

S. H. McClure, M. J. Cree, A. A. Dorrington, and A. D. Payne, "Resolving depth-measurement ambiguity with commercially available range imaging cameras," *Proc. SPIE-IS&T Electronic Imaging*, edited by D. Fofi, K. S. Niel, SPIE vol. 7538, pp. 75380K, (2010).

<http://dx.doi.org/10.1117/12.838786>

Resolving depth measurement ambiguity with commercially available range imaging cameras

Shane H. McClure, Michael J. Cree, Adrian A. Dorrington*, Andrew D. Payne
Dept. Engineering, The University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

ABSTRACT

Time-of-flight range imaging is typically performed with the amplitude modulated continuous wave method. This involves illuminating a scene with amplitude modulated light. Reflected light from the scene is received by the sensor with the range to the scene encoded as a phase delay of the modulation envelope. Due to the cyclic nature of phase, an ambiguity in the measured range occurs every half wavelength in distance, thereby limiting the maximum useable range of the camera.

This paper proposes a procedure to resolve depth ambiguity using software post processing. First, the range data is processed to segment the scene into separate objects. The average intensity of each object can then be used to determine which pixels are beyond the non-ambiguous range. The results demonstrate that depth ambiguity can be resolved for various scenes using only the available depth and intensity information. This proposed method reduces the sensitivity to objects with very high and very low reflectance, normally a key problem with basic threshold approaches.

This approach is very flexible as it can be used with any range imaging camera. Furthermore, capture time is not extended, keeping the artifacts caused by moving objects at a minimum. This makes it suitable for applications such as robot vision where the camera may be moving during captures.

The key limitation of the method is its inability to distinguish between two overlapping objects that are separated by a distance of exactly one non-ambiguous range. Overall the reliability of this method is higher than the basic threshold approach, but not as high as the multiple frequency method of resolving ambiguity.

Keywords: Range imaging, time-of-flight, ambiguity, phase unwrapping, image processing, phase

1. INTRODUCTION

Over the last few years interest in range imaging cameras based on the time-of-flight (TOF) principle has been growing. Recent developments have improved power consumption, image resolution, precision and accuracy of these cameras. One of the unique features of this technology is that it produces both an intensity and full-field depth image simultaneously at video frame rates.

Commercial range imaging cameras use infrared LEDs to actively illuminate the field of view with amplitude modulated light. When light is reflected back from objects in the scene, the phase of the modulation envelope of the reflected light is proportional to the distance travelled. To accurately convert the phase information to depth, four samples are typically taken, C_1 to C_4 , which are each phase shifted by 90 degrees. The active amplitude A and phase θ for each pixel in the sensor array is given by¹,

$$A = \frac{\sqrt{[C_4 - C_2]^2 + [C_1 - C_3]^2}}{2} \quad (1)$$

$$\theta = \tan^{-1} \left(\frac{[C_4 - C_2]}{[C_1 - C_3]} \right) \quad (2)$$

*adrian@waikato.ac.nz; phone +64-7-858-5062; fax +64-7-838-4835

The intensity image is similar to what you would expect from a traditional camera and is shown in figure 1 (a). This image is in grayscale and records the infrared response of the scene as most of the background light has been removed through internal processing and an external filter.

The phase information is directly related to the radial distance r to objects in the scene and can be converted to a depth image shown in figure 1 (b) by

$$r = \frac{\theta}{2\pi} \cdot \frac{c}{2f_m}, \quad (3)$$

where c is the speed of light and f_m is the modulation frequency of the camera.

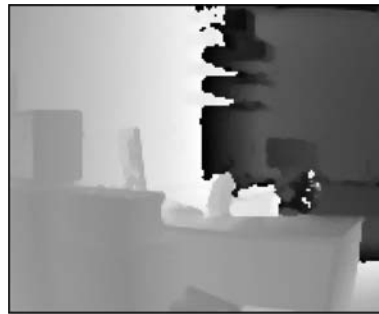
One of the main limitations intrinsic to this technology is the occurrence of ambiguity in the resultant range image. This is a fundamental problem that must be solved for many applications that do not have a controlled scene. Range ambiguity occurs when the camera is used to measure distances beyond half the modulation wavelength, typically at 3 to 10 m. Due to the cyclic nature of phase, as soon as 360 degrees is reached, the phase wraps back to 0 degrees. Because depth is calculated from this wrapped phase, the result is a wrapped depth image.

The larger the modulation frequency of the camera, the shorter the wavelength of the modulation envelope is, and the shorter the non-ambiguous range becomes. One simple approach to increase the ambiguity distance is to decrease the modulation frequency. Unfortunately, as modulation frequency is decreased, measurement uncertainty is increased, reducing the precision of the measured depth².

To illustrate this wrap around effect, imagine an object 7 m in front of the camera. If we use a camera modulated at 30 MHz, the non-ambiguous depth within the scene is limited to 5 m. In this case, the object's true depth at 7 m is wrapped around after 5 m, and appears to be at 2 m. Figure 1 (b) illustrates the wrapping of a depth image. The depth is represented with black at 0 m, progressing to white at a distance of 5 m. The white to black transition in the center of the image indicates where the phase wraps around on a continuous surface.



Intensity image (a)



Depth image (b)



Digital photo (c)

Figure 1. A typical office scene captured using the SR4000 in both intensity (a) and depth (b). Because this scene extends beyond the non-ambiguous range of the camera, the depth is wrapped. For reference, a digital photograph of the scene is shown in (c).

The aim of this paper is to explore how well ambiguity can be resolved using only the information supplied in a standard capture, namely, to estimate unwrapped depth using only the wrapped depth and intensity images. This has the advantage of requiring no additional capture time, and can be used with any existing commercially available range imaging camera. This solution is intended to be used for navigation of a mobile robot from range imaging data.

In this paper, captures have been made with two commercially available range imaging cameras. The first camera is the Mesa Imaging SR4000 (Zurich, Switzerland). This camera has a resolution of 176 by 144 pixels and was used for the majority of range images in this paper. The second camera used is the Canesta XZ422 (Sunnyvale, CA, USA). This has a resolution of 160 by 120 pixels and was used to capture the images in figure 12.

In section 2 we outline various methods currently used to resolve depth ambiguity. Our proposed method is described step by step in section 3. The performance and reliability of this method is then evaluated and compared to both basic intensity thresholding and a multiple frequency method of resolving ambiguity in section 4. Section 5 presents our conclusions and describes possible future work.

2. RESOLVING AMBIGUITY

2.1 Multiple Frequencies

Previous research at the University of Waikato investigated the simultaneous use of multiple modulation frequencies to extend the non-ambiguous range³. This method superposes two different modulation frequencies within a single capture, requiring just 1 additional sample (for a total of 5 samples) to resolve the range images at both frequencies. Experiments showed that this can give an accurate and precise depth measurement with only a small cost in capture time³. Unfortunately this solution uses specialized hardware and is not supported by commercially available range imaging cameras. For this reason, another approach has been sought that can be applied to the current camera technology.

One possible method to extend the non-ambiguous range is to use multiple modulation frequencies consecutively. In this case two captures are made, one after another, at different modulation frequencies so that each has a different non-ambiguous range. The depth measurements from both captures can then be compared to find the correct depth that will be common to both. This method is able to extend the non-ambiguous range to the product of the two individual ranges. Using this approach requires a total of eight samples and therefore twice the capture time for a single depth measurement. This reduces the frame rate of the camera and increases the amount of motion blur in the final capture.

2.2 Phase unwrapping

The range imaging ambiguity problem is essentially a phase unwrapping problem, however range data may involve abrupt changes in depth for which traditional phase unwrapping algorithms are not well suited. Two classes of methods are available for two dimensional phase unwrapping: path following methods and minimum norm methods.

Range image data from a TOF camera has been tested using the Goldstein 2D branch cut method⁴. This showed that while the method sometimes failed, a general improvement could be made. The main limitation with phase unwrapping techniques is the cost required to compute them. This makes them unsuitable for processing data in real time.

2.3 Intensity threshold

As range imaging cameras rely on active illumination, the active intensity of each pixel gives an initial estimate of its depth (the closer an object, the brighter it appears). This relationship is shown in figure 2 (a) and occurs because intensity drops at a rate of distance squared for a surface with Lambertian reflectance⁵. This is a well known concept and it has been suggested that the intensity measurement can therefore be used to help resolve depth ambiguity⁶.

If this ideal relationship was observed with a range imaging camera, we would get the result shown in figure 2 (b). In this figure we have wrapped the measured depth using a non-ambiguous range of 5 m. Any objects with a true depth within the non-ambiguous range (0 – 5 m) have their depth recorded correctly and are said to be in the first ambiguity bracket. Objects in the second ambiguity bracket (5 – 10 m) have their depth wrapped once, while objects in the third ambiguity bracket (10 – 15 m) have their depth wrapped twice.

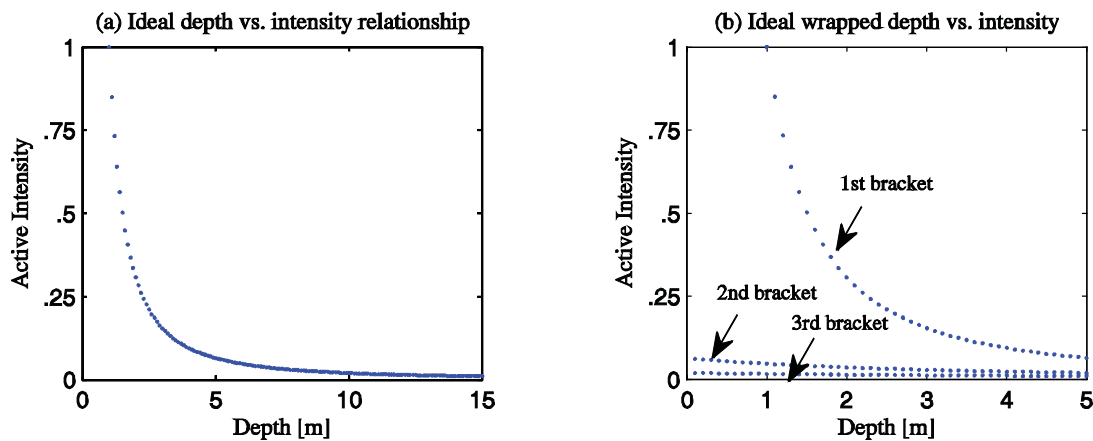


Figure 2. The ideal depth versus intensity relationship is shown in (a). When this is measured with a non-ambiguous range of 5 m, the depth wraps around into a different ambiguity bracket every 5 m (b).

In reality, the intensity I of each pixel is dependent on many factors, including depth, reflectivity and surface orientation⁷ and is given by

$$I = \Gamma \cdot \frac{\mathbf{n} \cdot \mathbf{l}}{r^2}, \quad (4)$$

where Γ is pixel reflectance, \mathbf{n} is the surface normal, \mathbf{l} is a unit vector from the camera light source and r is the depth to the pixel. This intensity value is equivalent to the amplitude calculated in equation 1. When all these factors are in effect, the resulting relationship of depth versus intensity of a typical scene is shown in figure 3. Here, the inverse square law is still quite apparent however the effects of reflectance and surface normal direction obscure its effect to a marked degree.

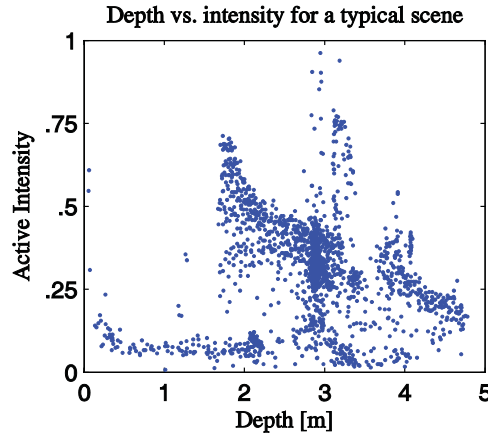


Figure 3. A typical depth-intensity relationship of a scene captured using the SR4000. Only one tenth of the total data points for a single capture are shown here to improve clarity.

This relationship has been demonstrated to help extend the range of a SR3000 range imaging camera⁸. The algorithm first normalized the intensity image for the camera optics and light source. A constant threshold was then applied to the entire image on a pixel by pixel basis to determine which pixels were beyond the first ambiguity bracket. If the intensity of the pixel was below the threshold value, the pixel had a value equal to the non-ambiguous range added to its depth. This approach is able to increase the effective range of the camera however it is very sensitive to the reflectance of objects which can cause the depth to be unwrapped incorrectly.

3. METHOD

Our proposed method aims to expand on the intensity threshold concept to form a more intelligent and robust way to classify pixels that lie beyond the first ambiguity bracket. The key focus of the method is to be able to resolve ambiguity in the presence of variations in an object's reflective properties.

In our algorithm, we segment the scene by detecting edges in the depth image. These edges are further processed to form completely closed groups that outline different objects in the scene. The average intensity of each group can then be used to decide whether each object is within the first ambiguity bracket or not, eliminating the effect of local variations in intensity. This enables objects that would normally be broken into pieces because of changes in material, color, texture etc to be classified correctly. In addition, multiple objects that are connected together, such as objects on a table, can be grouped together and their overall average intensity used to classify the entire group more robustly. The individual steps to achieve this are outlined in the algorithm below.

3.1 Edge detection

The initial step required to segment different objects in the scene is to detect the edges in the range image. When detecting edges from intensity data, edges can be caused by abrupt changes in color, depth, texture and surface reflectance. Edge detection from range images is much more powerful for segmenting objects because edges are known to only correspond to changes in depth.

To detect edges, the first step was to calculate the gradient of the depth image. This gradient image gives an indication of the magnitude of the edges in the scene. There are many possible ways to calculate this, and the external morphological half gradient was used. The external gradient is found by calculating the dilation of the depth map, shown in figure 4 (a) and subtracting the original depth image from it. This results in a very fast technique that selects the edge pixels which are closest to the camera, preferentially selecting edge pixels that are beyond the first ambiguity bracket. The gradient image is shown in figure 4 (b) where the dark lines represent areas with a large change in depth.

In order to convert this information into a more useable form, the gradient image is thresholded to produce a mask of all edges over a certain magnitude. It was found that the threshold value could be set anywhere from 20 cm to 100 cm with little change in results, however 40 cm was found to be good value to minimize false groupings. This mask, shown in figure 4 (c) is a binary image that shows the borders between separate objects. Any gaps in the borders however will cause two otherwise separate objects to join. To stop this from happening, an edge linking technique needs to be used.

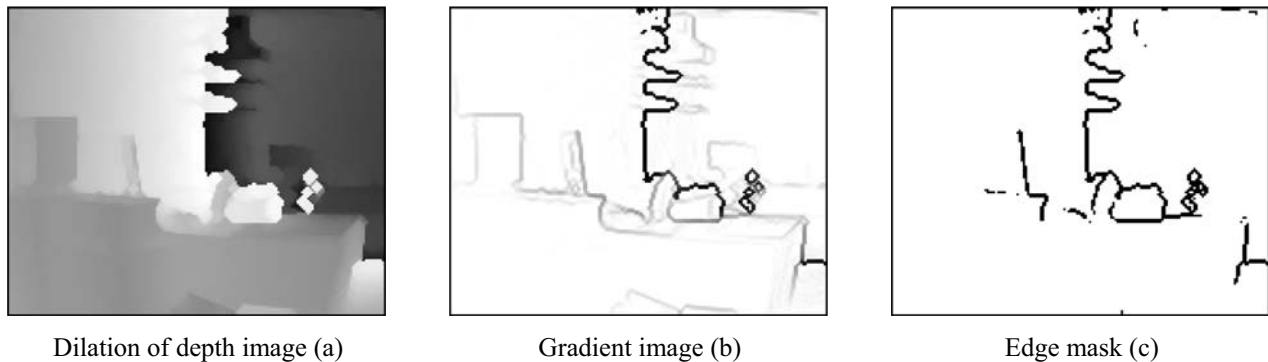


Figure 4. Image (a) shows the dilation of the depth image from figure 1. This is used to calculate the gradient image (b) which gives an indication of edge strength in the scene. The edge mask (c) indicates which pixels are large enough to be considered edges.

3.2 Edge linking

The purpose of edge linking is to bridge any gaps in the edge mask. These breaks can either be formed from discontinuities in the range image itself, or from a poorly chosen edge detection threshold. The edge linking algorithm designed for this algorithm was based on the watershed transform. The watershed transform often over segments an image. To overcome this weakness, it is often used with pre-processed data or images with added markers.

To perform the edge linking stage, a morphological distance transform was taken from the edge mask. This function sets the value of each pixel to the Euclidean distance from the closest edge pixel. The complement of this image was then taken to form an artificial image to use in the watershed transform, shown in figure 5 (a). The watershed transform, based on the Fernand Meyer algorithm⁹, was applied to produce the image in figure 5 (b). This result is similar to the original edge mask, except there are now complete edges that form distinctly separate groups. Each of these groups is intended to follow the outline of an object, or a section of an object.

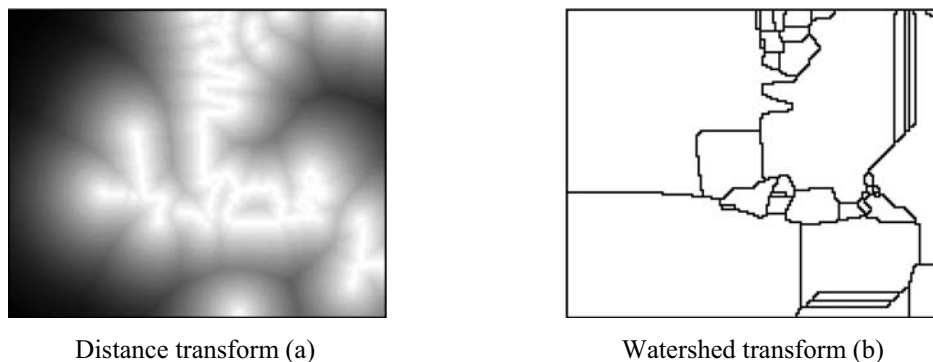


Figure 5. The distance transform (a) shows the distance of each pixel to the nearest edge in the edge mask from figure 4. This serves as an artificial image for the watershed transform (b) to form distinctly separate groups.

3.3 Group reduction

Despite using the artificially created image in the previous stage, the result is still over segmented. To help mitigate this effect, a technique to join some of the groups was developed.

The average depth from each group from the watershed transform was calculated and used to form a depth mosaic image shown in figure 6 (a). In order to analyze how similar each group was to its neighbors, some measure of the edge strength between groups was required. The morphological gradient of this image was calculated by taking the difference between the dilation and erosion of the image. The result is shown in figure 6 (b), where the areas of homogenous depth are shown in white, and the edge strength of each border is shown in shades of gray.

This border strength image is thresholded with a global constant to only select borders above a certain strength. The threshold value chosen here was the same as used in section 3.2. This has the effect of joining many of the groups that the watershed algorithm erroneously split into multiple groups. The remaining groups are shown in figure 6 (c) where it contains a smaller number of groups than the original watershed transform.

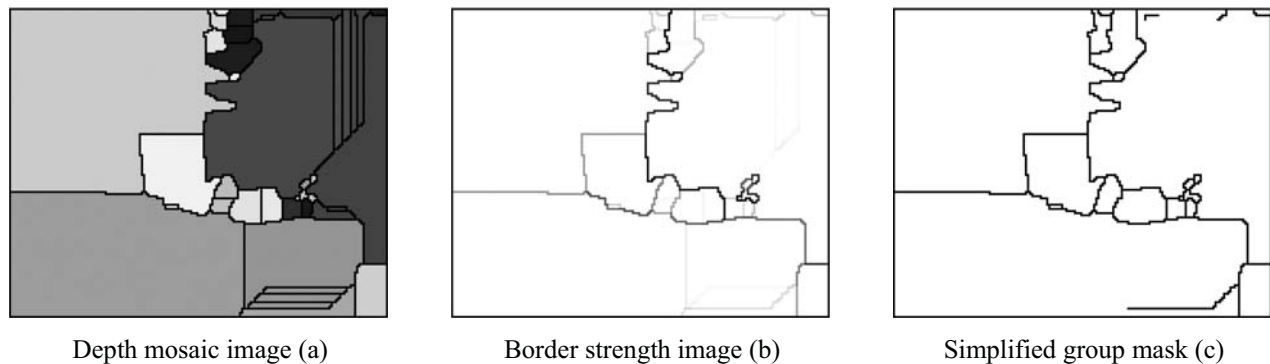


Figure 6. The depth mosaic image (a) was found by averaging the depth of each group found in figure 5. The difference in depth between adjacent groups was calculated in (b). This border strength image was thresholded to create the simplified group mask (c).

3.4 Group analysis and intensity threshold

To actually unwrap the depth map, any pixels beyond the first ambiguity bracket needed to be identified. This was done on a group by group basis. Any group with an average intensity below this threshold would be labeled as belonging to the second ambiguity bracket. This threshold was chosen manually for each scene.

To perform the threshold, the intensity of each pixel was first normalized by multiplying the intensity by its distance squared. This was required to compensate for the drop in active light intensity with depth, as explained in section 2.3, allowing a constant threshold to be applied. In the case of the SR4000, the CONV_GRAY function provided by the manufacturer was used to perform this adjustment. This setting normalizes the image for distance and corrects the data for the non-uniform illumination of its light source. This adjusted intensity value can be used to create an intensity mosaic image shown in figure 7 (a). In this image the value of each pixel is set to the average intensity of its group.

Figure 7 (b) shows the results of applying a constant threshold to the intensity mosaic image. Any objects with an average intensity below the threshold level are deemed to belong to the second ambiguity bracket and are shown in black. The image in figure 7 (c) shows a confidence map generated by taking the absolute difference between each group's intensity and the threshold value. The darker the group, the less confident the decision made about its depth. This image has been scaled to assist viewing but can otherwise be used to indicate the confidence with which each group has been classified.

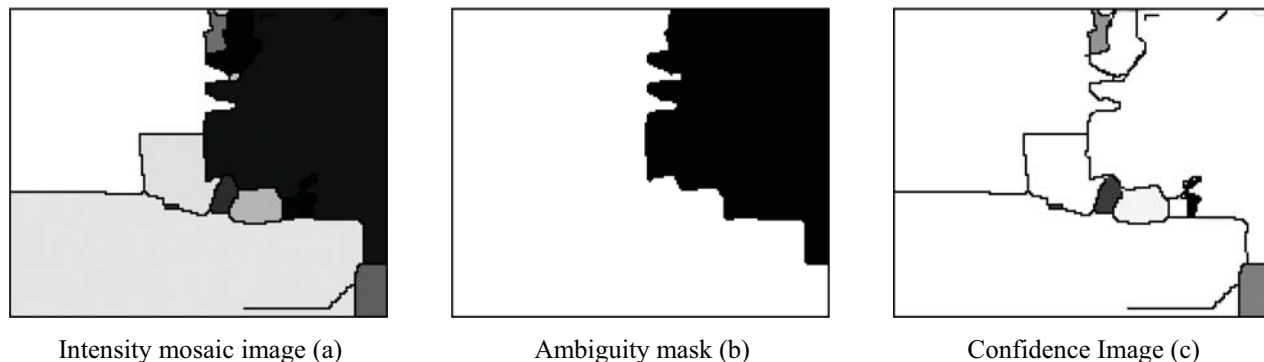


Figure 7. The intensity mosaic image (a) is found by averaging the intensity for each group. This is thresholded to find pixels in the second ambiguity bracket (b). The confidence of this decision can easily be computed and is shown in (c).

3.5 Phase unwrapping and final processing

To unwrap the ambiguous data marked above, a value equal to the non-ambiguous range is added to these pixels. This process unwraps the depth shown in figure 8 (a) to that of figure 8 (b). Because of the edge linking technique, the original edges have not been completely preserved. This manifests itself as the line of mislabeled pixels across the border of the ambiguity mask. To correct these artifacts, a final stage of post processing is used.

To detect the remaining pixels that were not adjusted, a bottom hat filter is used. This finds the difference between the closing of the image and the original image itself, extracting any dark pixels in regions of bright pixels. In a range image, these isolated dark pixels correspond to pixels in front of their neighbors. The bottom hat image is thresholded to detect any isolated pixels with more than 4 m variation. Using this image as a mask, the pixels have a distance equal to the non-ambiguous range added to them, correcting their depth.

This step is then repeated using a top hat filter to extract any bright pixels in regions of dark. These isolated bright pixels correspond to pixels that have been labeled as belonging to the second ambiguity bracket, while they should have been in the first bracket. The result of this process is shown in figure 8 (c) where the depth map has corrected all stray values.

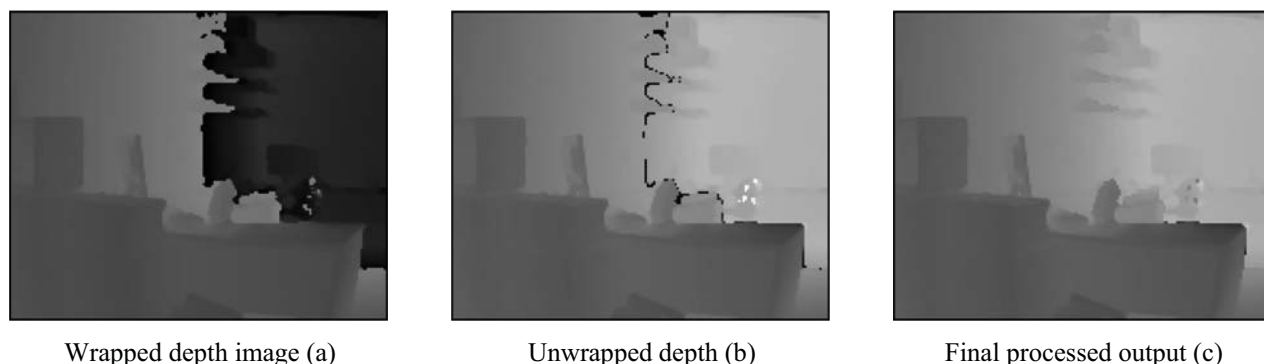


Figure 8. The wrapped depth in (a) is unwrapped in (b) using the proposed algorithm described in this method. Some final processing using top hat and bottom hat filters clean up outlying pixels to produce the final depth map (c).

4. RESULTS AND DISCUSSION

To test our algorithm, a variety of office and laboratory scenes were captured that were typical of our intended application of navigating a mobile robot from range imaging data. These scenes were unwrapped using our method and a basic intensity threshold technique. The performance of the algorithm was then compared with an alternative multiple frequency approach.

4.1 Proposed method versus basic intensity threshold

When compared to the basic intensity threshold approach there are some key advantages and disadvantages to consider. Figure 8 shows the results of unwrapping depth using both the basic intensity threshold approach and the proposed grouping algorithm. The first two columns show the raw data captured with the camera. The third column shows the unwrapped depth using the basic intensity threshold approach, and the fourth column shows the unwrapped depth using the proposed grouping method.

In both scenes, the basic intensity thresholding technique has mislabeled some of the pixels. Objects within the first ambiguity bracket with low reflectance have intensities below the threshold value, and are subsequently unwrapped into the second bracket. The equivalent problem occurs with high reflectance objects that are in the second ambiguity bracket. Because the wall is almost square-on to the camera, it reflects the light very well and has been set to the first ambiguity bracket using this basic approach.

By thresholding each pixel against the average intensity of its group, the proposed algorithm is much less sensitive to local variations in reflectance. In both of these scenes the dark pixels in the foreground and light pixels in the background have been correctly unwrapped using the proposed grouping method. One exception is illustrated by the computer monitor near the center of the first scene in figure 9. Here the computer monitor was not grouped with the rest of the table, and because the entire monitor has a very low reflectance, it was incorrectly unwrapped into the second ambiguity bracket.

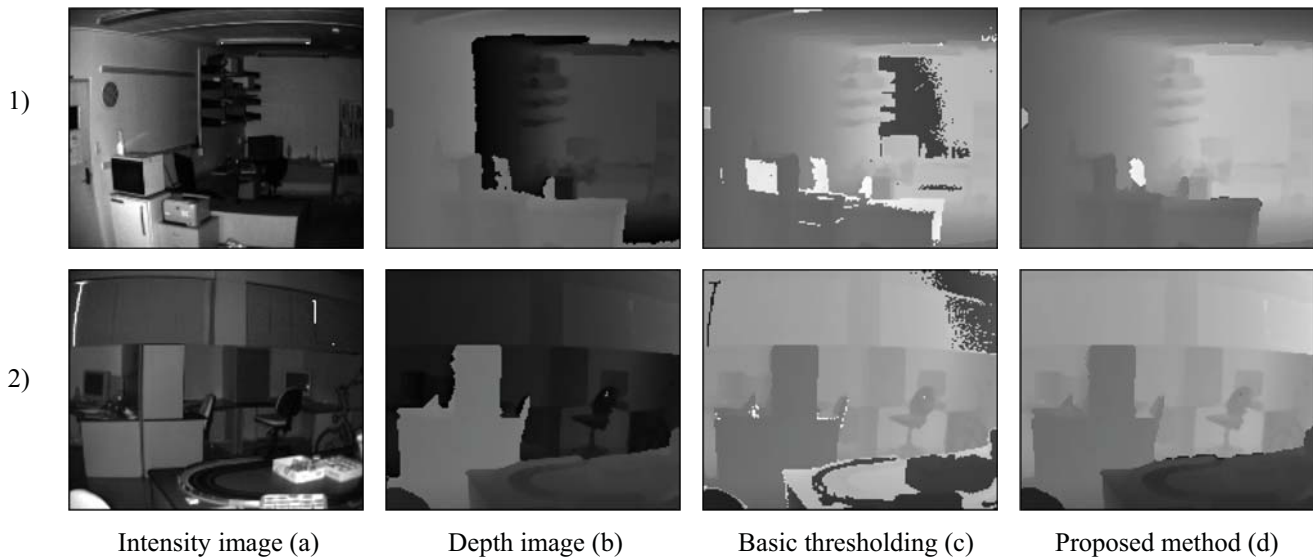


Figure 9. Two typical scenes that contain wrapped depth are shown. Columns 1 and 2 show the raw data captured, while columns 3 and 4 show the unwrapped depth images using the basic intensity threshold and the proposed grouping algorithm.

One major drawback of the proposed grouping method is the amount of time it takes to run the algorithm. Basic intensity thresholding on a per pixel basis is very quick and can be run at the frame rate that the camera is operating (20 to 30 fps). The grouping algorithm however was only able to process data at a rate of 5.9 fps. This was achieved on a 2.4 GHz quad core Pentium 4 processor with 4 GB of RAM running MATLAB 2008a. No effort was made to optimize the processing speed.

In order to group pixels into an object, we made the necessary assumption that connected pixels at similar depths belong within the same ambiguity bracket. This enables pixels to be joined, irrespective of their intensity (and therefore

reflectance). This assumption is not strictly correct however as two objects that are separated by a distance equal to the non-ambiguous range will also appear to be at identical depths. This can cause the algorithm to form incorrect groups and is a major limitation with the proposed grouping approach.

Figure 10 shows a scene where part of the table in the foreground is 5 m (one non-ambiguity range at 30 MHz) in front of the back wall. The grouping algorithm joined the table and the back wall together and this resulted in an incorrectly unwrapped group. It is worth noting that even in this case the rest of the image has been unwrapped correctly.

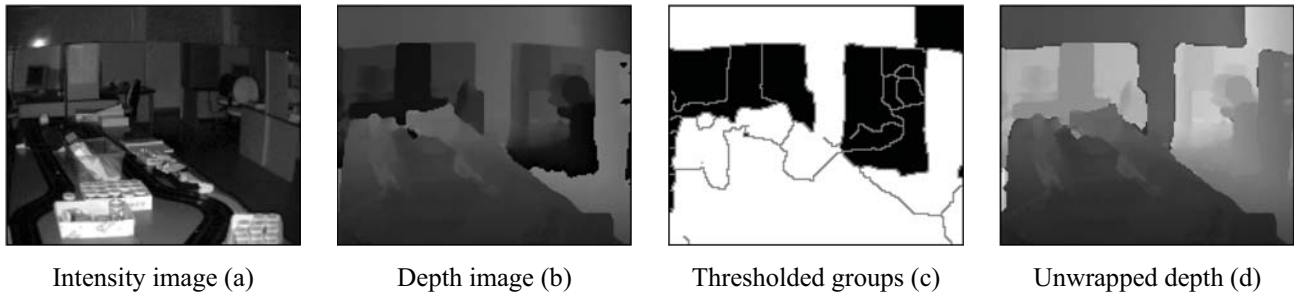


Figure 10. Images (a) and (b) show a scene where the back wall and table overlap with a section separated by a distance of 5 m, equal to the non-ambiguity range. These two groups are joined in (c) and the resulting unwrapped depth (d) contains mislabeled data.

4.2 Surface normals

Before objects are thresholded, the intensity is normalized for both the drop of intensity with depth, see section 2.3, and the characteristics of the lens and light source. At this stage it would also be useful to normalize the data for the shape of its surface, eliminating one more source of variation in the intensity. Surface normals have been previously shown to provide useful information in a similar application to improve the accuracy of range images⁷. In our research two techniques were used to calculate surface normals: the cross product of two tangent vectors and least squares plane fitting. The results from these two methods were processed using morphological operators to reduce the effect that object edges have on the calculations.

In order to improve the thresholding results, objects with a large angle of incidence to the incoming light need to have their intensity increased more than an object with a small angle of incidence. When this step was implemented however, it was found that more distant objects tended to have their surface normals over-estimated. One possible explanation is the increase in spatial variation of pixels as the pixel sampling area increases with distance. This had the opposite of the desired effect, increasing the brightness of objects further away, making the thresholding problem more difficult. For this reason the surface normals were not calculated in the final algorithm.

4.3 Multiple ambiguity brackets

One special case is a scene where the camera can view beyond the first two ambiguity brackets. Figure 11 shows a corridor that extends over 20 m. The algorithm was used to correctly classify all the pixels beyond the first ambiguity bracket. This includes all pixels in the second, third and even fourth ambiguity brackets. Theoretically, the algorithm could then be reiterated to unwrap the next ambiguity bracket. In practice however, there is insufficient difference between the pixel intensity values to do this accurately. This occurs as the change in intensity due to the inverse square law has less effect than the variations in surface direction and reflectivity at these distances. This limitation means the method cannot accurately extend the range imaging camera's effective distance beyond twice its non-ambiguous range.

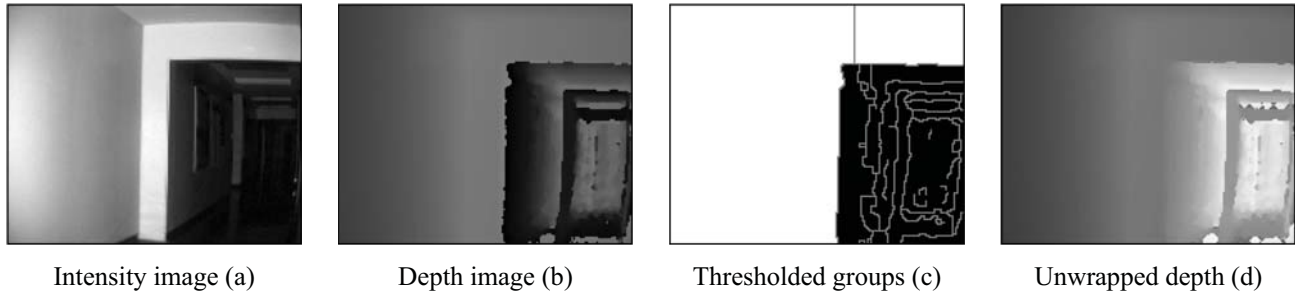


Figure 11. Images (a) and (b) show a corridor scene that extends across multiple ambiguity brackets. All groups beyond the first ambiguity bracket have been identified correctly in (c) however the unwrapped depth (d) still contains a wraparound from the third ambiguity bracket (10 to 15 m), and fourth ambiguity bracket (15 to 20 m), that cannot be accurately classified.

4.4 Inherent problem with proposed method and basic intensity threshold

One of the key limitations of both post processing techniques described above is their sensitivity to an object's reflectance properties. The grouping algorithm was able to reduce this sensitivity by a marked degree however there is no way to make either technique completely robust.

An interesting scene illustrates this point in figure 12 where a board has been set up 5 m in front of the back wall (one non-ambiguity range in front). Because the board reflects less light than the back wall, it appears at the same depth and the same intensity as captured by the range imaging camera in figure 12 (a) and (b). There is simply not enough information in these two images to correctly unwrap this scene and a technique such as multiple frequencies is required. For reference the scene has also been captured using a digital camera in figure 12 (c) where the board can be clearly seen in the visible spectrum.

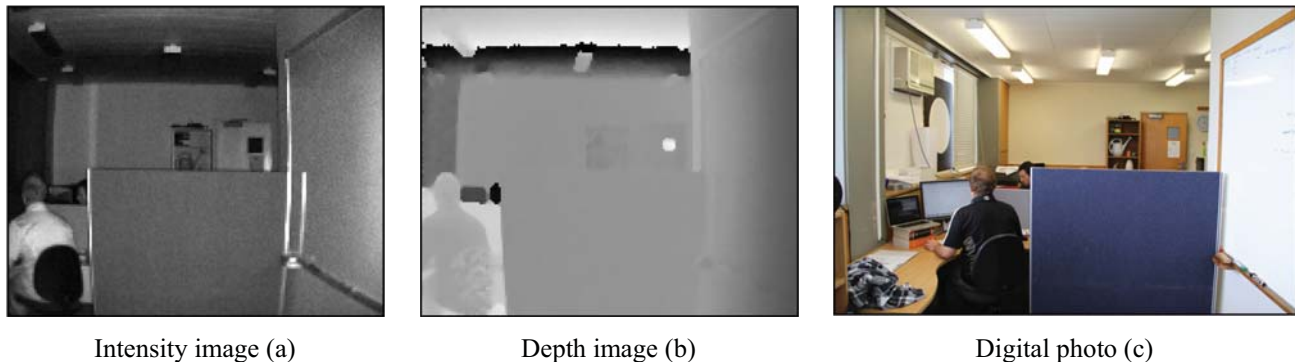


Figure 12. The active intensity image (a) and depth image (b) cannot be used to detect the difference between the board at 4 m and the back wall at 9 m (one non-ambiguity distance apart). A digital photograph (a) shows the scene in the visible spectrum for reference.

4.5 Proposed method versus a multiple frequency approach

Figure 13 shows two scenes captured using the Canesta XZ422 range imaging camera. The first two columns show the raw intensity and range information. The third column is the unwrapped range using the camera's multi-frequency mode, and the fourth column is the results using our proposed grouping method.

Looking at the first scene we can see that both techniques unwrap the depth accurately. Most of the extra noise in the multiple frequency approach can be attributed to the spatio-temporal technique used by the Canesta camera¹⁰ and can be cleaned up with filtering techniques similar to those in section 3.5.

In the second scene, we have added another object in front of the background at a distance equal to the non-ambiguous range. As these objects appear to be at the same depth, the proposed algorithm groups the foreground and background objects together, causing them to be unwrapped incorrectly. In comparison this scene does not affect the multiple frequency approach and the range image is still unwrapped correctly.

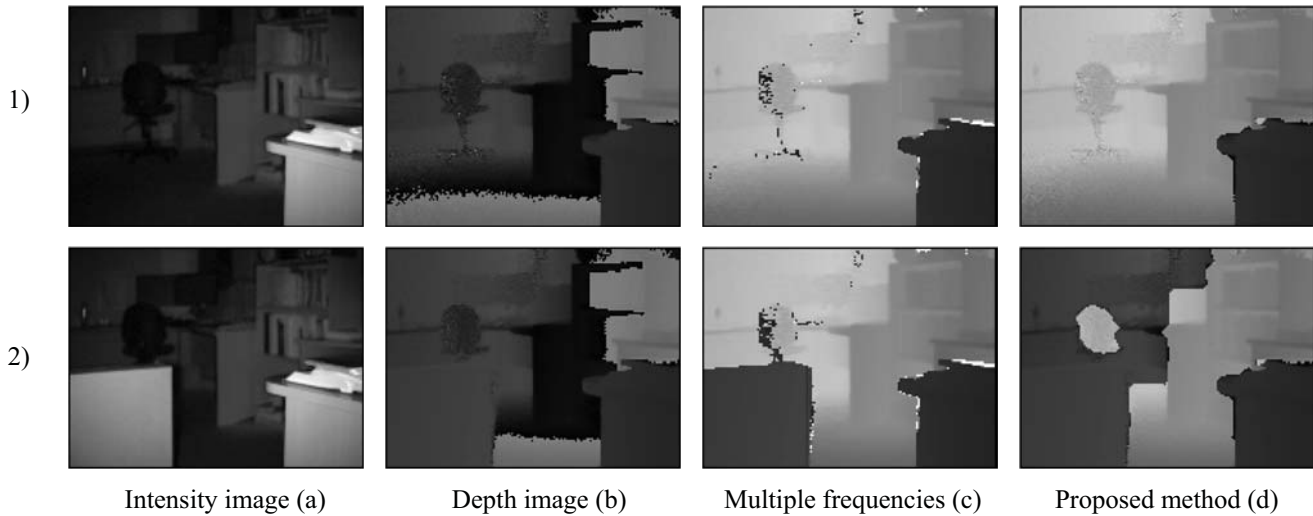


Figure 13. Two scenes captured using the Canesta XZ422 camera. In scene 1 both methods are able to unwrap the depth image correctly. In scene 2 an object was placed at a distance equal to the non-ambiguous range in front of the back wall. In this case the multiple frequency approach remained functional while the proposed method no longer worked correctly.

5. CONCLUSION

We have presented a practical solution to resolve depth ambiguity in range images. This was achieved using both the depth and intensity information available from existing commercially available range imaging cameras. This paper proposed a novel approach to determine ambiguity by first dividing the scene into groups of contiguous depth. Grouping the scene before performing thresholding significantly reduced the sensitivity to the reflectance of objects when compared to the basic intensity threshold technique.

This approach is very flexible as it can be used with any range imaging camera. Furthermore, the acquisition time is not extended, keeping the artifacts caused by moving objects at a minimum. The main limitation of this technique is how depth is unwrapped when objects are exactly one non-ambiguity distance apart. In these cases, foreground and background objects would be joined into a single object and subsequently mislabeled. This effect becomes more noticeable as the non-ambiguity range is decreased. This makes the algorithm most suited to cameras with a non-ambiguity range around 5 to 10 m, where the effective range can be extended to twice this distance.

Compared to the multiple frequency approach, the grouping algorithm is able to produce similar results for various different scenes that were tested. The grouping algorithm also has the advantage of requiring half as many captures to create a single depth and intensity pair, reducing capture time and motion artifacts. This makes it well suited to robot navigation where the camera is moving, and where a certain amount of false data is tolerable.

While the grouping algorithm does provide a viable solution for many applications, there are assumptions made about the scene that are not always correct. This means that the algorithm fails to unwrap certain special case scenes and as such will never be completely robust. The multiple frequency approach provides a more reliable solution to the depth ambiguity problem as it can be used for a wider variety of scenes. The multiple frequency approach is also able to extend the non-ambiguous range well beyond what the grouping method is capable of.

At present, the intensity threshold has been set manually in each scene. Possible future work could evaluate how robustly an appropriate threshold can be found using just the exposure settings of the camera and the intensity and depth images. Further improvements to the algorithm could be gained by researching alternative depth cues. These clues, such as object occlusion or surface normals, may become more important and useful with an increase in the resolution of range imaging cameras. The algorithm cannot process data at the frame rate that the camera, further work could be spent to simplify the algorithm and to increase its efficiency.

ACKNOWLEDGMENTS

The authors would like to thank the entire Chronoptics group for their contribution towards this project. This research is supported by the University of Waikato Strategic Investment Fund.

REFERENCES

- [1] Oggier, T., Lehmann, M., Kaufmann, R., Schweizer, M., Richter, M., Metzler, P., Lang, G., Lustenberger, F. and Blanc, N., "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger)," SPIE Proc Vol. 5249, 634-645 (2003).
- [2] Büttgen, B. and Seitz, P., "Robust Optical Time-of-Flight Range Imaging Based on Smart Pixel Structures," IEEE T CIRCUITS-I vol. 55, 1512-1525 (2008).
- [3] Payne, A. D., Jongenelen, A. P. P., Dorrington, A. A., Cree, M. J. and Carnegie, D. A., "Multiple Frequency Range Imaging to Remove Measurement Ambiguity," Optical 3-D measurement techniques IX, (2009).
- [4] Jutzi, B., "Investigations on ambiguity unwrapping of range images," ISPRS Vol. 38 prt 3/W8, 265-270 (2009).
- [5] Trucco, E. and Verri, A., [Introductory Techniques for 3-D Computer Vision], Prentice Hall PTR, Upper Saddle River, NJ, (1998).
- [6] MESA Imaging AG, "SR4000 user manual," version 0.1.2.2, Zurich, Switzerland, (2008).
- [7] Bohme, M., Haker, M., Martinetz, T. and Barth, E., "Shading constraint improves accuracy of time-of-flight measurements," IEEE CVPRW, 1-6 (2008).
- [8] Poppinga, J. and Birk, A., "A Novel Approach to Efficient Error Correction for the SwissRanger Time-of-Flight 3D Camera," LECT NOTES COMPUT SC Vol. 5399, 247-258 (2009).
- [9] Meyer, F., "Topographic distance and watershed lines," Signal Processing vol. 38(1), 113-125 (1994).
- [10] Canesta, "SDK, Reference Manual, 2IS," 12443-01 Rev 4, Sunnyvale, CA, (2009).