Large-Object Range Data Acquisition, Fusion and Segmentation

Reyhaneh Hesami1 MIEEE, Alireza Bab-Hadiashar1 SMIEEE and Niloofar Gheissari2
1Faculty of Engineering and Industrial Sciences, Swinburne University of Technology Hawthorne, Victoria, 3122, Australia rhesami@groupwise.swin.edu.au
2National ICT Australia, Locked Bag 8001, Canberra, ACT, 2601, Australia. Niloofar.Gheissari@nicta.com.au

Abstract

This paper presents the design of a low cost large-scale 3D scanner. A pan-tilt unit, a camera and a laser distance meter are combined to capture the required range and intensity data. Innovative methods for calibrating the device and fusion of the intensity and range data are introduced. Once a reliable data achieved, a parametric segmentation technique has been adopted to segment the range images into planar surfaces. Experimental results show that even for data corrupted by a large number of outliers (due to people movement in front of the imaged buildings), range and intensity data are highly consistence and segmentation algorithm has been able to properly segment coplanar areas.

1. Introduction

The availability of a set of accurately fused range and intensity images facilitates the process of 3D environmental modelling. A diverse range of applications becomes viable or significantly enhanced by such a representation of the environment. Examples of such applications are: volumetric and surface recording of historical (perhaps fragile) sites, testing for subsidence of unstable ground areas, virtual reality in environment building and modelling, surveillance camera placement design, architectural reproduction, computer game scenario construction, superimposing exterior sets over studio sets for film production, mining, monitoring coastal zones, etc.

Due to the wide variety of application fields, achieving a complete 3D digital representation of the surrounding environment has, over the years, received substantial attention (1-8) to cite a few. In the past decade, a number of high quality 3D scanners for long distance have been commercialised to fulfil many of the above applications. The critical problem of those 3D scanners is the fact that they are often very expensive (e.g. LMS-Z360i manufactured by Riegl costs more than 150K USD). The main goal of this research is to design and build a relatively inexpensive (less than 10K USD) 3D data acquisition system. The design criteria for the above device has been to:

- Facilitate scientific research at an affordable cost
- Provide adequate accuracy and resolution for large-scale civil applications
- Support easy operation, portability and flexibility

Classical 3D imaging systems are based on active or passive rangefinding techniques [1], [9], [10], [11] such as the time-of-flight method (active) [10], the stereovision method (passive) [12, 13] and the depth- from-defocus method (passive) [14]. However the design of current 3D image sensors is based on the “light-section” method (active) [15, 16] which provides an ultra high speed acquisition process.

Although the stereovision method provides affordable and simple system design, it is computationally expensive and fairly sensitive to the environmental conditions. Since in this method the range resolution is proportional to the distance, for longer ranges, the range data is relatively coarse. As such, its use is generally limited to indoor (short distance) measurement where lighting can be controlled.

In the depth-from-defocus method, the distance is calculated by camera focal tuning. Due to the fact that the range precision and resolution are related to the target condition, this method is basically applied to the objects with explicit edges and surface patterns.

In the time-of-flight method, a projected light (or sound) is transmitted to the target and a receiver views its reflection with a delay, which corresponds to the distance. The range is estimated by measuring this delay. Thus, in this method, the time resolution of the measurement device electronics determines the range resolution, which is independent of the properties and distance of the target. Typically, laser time-of-flight rangefinding is a preferred method.
method for large-object range measurements for two main reasons: firstly, they provide a direct range computation and this reduces the computational cost and acquisition time. Secondly, due to the fact that the accuracy of laser time-of-flight measurement is not very sensitive to the ambient illumination, the imaging (measurement) process is rather straightforward. This is more important in outdoor scene where the ambient light is not easily controllable (or predictable). In general, range imaging using this method has occupied mechanical scanning such as rotating or oscillating mirrors, pan-tilt units, opto-mechanical heads or micro-mechanical mirrors. The inherent drawback of this method of imaging is that the image does not represent the real scene of any specific instant because it is not obtained as an instantaneous shot. Also the mechanical parts consume additional energy.

Here, we have developed a rangescanner using a laser time-of-flight distance meter and captured several range images of different historical buildings and outdoor scenes. To be able to reconstruct the textured 3D model of the environment, the range and intensity data were fused. To examine the usability of the range data in the computer vision context, we have applied a robust segmentation algorithm to the captured range data. The algorithm segments the data into instances of planar surfaces. Visual inspections of our results show that the main features of the outdoor objects (buildings) have been captured by the generated range images.

The rest of the paper is organised as follows. Section 2 presents the detailed design of our rangescanner device including system set-up, calibration, data fusion, and data validation techniques. The segmentation method is explained in section 3 and the experimental results are shown in section 4. A brief conclusion is presented in section 5.

2. Rangescanner Device

In this section we describe the step-by-step design of a range scanner, which uses a time-of-flight laser distance meter. We have developed a mechanically stable and low-cost system that rotates the laser rangefinder to produce a 3D data set of the scene as a point cloud.

2.1. Assembly, Alignment and Calibration

The experimental set-up of the range scanner system is shown in figure 1. The scanner is composed of a laser range sensor (ILM300HR), a digital camera and a computer-controlled pan-tilt unit (PTU-46-70). The laser rangefinder (LRF) can measure distance of up to 300 meters with typical frequency of 1KHz and accuracy of 30cm. The camera captures intensity images of 720 × 576 pixels. The PTU has a maximum resolution of 0.0128° and the resolution of the scanner device is directly related to this angle. The PTU provides maximum horizontal and vertical Field of View ($\Phi$, $\Psi$) of [-159°, 159°] and [-30°, 30°] respectively.

![Figure 1- System Set-up](image1.png)

In our experiments, we have a range image of 720 × 400 points with 30° horizontal and 15° vertical Field of View. Here, ($i,j$) refers to indices of the horizontal and vertical coordinates of each point in the range image and $r_{ij}$ is the corresponding range.

![Figure 2: The overall system architecture (top) and the system coordinates (bottom - $r>>d$, Oc is the centre of camera coordinate system and $O_l$ is the centre of LRF coordinate system)](image2.png)
The raw range data is measured in a spherical coordinate system and is then converted to the Cartesian coordinate system before it is processed. To simplify the sensor fusion, the camera is modelled as a pinhole camera and its optical axis is aligned with the rangefinder optical axis. The camera is mounted on the y-axes of the laser rangefinder. Figure 2 shows the overall system architecture and alignment of the sensors.

2.2. System Calibration

Calibration is an important step of any measuring devices. In this case, the main purpose of calibration is to represent the range and the intensity sensor outputs in a single coordinate system. To achieve this goal, a novel semi-automatic calibration procedure is developed.

The rangescanner, in calibration mode, starts to capture a live video of the scene including the calibration target board while tracking the laser beam. The calibration target board is a 1.5mx1.5m white board and the laser beam is tracked while it reflects of this board. To follow the laser beam in the captured image, an appropriate infrared filter is mounted on the camera lens and the position of laser is reordered on the image plane. By comparing the position of laser for a number of different points, the average rate of changes in the image plane compare to the position on the calibration board can be calculated. The starting point for the range acquisition is then calculated based on the above rate, the desired size of and the desired resolution of the range data.

3. Range and Intensity Data Fusion

To be able to visualise 3D shapes together with their texture, we need to fuse the intensity value for each pixel with its range value. In fact, by assigning (mapping) an intensity to each range value, we achieve a six dimensional data set \((X_{ij}, Y_{ij}, Z_{ij}, R_{ij}, G_{ij}, B_{ij})\) where \(R_{ij}, G_{ij}, B_{ij}\) are the intensity values.

For the fusion task, we assume a pinhole model for the camera and compensate for radial and tangential distortions prior to mapping 3D data into the intensity image plane. Camera calibration is performed based on Bouguet’s camera calibration technique [17] to derive intrinsic and extrinsic camera parameter. In this technique, the main initialisation stage is from [18] and the intrinsic camera model is inspired from [19]. (The notation has been used in this paper for data fusion, is derived from [17] with slightly differences).

In order to correspond the points of the range data to the points of the intensity image \((x_{ij}, y_{ij})\), we consider the coordinate system of the laser rangefinder as the world coordinate system and denote every point on that coordinate system by \((X_{ij}, Y_{ij}, Z_{ij})\) and in the camera coordinate system by \((X_{cij}, Y_{cij}, Z_{cij})\). As shown in figure 2, we have:

\[
\begin{bmatrix}
X_{cij} \\
Y_{cij} \\
Z_{cij}
\end{bmatrix} = \begin{bmatrix}
x_{ij} \\
y_{ij} + d \\
Z_{ij}
\end{bmatrix}
\]

and we define the normalized image projection matrix \((M_{n,ij})\) as:

\[
M_{n,ij} = \begin{bmatrix}
X_{cij}/Z_{cij} \\
Y_{cij}/Z_{cij}
\end{bmatrix} = \begin{bmatrix}
x_{ij} \\
y_{ij}
\end{bmatrix}
\]

Distorted normalized image projection matrix \((M_{d,ij})\) is defined as:

\[
M_{d,ij} = (1+k_1r_{ij}^2+k_2r_{ij}^4+k_3r_{ij}^6)M_{n,ij} + T_{d,ij}
\]

where \(k_1\) to \(k_3\) are distortion coefficients and \(n_{ij}\) and \(T_{d,ij}\) (tangential distortion) are given as:

\[
r_{ij}^2 = x_{ij}^2 + y_{ij}^2
\]

\[
T_{d,ij} = \begin{bmatrix}
k_2y_{ij}^2 + k_3x_{ij}^2 + k_4(y_{ij}^2 + 2x_{ij}^2)
2k_3x_{ij}y_{ij} + k_4(2y_{ij}^2 + 2x_{ij}^2)
2k_2y_{ij}^2 + k_4(x_{ij}^2 + 2y_{ij}^2) + 2k_3x_{ij}y_{ij}
\end{bmatrix}
\]

We can now use the following camera model to calculate the relationship between our range and intensity coordinate systems:

\[
\begin{bmatrix}
x_{p,ij} \\
y_{p,ij}
\end{bmatrix} = \begin{bmatrix}
f_x & c_x \\
f_y & c_y
\end{bmatrix} \begin{bmatrix}
M_{dx,ij} \\
M_{dy,ij}
\end{bmatrix} \begin{bmatrix}
x_{c,ij} \\
y_{c,ij}
\end{bmatrix}
\]

In the above model, \(M_{dx,ij}\) and \(M_{dy,ij}\) are calculated using a six-order distortion model and \(C_x\) and \(C_y\) are the coordinates of the principal point.

3.1. Validation Technique

The accuracy of the camera parameters plays an important role in the data fusion algorithm and therefore we need to validate the camera calibration results. This is an important step as this determines the accuracy of the data fusion. For the verification purposes, we used an appropriate infrared filter and measured the coordinates of the laser beam in the image plane, \((v_p, y_p)\), for various points at different distances from the camera. Comparing the observed laser beam coordinates and its estimated ones (using the methods described in section 2.2 and 3), we have determined that the observed results are within the uncertainty bounds [17] computed from camera calibration parameters.

4. Model-based Segmentation of Large Scale objects

Model-based or parametric range segmentation refers to partitioning a range image to several regions where each presents a homogenous 3D surface (in the real-word) and the union of all regions generates the entire image
The existing range segmentation algorithms can be classified into three main categories: region based, edge based and hybrid techniques. Region based range segmentation algorithms are based on the fact that there is a spatial coherency in the real surfaces, which are imaged by the rangescanner. They use this spatial coherency to assign each pixel to a specific region. On the other hand, edge based range segmentation algorithms extract the edges from the range image and classify them into several groups. Finally, hybrid range segmentation algorithms combine the above two techniques to overcome the drawbacks of each and attempt to take into account all the available information.

4.1. Large-Scale Range Segmentation

Range segmentation of large-scale objects poses a number of challenging issues. The first one is the fact that capturing high-resolution data is very time consuming and therefore the range images are often sparse. Moreover, the disparity of depth in outdoor scenes are often substantial. This means that the difference in range can easily be much larger than the object features that are of interest in a typical object of an outdoor scene. Such disparity in depth has serious implications in terms of the scale of the noise as in many practical cases the level of the noise is proportional to the magnitude of data and most heuristic methods to estimate the scale of noise would be likely to fail in such circumstances.

Another major difficulty in capturing 3D data of large-scale objects is the fact that the environmental aspects are often very difficult to control. In our experiments, and particularly in capturing data of buildings of interests, we noticed that it is very difficult to control the flow of tourists, emergency vehicles, etc. and the images are almost always marred by such influences. It is therefore very important to have techniques that are robust to the influence of outliers and can resolve the occlusion issue (for example, a tourist walking in front of the rangefinder and breaking the continuity of the captured data).

To overcome these issues, we have adopted a robust parametric range segmentation technique that has previous been devised for range images of small (laboratory) objects containing both planar and curved surfaces [21]. The algorithm assumes that every surface in the scene can be described as a parametric model. The advantages of using a robust parametric technique in this case are two fold. First, the occlusion issue (including the artefacts caused by changing environment) can be satisfactorily resolved both by the continuity constraint imposed by the parametric model and the use of robust technique that diminishes the effect of bad data. The second advantage of using a parametric technique is that the algorithm generates high-level information in terms of surface descriptions, which is very useful for 3D reconstruction applications.

4.2. The Range Segmentation Algorithm

Here we briefly describe the range segmentation algorithm (adopted from [21]) used in our experiments. The algorithm employs a threshold K, which is the size of the smallest region that can be regarded as a separate region. This means those structures which contain less than K percent of the whole data are ignored.

In this algorithm, first the pixels whose associated depths are not valid (due to the limitation of the rangefinder used for measuring the depth) are eliminated. These points are usually marked by the rangefinder with an out-of-range number. Then, the following tasks are iteratively performed until the number of remaining data becomes less than the size of the smallest possible region in the considered application:

A localised data group inside the data space in which all the pixels appear on a flat plane is found. Even if there is no planar surface in the image, a very small local area (here 15×15) can be always approximated as a planar surface. To implement this stage and find such a data group, a number of random points, which all belong to the same square of size 15×15 (this square is only for the sake of local sampling) are chosen. Using these points, an over-determined linear equation system is created. If the number of inliers is more than half of the size of the square, then, this square is marked as an acceptable data group. The size of the square (15×15) is not important, however it needs to be large enough to contain adequate sample points. In our experiments, 30 samples were used. Then a planar model is fitted to all the accepted data groups and the residual for each point is calculated. The above two steps are repeated many times (around 8000 times) and the data group that has the least Kth order residual is accepted (the choice of K depends on the application [22] and is set to 2% for our experiments). Although the algorithm is able to detect the true underlying surface model, for simplicity, we fit a planar model to the whole data (not segmented parts); compute the residuals and estimate the scale of noise using the Modified Selective Statistical Estimator (MSSE) as explained in [22]. This technique assumes the noise is Gaussian and reduces the problem of estimating the scale of noise to a Gaussian distribution hypothesis test. Those points whose squared residual is greater than a threshold (T – specified based on the level of significance in the normal distribution) multiple of the scale of noise are rejected. Then, a new segment containing all the outliers to this fit regardless of their geometrical location is generated. As a result, the algorithm has the advantage of detecting and resolving occlusion while segmenting the data. Different parts of embedded objects can also be
separated in this stage. In the next step, by apply a hole-filling (here, we use a median filter of 10 by 10 pixels) algorithm to all inliers; holes resulting from invalid and noisy points (points where the rangefinder has not been able to correctly measure the depth) are removed. This step is mainly cosmetic and has no effect on the segmented surface’s parameters because the fitting has already been performed. Finally, the segmented parts are eliminated from the data and the above steps are repeated.

5. Experimental Results

We have conducted a number of experiments to evaluate the performance of the proposed rangescanner system. Here, we only present two data sets captured of cultural heritage objects. The first set (Figure 3-6) was obtained from the front side of the Melbourne Exhibition Centre (a world heritage listed site) and the second set (Figure 7-10) represents 3D information of the front side of the Shrine of Remembrance (a significant building in Melbourne, Australia).

Figure 3 and 7 show the intensity images while the colour-coded range images are shown in figure 4 and 8. In the latter figures, parts in navy indicate the areas for which the rangefinder could not measure the depth. Such areas include: very close objects, sky, glass windows and shiny items.

The visual inspection of figure 5 and figure 9 shows that the range and the intensity data are highly consistent. In these figures, the areas with no range or intensity (or neither) data are shown in white.

Figure 6 and 10 present the range segmentation results. It is important to note that, although these data contain a large number of outliers due to the random movement of visitors in those places, the algorithm has correctly detected the coplanar areas and has grouped them appropriately.

![Figure 3 - Melbourne Exhibition Centre: Intensity image (720x576) pixels.](image)

![Figure 4 - Melbourne Exhibition Centre: Colour-coded range image (720x400) pixels.](image)
Figure 5 - Melbourne Exhibition Centre: Range and intensity data fusion (720×400) pixels. White regions in the fusion result show lack of valid range or intensity data.

Figure 6 - Melbourne Exhibition Centre: Range segmentation image (720×400) pixels. Each segment is shown in different colour shades in grey scale.

Figure 7 - The Shrine of Remembrance: Intensity image (720×576) pixels.
Figure 8 - The Shrine of Remembrance: Colour-coded range image (720x400) pixels.

Figure 9 - The Shrine of Remembrance: Range and intensity data fusion (720x400) pixels. White regions in the fusion result show lack of valid range or intensity data.

Figure 10 - The Shrine of Remembrance: Range segmentation image (720x400) pixels. Each segment is shown in different colour shades in grey scale.
6. Conclusions

In this paper, we have presented our design of a 3D large-scale rangescanner. The main purpose of this research is to develop a reliable and accurate 3D rangescanner system for large-scale objects with minimum complexity and cost (e.g. components, design and computational and calibration complexities).

A parametric segmentation technique has also been adopted to segment the range images using planar surfaces. The results of our experiments show that although developing accurate and detailed 3D models of large-scale (outdoor) objects poses a substantial challenge, effective 3D geometrical models of large buildings can be reconstructed using the above approach.

7. Reference List