Rapid 3D Measurement and Influences on Precision Using Digital Video Cameras

Willie van der Merwe, Kristiaan Schreve

Department of Mechanical and Mechatronic Engineering
University of Stellenbosch, South Africa
kschreve@sun.ac.za

Abstract

Quality assurance and reverse engineering have become an almost inseparable part of the mass production industry. Non-contact measurement methods are playing an ever more important role. This study implements a rapid measurement system using two digital video cameras. Three different methods, using either laser tracking or structured light patterns, were developed and employed to solve the coordinate extraction and correspondence matching problems. The system achieves calibration in less than a minute and accumulates point correspondences at 12 frames per second. Accuracies of better than 0.4 mm are achieved using a single pair of images with 640 x 480 pixel resolution each.

1. Introduction

Optical measurement techniques have traditionally been bound to specific applications requiring expensive and specialised equipment. With the rapidly developing digital technologies in the market, computers and off-the-shelf digital cameras are continually improving in both speed and capability while also becoming less expensive. Using digital cameras and video cameras as the main data receiver component, the literature reports quite a few techniques that lend themselves to accurate vision metrology [2] [6] [11].

This study shows what can be achieved with the effective combination of simple techniques, readily available software and hardware. The ultimate goal is to build a working vision metrology system capable of rapid measurements, however currently the accuracy is not yet sufficient for this purpose. It is shown that measurement accuracies better than 0.4 mm (for a 235 x 190 x 95 mm volume) can be reached. It also shows that data-sets of thousands of measurements can be made within minutes using the automated and semi-automated processes of calibration, coordinate extraction and stereo-matching developed for the system. Three different methods of correspondence matching are explored and results on measurement precision presented.

2. System Design

As covered in sections 3 and 4, the parameters that mathematically describe the camera model will, to a certain degree, influence the precision of the measurement system. There are however also other factors influencing the precision of calibration and measurement that are mostly independent of the camera model. With the practical implications in mind, some important factors influencing precision, speed and cost are considered here.

2.1. Sub-pixel Target Extraction

In general, the greater the precision with which a feature is extracted, the greater the precision of the calibration.

Before the location of a feature can be determined, the other important consideration is the initial recognition of the features in an image. From an image processing point of view, the simplest way in which to aid automatic detection is by using high contrast features [8]. Examples of this are markers made of reflective material [6] or high contrasted black and white patterns. Using simple geometric shapes for the features, such as circles, rectangles and checkerboard patterns, can then further aid in the recognition phase.

For each of these shapes a different image processing method is used to extract precise target locations. For the rectangles or checkerboard patterns, corners can be initially detected using, for instance, Harris corner-detection. Sub-pixel refinement of the corner locations can then be made using interpolation between pixels [4]. Another method of refining the corner coordinates in these two cases is by using edge information to calculate line intersections [5] [10]. For circular features a number of locating methods are discussed and evaluated by [8].

The precision with which the coordinates of each of these shapes can be extracted using their corresponding methods is influenced differently by lens distortion and perspective effects of an optical system. Mallon and Whelan [5] found that circular patterns yield the least precise target location, being influenced by the lens distortion as well as the perspective effects. The best results were found for the line-intersection method which is invariant under perspective transformation, but is still influenced by lens distortion.

2.2. Angles of Convergence

With an increase in angles between rays formed by the same point, the precision of the calibration network will also increase. The practical implication is that the “base-to-depth” ratio should be as large as possible, up to 1:1, i.e. when the angle of convergence is 90 degrees. The base refers to the distance between camera centres and the depth refers to the perpendicular distance from the base-line to the point being measured. This effect of the converging angles is mentioned throughout the literature [7] [3], but no results were found to indicate the increase of calibration or measurement precision with an increased convergence angle.

2.3. Projective Coupling

Projective coupling refers to the correlation between the internal and external camera parameters. An example given by [9] is

73
the typical coupling between the principal point location, decentering distortion and the tip or tilt of the camera. Small changes in any of these parameters will still yield the same overall calibration result. For the case where there is a strong projective coupling, [7] as well as [10] makes a similar observation: there is a negligible difference in the final 3D precision if the principal point offset parameters are given different values (within a reasonable range). Remondino and Fraser [7] note this is also true for the decentering distortion terms.

3. Camera Model

The mathematical description has been well established throughout the literature. For this reason a brief description of the final camera model will be given here and the reader is referred to [4] or [3] for further information. In this paper the pinhole camera model is used [3] with a distortion model.

3.1. Linear Camera Parameters

In order to calculate the 3D coordinate of a corresponding point in a stereo-pair of images, the camera matrix, \( P \), is needed. This 3x4 matrix describes the projection (or mapping) of a 3D coordinate onto the image plane of the camera for which \( P \) was determined using only the linear description of the camera. Equation 1 is the compact notation for this mapping from the world coordinates \( X \) to the image coordinates \( x \).

\[
x = K[R][I-C]X
\]  

Starting from the left in Equation 1, the 3x3 matrix \( K \) describes the camera’s internal parameters in terms of pixels. These parameters describe the focal length, the image centre as well as the width to height ratio of the pixels. The following 3x3 matrix \( R \), describes the rotation of the camera and the 3x4 matrix \( [I-C] \) is constructed using the position of the camera centre. The final camera matrix containing all these elements is given in Equation 2 and the final mapping using \( P \) is given in Equation 3.

\[
P = K[R][I-C]
\]  

\[
x = PX
\]

In practice, \( P \) will be determined directly from the calibration process and the separated internal and external parameter components will not be needed. In this application, \( P \) is used directly for triangulation (along with the added non-linear parameters), which takes care of the projective coupling problem mentioned in section 2.3.

3.2. Distortion Model

Different mathematical models can be used for radial distortion, but they are most commonly described in the form of some polynomial expansion as a function of the distance from the radial centre, \( r \). The radial distortion model used here was taken from [4] and its vector form is shown in Equation 4.

\[
x_u = c + f(r)(x_d - c)
\]  

The undistorted image coordinate, \( x_u \), is computed by adding the coordinates of the centre of radial distortion, \( c \), to the coordinates of the corrected \( x \)- and \( y \)-distances. These corrected distances are calculated by multiplying the \( x \) - and \( y \)-distances from \( c \) to the distorted point, \( x_d \), by the correction function, \( f(r) \) in Equation 5.

\[
f(r) = 1 + k_1r + k_2r^2 + k_3r^3
\]  

4. Calibration

A very simple two-step method has been developed and implemented here. In the first step, the camera parameters are approximated using a linear method which ignores non-linear effects such as lens distortion. For this method, a 3D calibration object with known feature coordinates was designed and manufactured (Figure 1). The coordinates of the grid corners were determined on a Coordinate Measurement Machine (CMM). The repeatability of the CMM measurements of the grid was determined to be well below 0.1mm (within 95% confidence). The second step introduces non-linear effects of lens distortion with the model described in section 3.2. These parameters are determined through an optimisation function which minimises the back-projection error of the known 3D coordinates using the initial values from the first step.

4.1. Initialisation of Camera Parameters

If non-linear effects can be ignored, the camera matrix, \( P \), can be determined using a simple linear method if the image coordinates and their corresponding world coordinates are known. Used here is the direct linear transform (DLT) method as described by [3], but without the minimisation of geometric error.

For practical implementation of the solution, the linear system is first properly pre-conditioned. This is done by scaling and shifting both the image and world coordinates [3]. After normalisation the DLT algorithm calculates a normalised camera matrix. This matrix is de-normalised to retrieve the final camera matrix.

4.2. Refinement of the Camera Parameters

The values of the camera matrix from the DLT algorithm are now used as initial values for a robust and quickly converging minimisation function. This function must introduce the non-linear lens distortion into the thus far linear camera model.

With the camera matrix and a set of known world-coordinates available, there is an almost intuitive error to minimise: the difference between the calibration-feature coordinates initially extracted from the image and the back-projection of the world-coordinates onto the image plane.

There are different ways in which this error-set can be used to calculate an output for minimisation. Here it has been decided that the mean and standard deviation (SD) of the error-set will be added together and used as the value to be minimised. This has been established through trial and error as the best combination of values. Using the sum of these values gives a low mean value with a higher certainty in the error distribution. Using only the mean usually causes the standard deviation to be slightly higher and vice versa if only the standard deviation is used.

5. Image Processing

A number of image processing techniques are used and combined for the different stages of the process in order to speed it up. The first step is to automate the calibration phase and secondly the measurement phase in terms of automatic target extraction and correspondence matching between image pairs.
6.6 Automated Detection of the Calibration Grid

Figure 1 shows the calibration object used in the processing. Two well contrasted views are taken through a number of image processing techniques in order to detect each block on the object and find the matching blocks in the two views.

Figure 2 shows the initial image of the grid from one view. After a number of processing steps each block is detected, its approximate centre found and then fitted with an approximate square.

The final operation for each square is to precisely determine the corner positions as shown in Figure 3. To achieve this, a segment of each edge of a block is extracted using the positions of the approximated squares. Each edge-segment is then processed to detect the position of each edge pixel with high precision using the intensity peaks of the edge segment’s derivative image. The derivative image is in turn calculated by convolving the edge segment with a 1D derivative kernel. The edge positions are then used to apply a least-squares line-fitting to that edge. The corners are calculated as the intersection of the lines.

6.6.6 Rapid Correspondence Matching

Once the calibration stage is complete, the camera matrix and distortion coefficients can be used to determine the 3D coordinates of any two corresponding image coordinates. In order to solve the correspondence matching problem during the measurement stage, both a Digital Light Processing (DLP) projector and a moving laser dot is used to scan objects for measurement.

Three methods are proposed for the correspondence matching. Only the first, using a moving laser dot, can be used for practical measurements on non-planar objects. The other two are only for testing and comparing the achievable measurement precision. These last two methods use the DLP projector to project known patterns onto an object, currently only a planar surface.

5.2.1 Tracking a Moving Laser Dot

For tracking the laser, it is assumed that it is the highest intensity moving object in the image. Two consecutive images from the video stream are subtracted from one another. The position of values that are above a given threshold in the difference image is taken as the approximate position of the laser. A small region of interest (ROI) in one of the original images is taken around the approximate position determined from the difference image. Using a binary threshold on the ROI, the centroid of the black-and-white image is calculated.

5.2.2 Corner Detection Using Square Projections

This method is semi-automated and requires some user-input. Three squares in a single column are projected onto a flat surface using the highest possible intensity of the projector. This is done to get the best contrast between the white squares and the darker surroundings. The algorithm then finds the corners of each square and calculates the correspondences accordingly.

5.2.3 Projected Line-Crossings

This method uses horizontally and vertically projected strips of light. It is also only functional for a flat surface. The derivative images of the vertical and horizontal projections are added together. This last image gives the higher intensity areas where the vertical and horizontal lines cross. A small ROI is now extracted around each of the high intensity spots and the greyscale centroid is calculated. In order to find a corresponding coordinate in the stereo pair, the epipolar geometry of the cameras are used [3]. With only two images, this is an unstable way of searching for correspondences, with many erroneous correspondences being found, especially if there is a large number of crossing points positioned close together.

6. Experiments

6.1 System Description

The system was developed with the Python programming language using OpenCV (www.intel.com/technology/computing/opencv) for many of the image processing functions. Two Firefly MV IEEE 1394 cameras distributed by Point Grey Research are used for image capture. Each camera has a 640x480 resolution, with a frame-rate of 40 frames per second. One camera is greyscale, the other is colour, using a BGR Beyer-pattern

Figure 3: Block with sub-pixel corners detected.
and a edge-sensing colour processing method to create a three-channel colour image.

6.2. Definition of Errors

There are three errors that will be used as outputs for evaluation during the calibration and measurement experiments: back-projection and triangulation error for the calibration stage and deviation from a fitted plane for the measurement stage.

The back-projection error is the same error defined in section 4.2. The triangulation-error is calculated in much the same way as the back-projection error. The distances between the triangulated coordinates of the calibration grid corners and their known coordinates are also accumulated in an error set. Note that these errors can only be computed during the calibration stage because of the known world coordinates. It is one of the main reasons for the use of this type of calibration method: the achievable metric precision for the system can be established directly from calibration.

For the measurement phase, a flat surface will be scanned using all three methods: laser dot, projected squares and projected line-crossings. To evaluate the error, the triangulated surface coordinates will be fitted with a plane using a least-squares method. Each perpendicular distance from the plane (planar deviation) to a triangulated coordinate is accumulated in an error set and evaluated statistically using the standard deviation and visually using the histogram. The mean value is not used, because it is usually very close to zero due to the way the plane is fitted to the triangulated coordinates. Both [1] and [12] use the deviation from a flat surface as an estimate of the noise in the measurement system. The principle is that if a flat surface is reconstructed, any planar deviation indicates the basic measurement error that can be expected in the system.

6.3. Experimental Variables

The variables that will be used as inputs for the calibration experiments are the camera model complexity and the base-to-depth ratio. From the literature it is known that the optimum results should be achieved using a base-to-depth ratio of one and the most complex camera model. In this case it is a model containing two radial distortion coefficients and a drifting centre coordinate. Even though the optimum case is predictable, it will be tested in order to verify results already presented in the literature as well as evaluate the effect on precision for this unique system.

Because all code was custom-developed for this study, the camera model can be adjusted to contain different coefficients for distortion, allowing for the complexity to be increased systematically by adding more of the distortion model coefficients. It is assumed that the effect of the two variable parameters are independent of one-another. An experimental design testing the interdependence of the variables, such as a full-factorial experimental design, is therefore not used. For each variable, the other parameters are held fixed at their optimum values as given above.

6.4. Results

6.4.1. Model Complexity

The first test uses the DLT method directly with no distortion parameters. The first radial distortion coefficient, $k_1$, is then introduced, followed by the second, $k_2$, and finally the drifting radial centre, $c$, is also added. Table 1 and Table 2 give the back-projection and triangulation results of calibration respectively.

The very small difference in the triangulation error between the last two columns of Table 2 indicates that the tangential distortion has a much smaller effect on precision than the radial distortion. To clarify: when adding the drifting centre to the distortion model, the improvement in precision is two orders of magnitude smaller than the improvement gained for adding radial distortion.

When using only one distortion coefficient, the precision is still comparably close to the cases of greater precision. Using only the linear model, however, yields significantly less precise results, even with the iterative improvement that gets rid of statistical outliers.

<table>
<thead>
<tr>
<th>Camera model</th>
<th>Pinhole model</th>
<th>$k_1$</th>
<th>$k_1$, $k_2$</th>
<th>$k_1$, $k_2$, $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour Camera</td>
<td>Mean (pixels)</td>
<td>0.353</td>
<td>0.223</td>
<td>0.216</td>
</tr>
<tr>
<td>SD (pixels)</td>
<td>0.191</td>
<td>0.122</td>
<td>0.116</td>
<td>0.114</td>
</tr>
<tr>
<td>Mono Camera</td>
<td>Mean (pixels)</td>
<td>0.431</td>
<td>0.248</td>
<td>0.235</td>
</tr>
<tr>
<td>SD (pixels)</td>
<td>0.237</td>
<td>0.142</td>
<td>0.133</td>
<td>0.129</td>
</tr>
</tbody>
</table>

6.4.2. Base-to-depth Ratio

For the different test runs, the calibration object remains in the same position while the cameras are moved further from or nearer to one another across the baseline (the line along which the base distance is measured). Table 3 shows the results of the back-projection error for the two approximate base-to-depth ratios, while Table 4 shows the triangulation results.

Even though the 0.5 ratio yields better back-projection results for the colour camera (Table 3), this does not mean it will give better triangulation results. As expected, after five consecutive runs to get the average values presented in the tables, it is clear that for a greater base-to-depth ratio the triangulation is more precise.

<table>
<thead>
<tr>
<th>Base/depth ratio</th>
<th>1</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour Camera</td>
<td>Mean (pixels)</td>
<td>0.219</td>
</tr>
<tr>
<td>SD (pixels)</td>
<td>0.116</td>
<td>0.114</td>
</tr>
<tr>
<td>Mono Camera</td>
<td>Mean (pixels)</td>
<td>0.225</td>
</tr>
<tr>
<td>SD (pixels)</td>
<td>0.123</td>
<td>0.138</td>
</tr>
</tbody>
</table>
Table 4: Triangulation error for varying base-to-depth ratios.

<table>
<thead>
<tr>
<th>Base/depth ratio</th>
<th>1</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (mm)</td>
<td>0.157</td>
<td>0.214</td>
</tr>
<tr>
<td>SD (mm)</td>
<td>0.077</td>
<td>0.118</td>
</tr>
<tr>
<td>Precision, 95% conf (mm)</td>
<td>0.388</td>
<td>0.569</td>
</tr>
</tbody>
</table>

6.4.3. Planar Deviation

The results for the laser tracking method, the square corner matching and the projected line crossings are all presented in Table 5. Four times standard deviation (4 SD) of the error is used as the final output value to evaluate these measurements.

Table 5: Comparison of matching method precision.

<table>
<thead>
<tr>
<th>Matching method</th>
<th>Square corners</th>
<th>Laser</th>
<th>Line crossings</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D (mm)</td>
<td>0.103</td>
<td>0.235</td>
<td>0.263</td>
</tr>
<tr>
<td>4 SD (mm)</td>
<td>0.419</td>
<td>0.940</td>
<td>1.052</td>
</tr>
</tbody>
</table>

The best results by far are given by the square corner method. This is understandable, because it extracts the matching coordinates much more precisely than the laser or line-crossing method. The laser-dot’s form is not very stable from frame to frame, making the calculation of its centre quite unpredictable. Lastly, the line-crossing method performs worst. Different methods than the weighted centroid calculation might have to be used to achieve greater precision with the line-crossing method.

Figure 4 compares the histograms of the error-sets for each of the methods using the same x-axis scale for comparison. Note that for every method an area of about 210 x 240 mm is used. The spread of the histograms illustrate how the precision differs from method to method.

6.4.4. A Practical Measurement

The laser tracking method is used here to scan the profile of the bottle seen in Figure 5(a), along with different presentations of the 3D data. Even though not the most precise, this matching method is currently the only one capable of measuring more complex surfaces. This measurement is used for a qualitative evaluation only.

The point-cloud of the scanned profile consists of 15790 coordinates accumulated at about 12 fps. Points can of course only be constructed if the laser is visible in both images, which explains the loss of data around sharp bends. Note that the base-to-depth ratio used here is approximately 0.5 in order to increase the field of view common to both cameras.

7. Conclusions

A rapid optical measurement system has been developed and implemented for this project. It is capable of accumulating feature correspondences at 12 points per second with sub-millimetre precision. The precision achieved by calibration is better than 0.4 mm (in the case of the square corner method or better then 1 mm for the laser tracking method) for a 235 x 190 x 95 mm volume, using only one image pair and an image resolution of 640 x 480 pixels.

Most of the processes usually requiring time intensive user interaction in such a system has been automated using different image processing techniques in combination with the right hardware components. This includes the calibration phase as well as three different semi-automatic methods for solving the problem of rapid and precise correspondence matching.

8. References

Figure 4: Error histograms for matching methods.