The 16th Electronics New Zealand Conference (ENZCon), Dunedin, New Zealand, 18-20 November, 2009

Advantages of 3D Time-of-Flight Range Imaging Cameras in Machine Vision Applications

Adrian A. Dorrington⋆, Cameron D.B. Kelly, Shane H. McClure,
Andrew D. Payne, and Michael J. Cree

⋆Department of Engineering, University of Waikato
Private Bag 3105, Hamilton, New Zealand.
Email: a.dorrington@waikato.ac.nz

Abstract: Machine vision using image processing of traditional intensity images is in wide spread use. In many situations environmental conditions or object colours or shades cannot be controlled, leading to difficulties in correctly processing the images and requiring complicated processing algorithms. Many of these complications can be avoided by using range image data, instead of intensity data. This is because range image data represents the physical properties of object location and shape, practically independently of object colour or shading. The advantages of range image processing are presented, along with three example applications that show how robust machine vision results can be obtained with relatively simple range image processing in real-time applications.

Keywords: Range Imaging, Three-dimensional, 3D, Machine Vision

1 Introduction

Machine vision has been successfully used in industry in applications such as factory automation. One of the keys to its success is tight control of environmental conditions, particularly lighting control as well as object and background colours. There is also a demand for machine vision in other applications, such as people tracking and security, where it is difficult to control the environment and the objects. These applications, which are abundant in the literature, are less successful because of the difficulty in separating the wanted object from the background. The problem is confounded when the objects of interest have high contrast features, such as a person wearing a light shirt and dark trousers.

The main limitation in the current machine vision object detection problem is essentially the camera technology. Sophisticated algorithms have been developed to detect objects in the 3D world using a 2D camera, thus ignoring the most obvious of useful information, depth. Attempts have been made with reasonable success to use steroposis (multiple two dimensional cameras with triangulation techniques) to gain the depth information[1]. However, this approaches needs substantial computing power and has limited operating volume (defined by the intersection of the field of views of the cameras).

Time-of-flight (TOF) range imaging is a relatively new imaging technique that can be loosely described as a 3D camera. Such cameras produce a digital photograph or video like output that contains both depth and (monochrome) intensity information simultaneously for every pixel in the image. These cameras have great potential for machine vision applications because they natively capture the depth information that must be inferred with traditional imaging systems. Because the depth information captured with these cameras is (for the most part) independent of intensity and object colour, a machine vision system operating primarily on depth information can separate object from the background with relatively simple algorithms.

We introduce the basic operating principle of range imaging systems, highlight their advantages and disadvantages, and demonstrate some simple applications illustrating how they can be used for machine vision.

2 Background

2.1 Range camera operating principle

There are several different techniques that can be used to acquire range images, but by far the most common is the continuous wave amplitude modulated (CWAM) approach[2, 3, 4, 5]. As this is the only technique used in the currently available commercial products, it is the only one we describe here.

Range imaging is an active image approach. The scene is flood illuminated with an intensity modulated light source. Near infrared LEDs are the illumination source of choice in commercial products with modulation frequencies ranging from 10 MHz to 50 MHz. Laser diodes are used in research devices at modulation frequencies up to 100 MHz, and may make an appearance in future commercial devices because a higher modulation frequency provides better range measurement precision.

The image sensor gain is also modulated at the same frequency as the illumination source. Digital or square wave modulation is often used for both the illumination and sensor gain, so the sensor gain modulation can be thought of a high-speed shuttering during the integration period. This sensor gain modulation makes the pixel sensitive to the phase of the modulation envelope of the illumination scattered from the scene and collected by the camera lens. A change in phase (due to propagation delay) manifests as a change in pixel brightness. The measured phase change is converted to propagation time hence range (with knowledge of the speed of light). Light collected from objects close to the camera generate a bright pixel due to a small propagation delay resulting in a small phase change. More distance objects generate a dull pixel because of a large propagation delay that results in a large phase change.

Of course, pixel brightness is also influenced by object colour and background lighting conditions, so one capture is insufficient to accurately determine distance. Typically four captures are used in which the relative phase difference between the illumination and shuttering is stepped in $90°$ increments. This allows calculation of the background intensity, b, active intensity, a, and phase shift of the modulation envelope, φ, with

$$a = \frac{\sqrt{(A_0 - A_2)^2 + (A_1 - A_3)^2}}{2} \quad (1)$$

$$b = \frac{A_0 + A_1 + A_2 + A_3}{4} \quad (2)$$

$$\varphi = \tan^{-1}\left(\frac{A_0 - A_2}{A_1 - A_3}\right) \quad (3)$$

where A_0, A_1, A_2, and A_3 are the pixel brightness values from the four successive images. The object range can then be calculated from the phase, modulation frequency, f, and the speed of light, c, with

$$d = \frac{\varphi c}{4\pi f} \quad (4)$$

2.2 Advantages and disadvantages

Using range imaging cameras in machine vision applications has the significant advantage of providing distance information, thus allowing computer systems to perceive the world in 3D. It also provides the advantage of generating separate background and active intensity images. This may be of advantage in some applications because the active intensity image provides grayscale information that is not influenced by ambient lighting conditions and shadows.

These advantages come with a price. There are currently several significant disadvantages associated with range imaging cameras[6], of which the most most obvious is the need for active illumination. This increases power consumption and physical size, complicates thermal dissipation, but perhaps most importantly, limits the useful operating distance of the cameras. Objects too distant can be poorly illuminated leading to low quality range measurements. This is especially the case in a scene that contains both near and far objects because the illumination levels and camera integration time must be controlled to avoid saturation of the near, bright objects, resulting in very dull distant objects.

Another significant disadvantage is the low spatial resolution of commercially available range imaging cameras. At the time of writing, the highest resolution camera available is the CamCube 2.0 (PMD Technologies GmbH, Siegen, Germany) with 204×204 pixel resolution. However, as with traditional imaging technology, resolution is increasing steadily with each new model offering higher resolution as the technology matures.

Range imaging cameras also suffer from several range measurement artefacts. At the edge of objects light from two or more objects at different ranges may fall on the same pixel causing erroneous measurements. These errors are referred to as mixed pixels or flying pixels. Also, because of the cyclic nature of the phase measurement used to calculate distance, ambiguities can arise if objects occur at distances longer than half the wavelength of the modulation signal. Finally, significant motion can cause corrupt range data because the scene may change during the four acquisitions required to produce one range imagine. Research is being conducted to solve these shortcomings, and it is likely that future range imaging cameras will include features to minimise or resolve all of these effects.

3 Range Image Processing

Many traditional image processing techniques can be applied to range images[7, 8]. Indeed, operating on the range data not only has the potential to simplify operations such as object detection, but also opens new opportunities for more advanced and qualitative measurements. With range data it is possible to determine physical object parameters such as size and surface curvature. Consequently, range data image processing is likely to be much more robust and reliable than traditional image processing.

To illustrate the simplicity of range data image processing, we present a simple example. Figure 1 shows images acquired with the SwissRanger™SR4000 camera (MESA Imaging, Zuerich, Switzerland). The active intensity image, figure 1a, is similar to a traditional greyscale image but using the built in light source. Note that the object in the foreground has colourful high contrast surfaces and is in front of a visually complex background. The range image, figure 1b, instantly provides additional information about the scene. One of the first noticeable aspects to this image is that the object and the background are of low contrast compared to their appearance in the intensity image, because there is little difference in distance from the camera across their surface.

Segmenting the objet in this scene based on the intensity information presents a challenge for traditional image processing, but is a relatively simple task when processing the range data. For example, consider an edge detection on the intensity image, as shown in figure 1c. Because of the complicated nature of the scene, this detection provides little useful information. However, the same processing on the range data, figure 1d, clearly finds the edges of the gnome in the foreground and the bookshelf in the background.

The other aspect that is immediately apparent from the range image is that the gnome is closer to the camera than the bookshelf as darker pixel shades represent range values closer to the camera. This information in particular cannot be determined in the intensity image. In fact, it is possible to determine the shape, size, and relative location of objects in the scene because the SR4000 also provides a 3D output consisting of x, y, and z values calibrated in metres for each pixel.

Segmentation can be performed very simply in one of two ways. Firstly, if this image is the only data available, simple thresholding can be applied to the range image to segment the objects based on their distance to the camera, as shown in figure 1e. Alternatively, simple background subtraction can be used if a range image of the the scene before the objects were added is available.

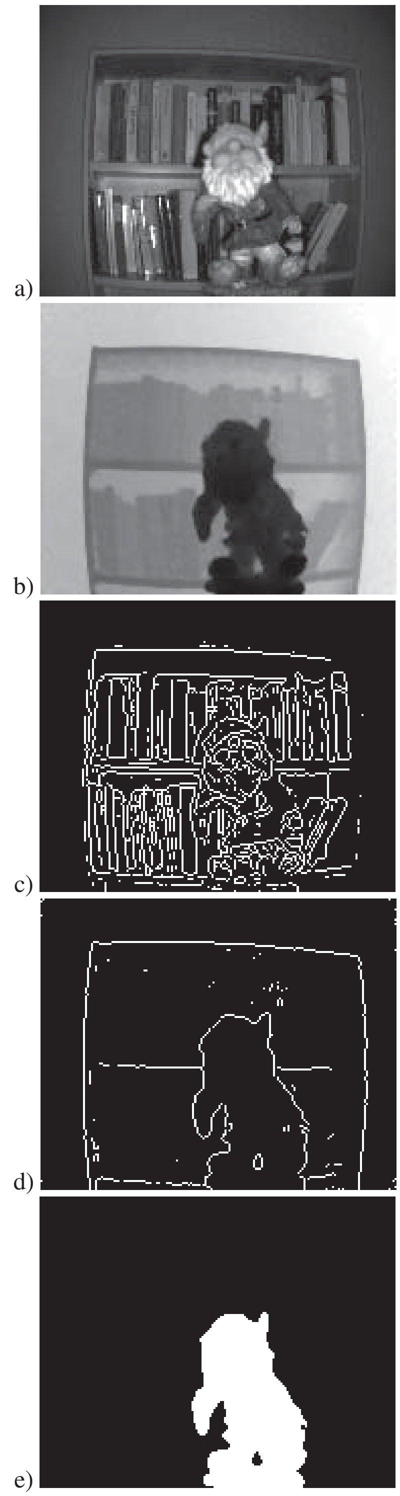

Figure 1: Intensity image (a), range image (b), edge detection on intensity image (c), edge detection on range image (d), and threshold segmented object mask (e) of a gnome in front of a bookcase.

4 Example applications

To illustrate the potential uses of range image processing, we present three examples. In all cases, the image processing code was written in MATLAB™, and is sufficiently simple to operate in real-time on a 2 GHz computer under Microsoft Windows XP™.

4.1 Height estimation

The first example is an algorithm that identifies objects and estimates their height off the ground. This is primarily aimed at measuring a person's height. Although there are

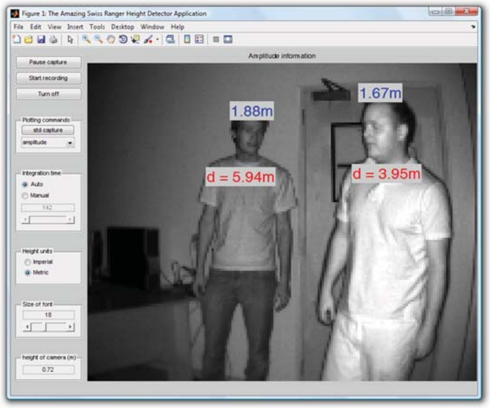

Figure 2: Screen shot of the height detection software showing a scene containing two people, each with their height and average distance from the camera displayed.

Figure 3: A user interacting with a push button menu by moving their hand in free space.

some constraints on object size and dimensions, the software does not discriminate between people and other similar sized objects. The height of a person (or other object) 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 2.

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 was 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 depth image. These edge pixels were 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 was separated into groups using the watershed transform. The pixels in each group were then analysed to determine the maximum height and average distance from the camera for each person.

4.2 Gesture control user interface

Another example that highlights the advantages of range imaging is a gesture control application. In this application, the user can activate and control a menu system, press virtual buttons, and move virtual slider controls simply by moving an empty hand in free space. A screen shot from the application is shown in figure 3. The action of pressing a button, for example, is very difficult to detect with traditional cameras because movement towards the camera produces very little change in the image. However, with depth data it is very easy to detect if an object moves towards or away from the camera.

The range image processing algorithm first uses background subtraction to isolate a potential user. Then, range thresholding produces a mask indicating all pixels that are within 0.2 m from the closets pixel to the camera, thereby segmenting the closest object to the camera. Morphological image processing techniques are used to remove any outlying pixels in the mask. The centre of the mask is found, and the 3D location of the centre is tracked. The location and motion of the centre is used to select and control objects in the user interface.

A push button and a slider bar are two examples of controls available. The push button works by tracking the center of the mask, a button push is recorded when the centre is over a button graphic and then moves forward by a configurable distance over at least 5 frames. The slider bar works by tracking the position of the centre relative to the position of the slider. When the centre is close, to the slider's position tracks with the centre. The user in figure 3 is selecting a push button in an interactive menu.

4.3 Robotic arm location and control

The final example demonstrates range imaging being used to identify and calibrate a low cost robotic arm. Range imaging is used because this type of robotic arm is inherently inaccurate in its positioning and has no feedback to provide control compensation. Two ping-pong balls were added to the arm to provide 3D reference markers, and were detected using range image processing techniques. Intensity and range images of the robotic arm are shown in figures 4 a and b. First background subtraction is used to segment the robotic arm. The locations of the ping-pong balls were then identified using a watershed segmentation technique (applied to the range data), as seen in figure 4c, and their 3D centres calculated from the 3D data, as shown in figure 4d. A model of the arm was then aligned with the centres of the ping-pong balls to relate the arms actual position and orientation to the programmed movement.

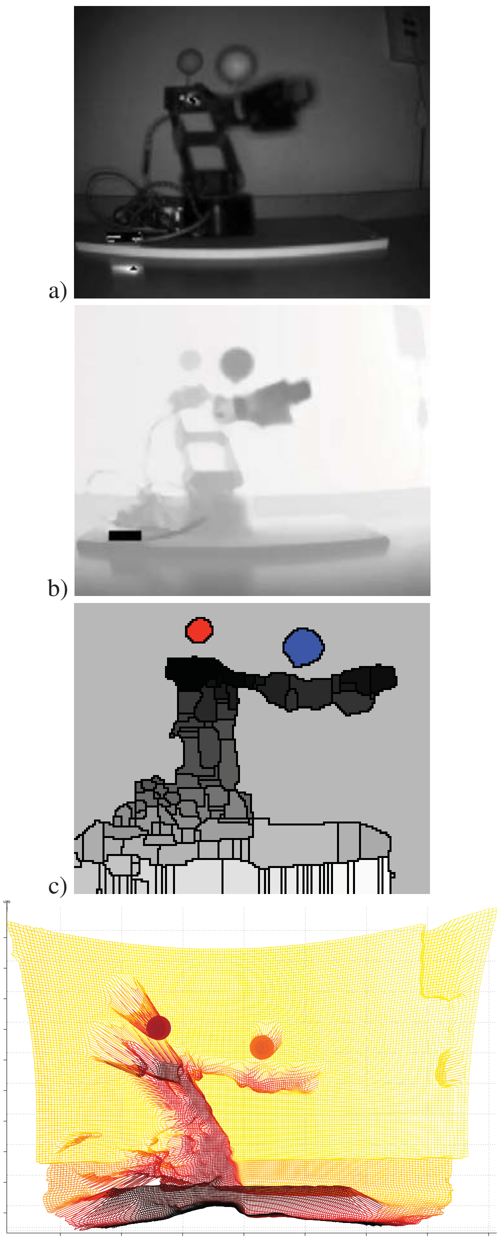

Figure 4: Images of the robotic arm showing the intensity image (a), range image (b), watershed with identified ping-pong balls (c) and 3D data wireframe with fitted shaded spheres (d).

5 Conclusions

We have shown how applying image processing techniques to range data, rather than intensity data, can have significant advantages. This is because range data represents the location and shape of objects in the scene, practically independently of their colour or shade. Comparisons of simple edge detection on both intensity and range data highlight marked differences in the outputs.

Three examples of applications using range image processing are also presented. These examples highlight how range image processing can be easily used in real-time applications such as user interfaces, object size detection possibly including security applications, and control system feedback.

References

[1] T. Acharya and A. K. Ray, *Image Processing: Principles and Applications*. John Wiley and Sons, 2005.

[2] D. Carnegie, M. Cree, and A. Dorrington, “A high-resolution full-field imaging system,” *Review of Scientific Instruments*, vol. 76, p. 083702, 2005.

[3] A. Dorrington, A. Payne, R. Conroy, M. Cree, and D. Carnegie, “Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera,” *Measurement Science and Technology*, vol. 18, pp. 2809–2816, 2007.

[4] J. P. Anthes, P. Garcia, J. T. Pierce, and P. V. Dressendorfer, “Nonscanned ladar imaging and applications,” vol. 1936, pp. 11–22, SPIE, 1993.

[5] R. Lange and P. Seitz, “Solid-state time-of-flight range camera,” *Quantum Electronics, IEEE Journal of*, vol. 37, pp. 390–397, Mar 2001.

[6] M. Hebert and E. Krotkov, “3d measurements from imaging laser radars: How good are they?,” *IVC*, vol. 10, pp. 170–178, 1992.

[7] A. A. Dorrington, M. J. Cree, D. A. Carnegie, A. D. Payne, R. M. Conroy, J. P. Godbaz, and A. P. P. Jongenelen, “Video-rate or high-precision: a flexible range imaging camera,” vol. 6813, p. 681307, SPIE, 2008.

[8] R. Tanner, M. Studer, A. Zanoli, and A. Hartmann, “People detection and tracking with tof sensor,” *Advanced Video and Signal Based Surveillance, IEEE Conference on*, vol. 0, pp. 356–361, 2008.