitably stable, this removes the need for the ad-hoc calibration and eliminates a significant source of potential errors. It would also be useful to acquire a ground truth measurement, possibly with a high quality laser scanner, to quantitatively assess the improvements in accuracy obtained with the multi-path separation algorithm.

REFERENCES

[1] Boehm, J. and Pattinson, T., "Accuracy of exterior orientation for a range camera", Proc. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII, Part 5, 103-108 (2010).

[2] Karel, W., Ghuffarb, S. and Pfeifer, N., "Quantifying the distortion of distance observations caused by scattering in time-of-flight range cameras", Proc. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII, Part 5, 316-321 (2010).

[3] Oprisescu, S., Falie, D., Ciuc, M. and Buzuloiu, V., "Measurements with ToF cameras and their necessary corrections", Proc. International Symposium on Signals, Circuits and Systems 2007 (ISSCS 2007), 1-4, DOI 10.1109/ISSCS.2007.4292691 (2007).

[4] Falie, D., "3D image correction for time of flight (ToF) cameras", Proc. SPIE 7156, 715615, DOI 10.1117/12.807125 (2008).

[5] Kavli, T., Kirkhus, T., Thielemann, J. T. and Jagielski, B., "Modelling and compensating measurement errors caused by scattering in time-of-flight cameras", Proc. SPIE 7066, 706604, DOI 10.1117/12.791019 (2008).

[6] Radmer, J., Fuste, P.M, Schmidt, H. and Kruger, J., "Incident light related distance error study and calibration of the PMD-range imaging camera", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2008 (CVPRW '08), 1-6, DOI 10.1109/CVPRW.2008.4563168 (2008).

[7] Chiabrando, F., Chiabrando, R., Piatti, D. and Rinaudo, F., "Sensors for 3d imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera", Sensors 9, 10080-10096, DOI 10.3390/s91210080 (2009).

[8] Jamtsho, S. and Lichti, D. D., "Modelling scattering distortion in 3D range camera", Proc. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII Part 5, 299-304 (2010).

[9] Fuchs, S., "Multipath interference compensation in time-of-flight camera images", Proc. 2010 International Conference on Pattern Recognition, 3583-3586, DOI 10.1109/ICPR.2010.874 (2010).

[10] Godbaz, J. P., Cree, M. J. and Dorrington A. A., "Mixed pixel return separation for a full-field ranger", Proc. Proc. Of the Image and Vision Computing New Zealand Conference (IVCNZ'08), Christchurch, New Zealand, ISBN 978-1-4244-3780-1, 1-6 (2008).

[11] Godbaz, J. P., Cree, M. J. and Dorrington A. A., "Multiple return separation for a full-field ranger via continuous waveform modeling", Proc. SPIE 7251, 72510T (2009).

[12] Godbaz, J. P., Cree, M. J. and Dorrington A. A., "New methods for mixed pixel component separation and bounding", In preparation.

[13] Payne, A. D., Dorrington, A. A., Cree, M. J. and Carnegie, D. A., "Improved measurement linearity and precision for AMCW time-of-flight range imaging cameras", Applied Optics 49(23), 4392-4403 (2010).

[14] Dorrington, A. A., Cree, M. J. and Carnegie, D. A., "The importance of CCD readout smear in heterodyne imaging phase detection applications", Proc. Image and Vision Computing New Zealand Conference (IVCNZ'05), Dunedin New Zealand, ISBN 0-473-10523-3, 73–78 (2005).