AN EVALUATION OF TIME-OF-FLIGHT RANGE CAMERAS FOR CLOSE RANGE METROLOGY APPLICATIONS

A. A. Dorrington*, A. D. Payne, and M. J. Cree

School of Engineering, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand
* a.dorrington@waikato.ac.nz

KEY WORDS: Photogrammetry, LIDAR, Metrology, Camera, Point Cloud, Accuracy, Precision, Three-dimensional

ABSTRACT:

Time-of-flight range cameras are an emerging technology that produces a digital photo or video like output in which every pixel contains both intensity and distance information. Typical currently available off-the-shelf range imaging cameras offer spatial resolutions of up to 40 k-pixels, distance measurement precision of less than 1 cm, and accuracy of a few centimetres. These specifications seem poor for metrology applications, but significant improvements can be made with temporal and spatial averaging. We investigate and compare the SR4000 (Mesa Imaging) and XZ422 (Canesta Inc.) range cameras for close range metrology applications. As an example, we first describe a real-time person height measurement system. We then present range camera measurements of a three-dimensional reference object and compare to a photogrammetric survey to evaluate the precision and accuracy obtained. Results show that, with appropriate averaging, the range cameras can provide sub-millimetre precision. The accuracy results, however, are not as encouraging, at a few millimetres for the XZ422 camera and tens of millimetres for the SR4000 camera, with poor radial lens calibration the likely source of errors for the SR4000. Finally, we comment on the promising possibilities for the future of range cameras, assuming that the technology continues to improve.

1. INTRODUCTION

Time-of-flight range cameras are an emerging technology that produce a digital photo or video like output where every pixel contains both intensity and distance information. These distance measurements are obtained with an active approach, typically using the Amplitude Modulated Continuous Wave (AMCW) indirect time-of-flight method, simultaneously measuring the distance to objects in the scene for every pixel in the image. Hence, these cameras offer advantages over traditional close range metrology techniques, providing three-dimensional data in a matter of seconds (with both illumination and detection) from a single viewing location.

According to manufacturer's specifications, typical currently available off-the-shelf range imaging cameras offer spatial resolutions of up to 40 k-pixels, distance measurement precision of less than 1 cm, and accuracy of a few centimetres (PMD Technologies, 2009; MESA Imaging, 2009). When compared to other metrology techniques, these specifications seem poor, however, it is important to consider that they are reported on a per-pixel and per-frame basis. Because range cameras are capable of video frame rates, assuming a static scene, hundreds of measurements can be averaged temporally in only a few seconds leading to more than an order of magnitude improvement in precision. In addition, many metrology applications involve objects or targets that span multiple pixels, such as simple retro-reflective circular targets, allowing the use of spatial averaging or smoothing algorithms to further improve measurement precision.

Although currently this technology has several limitations, it is maturing rapidly, and is showing promise as a simple, low cost, and fast solution to many machine vision and metrology applications. For example, spatial resolution continues to increase, with PMD Technologies (Siegen, Germany) currently offering cameras with 204 by 204 pixels (PMD Technologies,

2009). If this rate of improvements continues, VGA resolution range imaging cameras may be available in the near future.

In this paper we first briefly review range imaging camera technology, and discuss measurement precision and accuracy. We then investigate the potential of two different off-the-shelf range imaging cameras for simple metrology applications. As an example application, we first demonstrate a basic person height measurement system. This software uses simple image processing techniques on both the range and intensity information to detect people in the scene and estimate the height of each person in real time. To investigate the cameras' performance in more demanding applications, we also present measurements of a "staircase" shaped calibration object, with 64 circular retro-reflective targets attached, and compare these results to photogrammetric survey.

2. TIME-OF-FLIGHT RANGE IMAGING

2.1 Background

Typical off-the-shelf range imaging camera's employ the AMCW indirect measurement approach. This is an active imaging system that relies on flood-lighting the scene of interest with amplitude modulated illumination (typically in the region of 10 to 100 MHz) and detecting the phase shift in the modulation envelop of the returned light (Lange, 2000; Büttgen et. al., 2005). The phase measurement is performed with a specialised gain modulated imaging sensor that can be conceptually described as high-speed shuttering during image integration, operating at the same rate as the illumination modulation. Measured phase shift can be related directly to propagation time, hence object distance, at each pixel by

$$d = \frac{\varphi c}{4\pi f},$$

(1)

where d is the distance measured, φ is the measured phase, c is the speed of light, and f is the modulation frequency (Büttgen et. al., 2005).

The native output of a range imaging camera comprises two separate images. One is like a traditional greyscale digital photograph containing the so-called "active brightness" image that is generated from the amplitude of the detected modulation envelope. The second is the range image containing the radial distance values from the perspective centre of the camera to the object for each pixel (as per equation 1). The camera manufacturer usually provides a perspective projection function to transform the distance image into a three-dimensional Cartesian coordinate point-cloud

2.2 Measurement precision

Note from equation 1 that the distance measurement is dependent on the phase measurement and the operating frequency, hence, measurement precision is governed by quality of the phase and the modulation frequency, and is defined as (Büttgen et. al., 2005)

$$\sigma = \frac{c}{4\pi f} \cdot \frac{\sqrt{(A+B)/2}}{c_d A} \qquad (2)$$

where A and B are the modulation envelope amplitude and the DC offset of the collected light respectively, and c_d is the demodulation contrast, which describes the quality of the light source modulation and the efficiency of the image sensor shutter (Büttgen et. al., 2005).

For camera manufacturers, the easiest way to improve measurement precision is to increase the modulation frequency, however, the maximum practical operating frequency is limited by both bandwidth limits of the electronics and ambiguity effects. The power-bandwidth product of LEDs is approximately constant (Mynbaev and Scheiner, 2001), therefore the high intensity infra-red Light Emitting Diodes (LEDs) typically used as the illumination source restrict the maximum amplitude modulation frequency to around 40 to 50 MHz.

Ambiguity effects arise because of the cyclic nature of the modulation envelope phase measurement (Büttgen and Seitz, 2008; Payne et. al., 2009; McClure et. al., 2009). The phase, φ, has a range of 0 to 2π, hence the maximum operating distance is limited to half of the wavelength the modulation frequency (because the light must complete a return journey). Measurements attempted at larger distances result in errors, termed ambiguity or aliasing errors. Techniques to remove these ambiguity effects have been demonstrated in the laboratory, but have found only limited implementation in off-the-shelf cameras (Payne et. al., 2009; McClure et. al. 2009, Dorrington et. al., 2007). Camera manufactures typically offer selectable modulation frequencies, allowing the user the trade-off between operating distance and precision.

2.3 Measurement accuracy

The measurement accuracy of AMCW range cameras is influenced by both the accuracy of the distance measurements and imaging artefacts such as lens distortion. During the perspective projection process, lens distortion is corrected in the normal way with calibration files shipped with the cameras. In order to avoid the need for end-user calibration, most cameras are factory calibrated and shipped with fixed focus non-removable lenses. Unfortunately this constrains their use in a number of practical applications because the field of view cannot be adjusted. Focus is not usually a problem due to the low spatial resolution of the cameras.

In principle, the use of sinusoidal modulation for the illumination and sensor gain produces perfectly linear distance measurements, leaving only a constant delay in the electronics affecting distance accuracy. In practice, however, sinusoidal modulation in not achieved because both the light source and the sensor have a non-linear response. In fact, it is common to use square wave modulation to simplify implementation using digital electronics and reduce power consumption.

Such amplitude modulation signals contain many harmonics that, with the AMCW homodyne technique, cause phase measurement non-linearities (Lange, 2000; Payne et. al., 2008; Dorrington et. al. 2008) manifesting as distance accuracy errors. These errors are normally compensated for with factory distance measurement calibrations. However, as it is possible for the response of the light source or image sensor to change with temperature, the effectiveness of these calibrations is limited for applications that require high accuracy.

A further cause of accuracy errors is the stability of the camera's modulation clock. If the actual operating frequency is not measured or tracked, drift or offset errors in the clock generator will propagate to the distance measurements, due to the inverse relationship in equation 1.

2.4 Imaging artefacts

Accuracy is also affected by imaging artefacts that are scene dependent. In general these artefacts arise from multi-path effects, which can be divided into two categories depending on the mechanism behind the effect. In the first category we consider effects arising from illumination and returned light occurring strictly within a particular pixel's field of view, and in the second we consider illumination light detected in one pixel that was intended for another.

Distance measurement errors at object edges are common because a pixel can collect light from two or more (foreground and background) objects simultaneously. These multiple returns interfere with one another resulting in erroneous measurements (Larkins et. al., 2009). This is called the "mixed pixel" or "flying pixel" effect.

Ideally, all light that is detected by a given pixel results from illumination within that pixel's field of view. However, normal objects are not retro-reflective, and the illumination is scattered over a wide angle, therefore it is not uncommon for a given pixel to not only detect the illumination intended for that pixel, but to also detect illumination that has been scattered from other parts of the scene. This so called "multi-path" effect can cause interference and errors in the distance measurements (Guðmundsson et. al., 2007, Falie and Buzuloiu, 2009). Multi-path error is most significant for darker objects that are close to much brighter objects.

3. REALTIME APPLICATION EXAMPLE

To illustrate the potential uses of range cameras in simple real-time applications, we present an example algorithm that

identifies objects and estimates their height off the ground. This is primarily aimed at measuring people's height. The height of a person can be determined as long as the top of the object remains in frame, regardless of their distance from the camera. An example output of the final application is shown in figure 1.
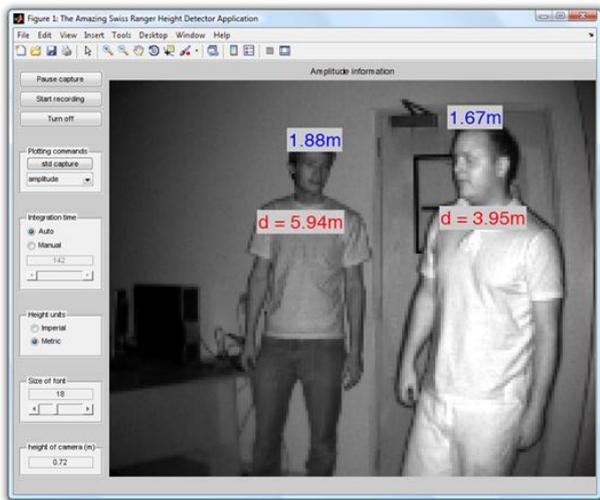
Figure 1. Screen shot from person height detection software. Average distance from the camera (labelled "d") and object height (not labelled) is displayed on the active brightness image in real time.

The image processing algorithm starts by capturing the background scene without any people. This forms the background image for all subsequent captures and is valid as long as the scene geometry remains unchanged. For each capture a foreground mask is calculated by thresholding the difference between the current capture and the background image, that is, a background subtraction algorithm. This marks the people in the scene. Any edges of the scene are then detected by thresholding the gradient of the range image. These edge pixels are deleted from the foreground mask to remove any mixed pixels and to help distinguish between overlapping groups. In order to measure more than one person at a time, the foreground mask is separated into groups using the watershed transform. The three-dimensional data for each group are then analysed to determine the maximum height and average distance from the camera for each person. The algorithm is implemented in Matlab®, and can process tens of frames per second on a 2 GHz class computer.

4. PERFORMANCE EVALUATION

4.1 Camera hardware

To determine if range imaging cameras can offer a reasonable alternative to traditional metrology technologies, we evaluate the precision and accuracy of two currently available off-the-shelf cameras, shown in figure 2, the Mesa Imaging SR4000 (Zurich, Switzerland) and the XZ422 Demonstrator (utilising the Jaguar image sensor) from Canesta (Sunnyvale, California, USA). A summary of the camera configuration is detailed in table 1.
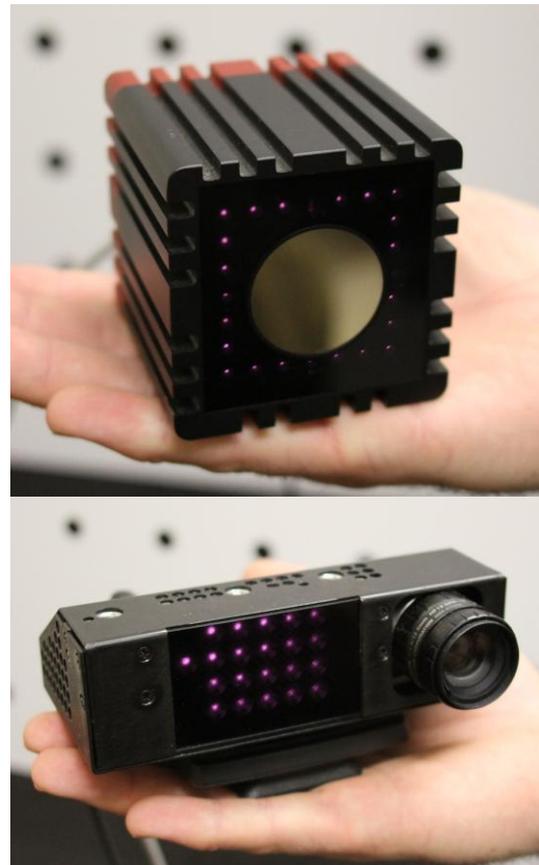
Figure 2. Range imaging cameras, SR4000 (top), and XZ422 (bottom). Note that the IR LEDs seen here as purple are not as visible to the naked eye.

Camera Model	SR4000	XZ422
Manufacturer	Mesa Imaging	Canesta
Image resolution	176×144	160×120
Illumination type	IR LED	IR LED
Modulation frequency	30 MHz	44 MHz

Table 1. Camera configuration summary

4.2 Reference object

A "staircase" shaped object (figure 3) of size 0.8 by 0.8 by 0.6 metres with 64 circular retro-reflective targets attached (16 on each of the 4 steps) was used to provide a constant reference structure suitable for measurement comparisons. A photogrammetric survey of this object was performed using a Canon 7D camera and the PhotoModeler software package to provide a high-precision ground-truth reference. Photomodeler reported average measurement precisions of 11.7, 12.4, and 24.7 µm in x, y, and z dimensions respectively. The corresponding maximum precision values were 12.8, 17.3, and 33.3 µm.

The retro-reflective targets on the reference object are useful to help avoid multi-path interference in the images. Because the targets are significantly brighter than any other objects in the scene, they are much less susceptible to interference from scattered light. Furthermore, if any scattered light impinges on the retro-reflectors, the majority of this light is reflected back to its source, and not back to the camera.

Figure 3. "Staircase" reference object.

4.3 Data acquisition

Both cameras were arranged (in turn) such that the reference object filled the field of view as completely as practical. The integration time of the camera was adjusted to acquire the strongest signal possible (thereby providing the best precision) without saturation. All acquisitions were acquired inside under fluorescent ambient lighting.

The SR4000 camera did not provide a sufficiently short integration time to avoid saturation on the retro-reflective targets, so a piece of paper was placed in front of the LEDs to diffuse the illumination. With the diffuser in place an integration time of 8 ms was suitable. The SR4000 has the option to apply a median filter to help reduce noise in the range image during acquisition. For all measurements taken for this paper, the median filter was disabled to remove any potential bias this may introduce to the results.

The integration on the XZ422 camera was set to 0.6 ms, much shorter than normal operation due to the use of retro-reflective targets. The XZ422 provides a number of different operating modes that allow the trade off of spatial image resolution vs. range image acquisition time. For this experiment, the "temporal" mode was selected to maximise the spatial resolution of the data. A consequence of this operating mode, and the sliding-window type processing, used by the camera is that only every fourth range image generated by the camera is unique. We temporally decimated all data collected from this camera by a factor of four to ensure that each range image is unique. All frame rates and frame counts specified henceforth relate to the decimated data.

A series of 250 images were acquired and averaged to improve overall precision. For this experiment, the SR4000 and XZ422 were run at 7 and 12.5 range images per second respectively, acquiring these images in less than 36 and 20 seconds respectively.

Imaged targets ranged from approximately 3 to 8 pixels in diameter. Three-dimensional centroid locations were found for each target by averaging the camera's calibrated x,y,z output, including only the pixels with an active brightness of at least 50% of the brightest pixel in a 9 by 9 pixel window. The set of target locations were roughly oriented to match the photogrammetric reference measurement, and then aligned with an iterative approach to find the smallest root-mean-squared residual (of all targets in all dimensions).

4.4 Results

Precision was determined by repeating the 250 image averaged measurement 100 times and finding the standard deviations for each target. Overall one-sigma precisions were found by averaging the x, y, and z, precision values over all 64 targets, and are summarised along with best and worst precision values in table 2.

	x (mm)	y (mm)	z (mm)
SR4000			
Average precision	0.055	0.038	0.271
Best precision	0.008	0.006	0.178
Worst precision	0.360	0.138	0.332
XZ422			
Average precision	0.343	0.230	0.385
Best precision	0.111	0.077	0.085
Worst precision	0.673	0.715	0.743

Table 2. Range camera one-sigma precision results.

Precision in the distance measurement (z) dimension is reasonably consistent between the two cameras, with the only notable difference being an approximately two times wider range of values for the XZ422. The XZ422 operates at a 47% higher modulation frequency than the SR4000, so if the phase measurement precision is the same for both cameras, one would expect a corresponding improvement in distance precision for the XZ422 camera. However, this precision improvement is not apparent, meaning the XZ422 camera must have worse phase measurement precision than the SR4000 in our test conditions. It is important to note that the phase measurement precision is dependent on both sensor and camera design, and the operating conditions (as discussed in section 2.2 above), so these results may change under different conditions.

In the imaging (x and y) dimensions however, the SR4000 is around an order of magnitude better in precision. The SR4000 has a slightly higher resolution, meaning more pixels per target and more effective averaging, but this cannot account for all of the difference observed. The results would suggest that the XZ422 has a nosier image, which could be due to the physical design of the camera, but also could arise from the much shorter integration time used (discussed in section 4.3 above).

Accuracy was evaluated by finding the residuals between the aligned range camera measurements and the reference photogrammetric measurement. These results can be seen visually in figure 4 and are summarised in table 3.

	x (mm)	y (mm)	z (mm)
SR4000			
Average accuracy	21.956	12.667	3.909
Best accuracy	1.513	1.190	0.030
Worst accuracy	40.180	24.580	9.403
XZ422			
Average accuracy	3.111	3.033	3.988
Best accuracy	0.116	0.040	0.327
Worst accuracy	6.151	5.690	7.520

Table 3. Range camera one-sigma accuracy results.

a)

b)

c)
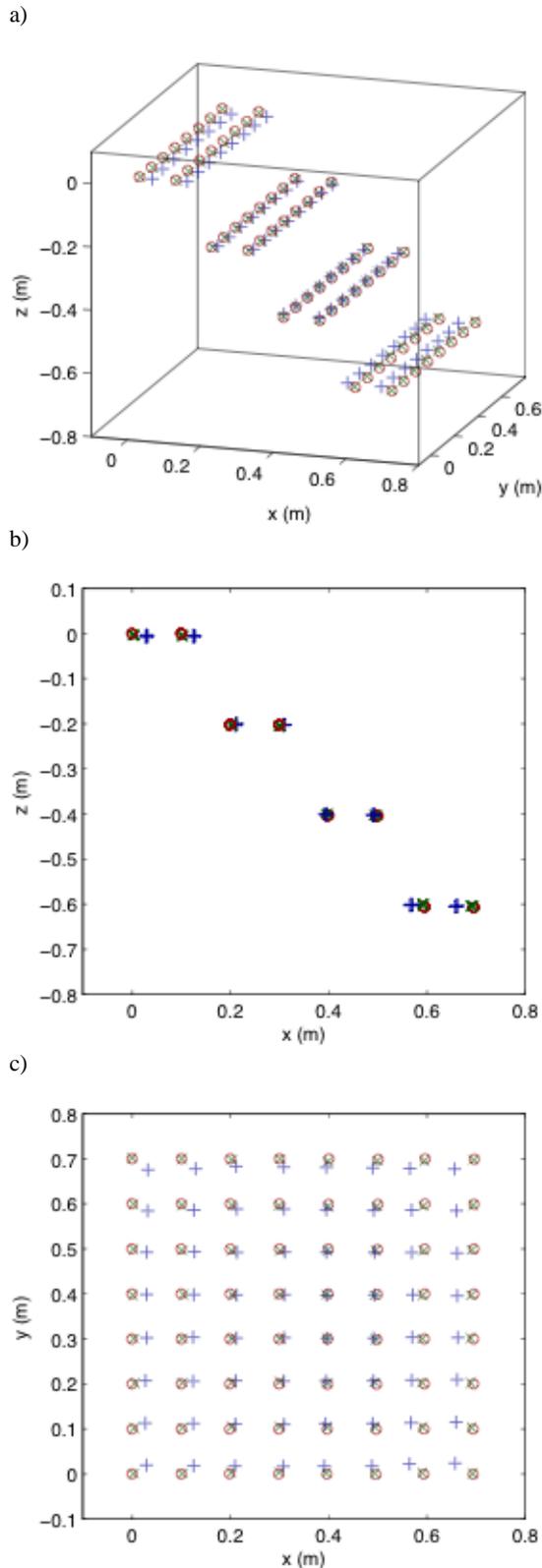
Figure 4. Accuracy plot of range camera acquisitions from the SR4000 (+) and XZ422 (×) compared to photogrammetric survey (O).

Over all, the XZ422 camera performs better than the SR4000. Although the SR4000 is comparable in distance (z dimension), as can be seen in figure 4b, it suffers from radial lens distortion, as seen in figure 4c. The XZ422 has similar accuracy errors in the spatial imaging (x and y) dimensions as it does in the distance (z) dimension; this is partially due to less lens distortion (evident in the raw images) as well as the manufacturer's lens calibration providing a better correction.

The XZ422 is the only camera tested that provided an indication of actual modulation frequency, which was 43.92 MHz in our case, equating to less than 0.2% error. The camera was placed approximately 1.5 m from the test object, so any accuracy errors introduced, prominently in the z dimension, by this clock drift are much less than those observed. Hence, other sources of accuracy error are dominant.

An indication of repeatability was obtained in a similar way to the precision experiment, but instead of leaving the camera still, it was moved between acquisitions. A series of 16 images, each from a different viewing location, were acquired with each camera. For comparison purposes, the standard deviation was calculated for each target location (separately in each dimension) across all images, and then averaged over all targets. These results are summarised in table 4. Note that cameras were moved over a range of approximately one metre across the x-axis, and approximately 1.5 m along the z-axis. The cameras were not moved substantially along the y-axis, so the y-axis repeatability values may be understated.

	x (mm)	y (mm)	z (mm)
SR4000			
Average repeatability	5.98	2.24	6.54
Best repeatability	2.05	0.70	3.45
Worst repeatability	10.66	6.12	10.96
XZ422			
Average repeatability	2.08	1.48	2.97
Best repeatability	1.31	0.85	1.33
Worst repeatability	3.79	2.48	4.66

Table 4. Repeatability of measured target locations between different viewing angles for both cameras. All values are in millimetres.

As expected, the SR4000 exhibited worse repeatability, most likely due to the poor calibration, but the values are still much lower than the accuracy errors (in table 3). Because the cameras were moved significantly, in relative terms, along the z-axis, approximately doubling the distance from the object, the object became appreciably smaller in the field of view. Since the object filled a smaller region of the sensor, the calibration errors may have become less significant. It should also be noted that when the cameras were further from the object, the targets covered fewer pixels, leading to less averaging and worse precision.

When interpreting these results, it is important to consider that this is only one specific example of each of these cameras, and may not be indicative of a typical camera. In particular, the SR4000 has previously been returned to the manufacturer for repair, and although it seems unlikely, it is possible that the calibration for this camera was invalidated.

5. CONCLUSIONS

The measurement precision and accuracy of two off-the-shelf range cameras, the Mesa Imaging SR4000 and Canesta XZ422, have been experimentally investigated as an alternative for close range metrology. The results show that multi-frame and multi-pixel averaging can provide sub-millimetre measurement precisions. This level of precision may be suitable for a number of applications, especially when only one projection/viewing position is available and measurements are required in a matter of seconds. The accuracy of the camera system, however, is lacking. Some promising results have been recorded with sub-millimetre best-case accuracies with the XZ422 camera. The SR4000 camera was found to be sorely lacking in radial lens distortion calibration, which could potentially be rectified with ease. This would significantly improve the imaging dimension accuracies, but have only a minor impact on improving the distance accuracy.

In this paper we used circular targets to obtain multiple pixel averages and improve the measurement accuracy and precision. It is important to note that a similar improvement could be obtained with full resolution using some knowledge or assumptions about the object being measured. For example, if rate of curvature limits are known, high spatial frequency distance measurement noise can be removed quite effectively with signal and image processing techniques (Bauer et. al., 2009).

Assuming the sensor technology will continue to develop, it is not unreasonable to expect VGA resolution range cameras in the near future. This would provide a four-fold increase in both horizontal and vertical resolution, which along with appropriate lens distortion calibration, should improve the imaging dimension accuracies to the sub-millimetre region. It is also not unreasonable to expect that modulation frequencies will increase as the technology matures, providing an improvement in distance measurement precision.

6. REFERENCES

Bauer, F., Schöner, H., Dorrington, A. A., Heise, B., Wieser, V., Payne, A. D., Cree, M. J., and Moser, B., Submitted 2009. Image Processing for 3D-Scans Generated by TOF Range Cameras, *IEEE/OSA Journal of Display Technology*.

Büttgen, B., Oggier, T., Lehmann, M., Kaufmann, R., and Lustenberger, F., 2005. CCD/CMOS Lock-In Pixel for Range Imaging: Challenges, Limitations and State-of-the-Art. In: *1st Range Imaging Research Day*, Zurich, Switzerland, pp. 21-32.

Büttgen, B., and Seitz, P., 2008. Robust Optical Time-of-Flight Range Imaging Based on Smart Pixel Structures. In: *IEEE Transactions on Circuits and Systems I: Regular Papers*, 55(6), pp. 1512-1525.

Dorrington, A. A., Cree, M. J., Payne, A. D., Conroy, R. M. and Carnegie, D. A., 2007. Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera, *Meas. Sci. Technol.* 18, pp. 2809–2816.

Dorrington, A. A., Cree, M. J., Carnegie, D. A., Payne, A. D., Conroy, R. M., Godbaz, J. P., and Jongenelen, A. P. P., 2008. Video-rate or high-precision: a flexible range imaging camera. In: *SPIE 6813 - Image Processing: Machine Vision Applications*, San Jose, CA, USA, pp 681307.

Falie, D., and Buzuloiu, V., 2009. Distance errors correction for the Time of Flight (ToF) Cameras. In: *Fourth European Conference on Circuits and Systems for Communications (ECCSC'08)*, Bucharest, Romania, pp. 121-124.

Guðmundsson, S. Á., Aanæs, H., and Larsen, R., 2007. Environmental Effects on Measurement Uncertainties of Time-of-Flight Cameras. In: *International Symposium on Signals Circuits and Systems - ISSCS*, pp. 113-116.

Lange, R., 2000. 3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology, Ph.D. dissertation, University of Siegen.

Larkins, R. L., Cree, M. J., Dorrington, A. A., and Godbaz, J. P., 2009. Surface projection for mixed pixel restoration. In: *Image and Vision Computing New Zealand (IVCNZ '09)*, Wellington, New Zealand, pp. 431-436.

McClure, S. H., Cree, M. J., Dorrington, A. A., and Payne, A. D., 2010. Resolving depth-measurement ambiguity with commercially available range imaging cameras. In: *SPIE 7538 - Image Processing: Machine Vision Applications III*, CA, USA, pp. 75380K.

MESA Imaging, 2009. "SR4000 Data Sheet Rev. 3.0", Zürich, Switzerland. http://www.mesa-imaging.ch/dlm.php?fname=pdf/SR4000_Data_Sheet.pdf (accessed 13 April 2010).

Mynbaev, D., and Scheiner, L., 2001. *Fiber-optic communications technology*, Prentice Hall Upper Saddle River, NJ, USA.

Payne, A. D., Dorrington, A. A., Cree, M. J., and Carnegie, D. A., 2008. Improved Linearity Using Harmonic Error Rejection in a Full-Field Range Imaging System. In: *SPIE 6805 - 3D Image Capture and Applications VII*, pp. 68050D.

Payne, A. D., Jongenelen, A. P. P., Dorrington, A. A., Cree, M. J., and Carnegie, D. A., 2009. Multiple Frequency Range Imaging to Remove Measurement Ambiguity. In: *Optical 3-D Measurement Techniques IX*, Vienna, Austria, Vol. 2, pp. 139-148.

PMD Technologies, 2005. "PMD [vision]® 3k-S Datasheet", Siegen, Germany. http://www.pmdtec.com/inhalt/download/documents/PMDvision3k-S_000.pdf (accessed 7 July 2008)

PMD Technologies, 2009. "PMD [vision]® CamCube 2.0 Datasheet V. No. 20090601", Siegen, Germany. http://www.pmdtec.com/fileadmin/pmdtec/downloads/documentation/datasheet_camcube.pdf (accessed 22 July 2009)

7. ACKNOWLEDGEMENTS