Evaluation of Camera Calibration Techniques for Quantifying Deterioration

Michael O’Byrne¹, Franck Schoefs², Bidisha Ghosh³, Vikram Pakrashi¹

¹Department of Civil and Environmental Engineering, University College Cork; Cork, Ireland
²Université Bretagne-Loire, Université de Nantes, Research Institute of Civil Engineering and Mechanics (GeM)/Sea and Littoral Research Institute (IUML), CNRS UMR 6183/FR 3473, Nantes, France; IEXED/CAPACITES, Nantes, France
³Department of Civil, Structural and Environmental Engineering, Trinity College Dublin; Dublin, Ireland
E-mail: michael.obyrne@ucc.ie, bghosh@tcd.ie, franck.schoefs@univ-nantes.fr, v.pakrashi@ucc.ie

ABSTRACT: Imaging systems offer an efficient way of obtaining quantitative information on the health status of structural components. They hold particular value for underwater inspections as they can be easily adapted for underwater use and they enable physical information to be captured from a scene for the purpose of later analysis. In order to make the visual data a part of a quantitative assessment, it is necessary to calibrate the imaging systems so that photographed instances of damage can be expressed and measured in physically meaningful real world units, such as millimetres, which can then be used by engineers in subsequent analyses. The imaging system employed in this study is a stereo rig. It consists of two synchronised cameras that capture images of the scene from slightly different perspectives, thereby encoding depth information. This paper evaluates and compares two main approaches for calibrating such a stereo systems, namely, the classical checkerboard calibration and self-calibration based on Kruppa’s equations. Conventional checkerboard calibration must be carried out on-site by photographing a planar checkerboard pattern that is held at multiple random poses, while self-calibration can be carried out after-the-fact and relies only on the static scene acting as a constraint on the camera parameters. The performance of each approach is assessed through a set of experiments performed on controlled real-world specimens as well as on synthetic data. Results indicate that checkerboard calibration is slightly more accurate than self-calibration; however, the practical advantages of using self-calibration may outweigh this reduction in accuracy. An understanding of the advantages and limitations associated with each camera calibration allows inspectors to rationalise the use of either approach as part of their inspection regime, and it helps them to fully capitalise on the benefits of image-based methods.

KEY WORDS: Imaging; Camera calibration; Self-calibration; Deterioration quantification; Structural Health Monitoring.

1 INTRODUCTION

Efficient monitoring strategies seek to collect and extract maximum useful information from structural performance data at a minimum cost and in a reliable manner, while limiting or supplementing the qualitative, subjective and unreliable aspects. An increasingly popular way of enhancing the quality of the data acquired from inspections relies on making more effective use of imaging devices. Imaging devices are capable of introducing a source of quantitative information that can offset some of the inherent limitations of conventional visual inspections.

Many marine structures are assessed visually either by trained divers or by Remotely Operated Vehicles (ROVs). Visual inspections carried out by divers almost always capture photographs to include in the inspection report, while ROVs are typically equipped with at least one camera/video system. In both cases, the acquired imagery is usually qualitatively reviewed and archived; however, it is rarely exploited to the fullest potential in a quantitative sense despite requiring only minimal effort to do so. The quantitative nature of the data obtained from image analysis is important and naturally lends itself to numerous applications, including for developing new damage models, or strengthening existing ones, which can be used to forecast the rate of propagation of damage as the structure continues to operate.

Methods of extracting quantitative information from images vary in terms of the level of sophistication. Naive approaches entail placing an object of known dimensions in the scene, such as a ruler, and then inferring the dimensions of nearby instances of damage by establishing a scale factor that relates pixel units to real world units, such as millimetres. More robust and advanced solutions can reconstruct the full 3D scene through the acquisition of multiple views. This is typically accomplished using either Structure-From-Motion (SFM) photogrammetry or stereo-based approaches, both of which require camera calibration to determine a relation between 2D image coordinates and the 3D world. Furthermore, capturing the 3D shape, instead of just a 2D projection as a standard camera does, opens the possibility of embracing a host of emerging technologies, such as virtual and augmented reality [1].

Structure-from-motion photogrammetry techniques are used to reconstruct a scene from a sequence of overlapping images acquired by a single moving camera [2]. The process is based on the automatic extraction of points of interest (a sparse set of features), the tracking of this sparse set of features across the image sequence, and the estimation of their 3D positions using multiple views [3]. Photogrammetry techniques have the advantage of being easy to implement on-site as they only require a diver to operate a single unconstraint camera and capture photographs as normal. However, they are limited insofar as the 3D reconstructions are scale ambiguous (i.e. the 3D scene can only be recovered up to an unknown scale factor), there is a requirement that the scene remains static, and the performance of the technique is heavily reliant on the successful tracking of features over time. Reliably tracking
features is difficult when operating in underwater conditions characterised by poor visibility [4]. Given these limitations, this paper focuses on calibration of stereo systems as it is felt that the effort and expense associated with underwater inspections warrants the use of more dependable solutions.

Stereo systems have been widely used for underwater inspections that have been carried out by divers [5] and ROVs [6]. These systems consist of two horizontally displaced digital cameras to simultaneously photograph a scene from slightly different perspectives, resulting in two images that are collectively referred to as a stereo pair. 3D information can be extracted by comparing information about the scene from the vantage point of the left camera with that of the right camera. The use of two cameras represents a good compromise between reliability and operational complexity as 3D shape information can be obtained from a single stereo pair once the stereo system has been calibrated, and unlike SFM, stereo systems can be used even if the scene does not remain static (i.e. it evolves over time) since both images in the stereo pair are captured simultaneously.

Camera calibration [2] is the process of finding the camera’s extrinsic and intrinsic parameters. It must be carried out for each set of inspection conditions and for each stereo camera configuration. For a stereo system, the extrinsic parameters describe the rotation and a translation of the right camera with respect to the left camera, while the intrinsic parameters consist of the focal length, the optical centre, also known as the principal point, and the skew coefficient. Once the intrinsic and extrinsic camera parameters are known, and, taking into account of radial distortion, image pixels can be back-projected to 3D rays in space.

A well calibrated stereo system is essential for ensuring that the reconstructed 3D shape accurately reflects the true real-world shape. This paper evaluates and compares two main approaches for calibrating stereo systems, namely, the classical checkerboard procedure and self-calibration that is based on Kruppa’s equations. This performance of each approach is assessed through a set of experiments performed on controlled real-world specimens as well as synthetic data. A measure of the accuracy of the estimated calibration parameters is found by comparing the sizes of the reconstructed shapes with the known sizes and by calculating the re-projection error. The contribution of this paper lies in the application of this study to the domain of underwater inspections where the usefulness of extracting quantitative information from visual data is greatly felt, as well as in the novel methodology that leverages the advantages of real world and virtual data.

The background of stereo-based 3D imaging is discussed in the following section. Section 3 presents the images that are analysed in this paper. Section 4 evaluates and compares the performance of each calibration procedure. Section 5 concludes the paper.

2 BACKGROUND

2.1 Camera Model

The most widely used camera model is the pinhole camera. In this model, the relationships between world coordinates \( \mathbf{X} \) and image (pixel) coordinates \( \mathbf{x} \) is modelled via the perspective transformation. Let \( \mathbf{x} = (u, v, 1) \) be the homogenous coordinates of a point in an image. The equation for relating these image points to real world points is:

\[
\begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix}
= \mathbf{P}
\begin{bmatrix}
    X_s \\
    Y_s \\
    Z_s \\
    1
\end{bmatrix}
\]

(1)

where \( s \) is a non-zero scale factor, \( \mathbf{X} = (X_s, Y_s, Z_s, 1) \) are world coordinates, and \( \mathbf{P} \) is a 3x4 projection matrix that completely represents the mapping from the scene to the image. \( \mathbf{P} \) encapsulates both the extrinsic and intrinsic parameters of a camera. It is given by:

\[
\mathbf{P} = \mathbf{K} [\mathbf{R} | \mathbf{T}]
\]

(2)

where \( \mathbf{R} \) and \( \mathbf{T} \) are the extrinsic parameters - representing the rotation matrix and the translation of the camera in the 3D scene, respectively. Both the rotation matrix and the translation vector have three degrees of freedom. \( \mathbf{K} \) is an upper-triangular matrix that represents the intrinsic parameters. It has five degrees of freedom. The five intrinsic parameters correspond to the focal length in pixels for the horizontal and vertical axes, \( f_x \) and \( f_y \), respectively, the skew parameter, \( sk \), and the two principle points \( x_0 \) and \( y_0 \) (the optical centre of the camera), as summarised in Equation 3.

\[
\mathbf{K} =
\begin{bmatrix}
    f_x & sk & x_0 \\
    0 & f_y & y_0 \\
    0 & 0 & 1
\end{bmatrix}
\]

(3)

2.2 Stereo Imaging

Recovering depth from stereo images involves a number of sub-problems, including rectification, matching, and reconstruction. Rectification is the process of transforming the left and right images of a stereo pair, such that corresponding points in each image are separated only by a horizontal distance and not by a vertical distance. Rectification can be carried out either with calibration information using the essential matrix, or without it using the fundamental matrix. The essential matrix, \( \mathbf{E} \), is a 3x3 matrix that depends only on the extrinsic parameters \( \mathbf{R} \) and \( \mathbf{T} \). The fundamental matrix, \( \mathbf{F} \), is a generalisation of the essential matrix. It relates corresponding points in the stereo images and may be estimated from seven or more point correspondences. The seven parameters represent the only geometric information about cameras that can be obtained through point correspondences alone. It does not require any knowledge of camera internal parameters.

All points in the rectified images should satisfy the epipolar geometry of a rectified image pair (i.e. that the images are aligned horizontally). This may be expressed as follows: if a point \( x_i \) in the left image corresponds to a point \( x'_i \) in the right image then they should satisfy the constraints in Equation 4 and 5. These constraints are geometrically equivalent; the only difference being that Equation 4 is more general as it is based on the fundamental matrix, \( \mathbf{F} \), while Equation 5 makes use of camera calibration information by including the calibration matrices for the left and right cameras, \( \mathbf{K} \) and \( \mathbf{K}' \) respectively,
as well as the relative position and orientation between the cameras which is captured by the essential matrix E:

\[ x^T F x_i = 0 \]  
\[ x^T K^{-T} E K^{-1} x_i = 0 \]

Rectification makes the task of matching pixels in the left and right image considerably faster as it confines the search to the horizontal direction only.

The second stage of the stereo process entails solving the stereo correspondence problem. Stereo correspondence is principally about finding matching points between the left and right stereo images, typically using some measure of similarity such as the Sum of Absolute Differences (SAD) or a correlation based metric. In this paper, a loopy Belief Propagation (BP) based technique is adopted, as described by Ho [7]. The output of this stage is a dense disparity map which represents the distance in pixels between corresponding points in the rectified images.

The final step uses this disparity map to reconstruct the 3D scene, as the disparities are proportional to the distance between the cameras and the 3D world points. If the camera’s intrinsic and extrinsic parameters are known then a full metric 3D scene can be computed.

It is worth mentioning that each stage of the stereo imaging pipeline is dependent on the preceding stages. For instance, inaccurate rectification adversely impacts on the level of success that can be attained in subsequent stages. An image that is poorly rectified will likely lead to unreliable matching, and consequently, will result in a low quality 3D reconstruction. This underlines the value of having reliable calibration data to start off with.

2.3 Checkerboard Calibration

Checkerboard based calibration involves capturing stereo photographs of a checkerboard/chessboard, that is positioned at several unknown positions and orientations in the scene, as depicted in Figure 1. A checkerboard pattern is used as the black and white grid structure naturally produces a lot of high contrast corner points that can be easily detected and precisely located by the calibration algorithms.

![Figure 1. Checkerboard based calibration.](image)

A seminal example of checkerboard calibration algorithm is given by Zhang [8]. Zhang’s method calibrates cameras by solving a particular homogeneous linear system that captures the homographic relationships between multiple perspective views of a checkerboard that is known to be planar. This calibration approach is popular as it is capable of achieving accurate results whilst being relatively easy to carry out on-site. Generally, it is advisable to capture more than ten photographs of the checkerboard under the same conditions that the inspection photographs will be captured under (i.e. same underwater medium, intrinsic camera settings, stereo-rig configuration etc.), in order to obtain reliable calibration parameters.

2.4 Self-Calibration

Self-calibration, also known as auto-calibration, is an attractive way of determining the intrinsic and extrinsic camera parameters as it does not require the diver/photographer to undertake any preliminary calibration procedures, nor does it require the assistance of a second diver or the use of additional props. Self-calibration refers to the process of obtaining a calibrated camera matrix using static scene as a constraint for the five degree-of-freedom pinhole camera model, as represented by the matrix \( K \) in Equation 3. Theoretically, a minimum of three views is needed for full calibration assuming the intrinsic parameters remain constant between views (i.e. the focal length, or any other parameter, is not adjusted between views). In reality, however, the principle points, \( x_0 \) and \( y_0 \), can usually be safely estimated to be at the image centre. Furthermore, most modern imaging sensors and optics provide further prior constraints such as zero skew and unity aspect ratio, which means that the horizontal and vertical focal length, \( f_x \) and \( f_y \), can be regarded as having the same value. Integrating these priors will reduce the minimum number of views required to two. This paper adopts the approach described by Faugeras et al. [9], whereby two positions of the stereo system were analysed. Kruppa’s equations provide an initial estimate of the focal lengths in the left and right camera matrices, \( K \) and \( K' \) respectively, and the principal points, \( x_0 \) and \( y_0 \), are initially assumed to be located at the image centre, while the skew is taken as 0. The camera matrices are then refined through non-linear optimization.

The only additional information required is the baseline, which is the distance between the two cameras (measured in millimetres). It provides the single necessary piece of information for obtaining correctly scaled 3D reconstruction.

There are some inherent drawbacks associated with self-calibration. Firstly, real cameras are affected by radial and tangential lens distortion. While checkerboard calibration is capable of estimating the distortion parameters and undistorting the images, self-calibration methods are not well equipped to deal with this additional layer of complexity. This is especially problematic when dealing with wide-angle lenses as the level of distortion tends to be particularly severe.

Secondly, self-calibration requires the extraction of a sparse set of corresponding points from the left and right frames of the stereo pair, \( x \) and \( x' \) respectively, in order to set up the static scene constraint problem. However, if not enough corresponding points are found (a minimum of 5 is required), or if erroneous correspondences are present, then the essential matrix cannot be determined, or it will be poorly estimated. This issue is mitigated by using the robust, state-of-the-art SIFT (Scale Invariant Feature Transform) algorithm [10] for
extracting and matching points of interest in the left and right images in conjunction with bundle adjustment, which serves to reject outlier matches.

3 DATA ANALYSIS

3.1 Applications of 3D shape recovery for underwater inspections

Recovering quantitative shape information is a challenging but useful task, and has wide applicability in many areas of Structural Health Monitoring (SHM). This paper focuses on an application specific to marine structures; that is, measuring the marine growth accumulation. Marine growth colonisation has adverse effects on the hydrodynamic performance of marine structures. The two key parameters that are needed for structural reliability computation are the thickness and the roughness [11]. Both of these vary around and among structural components and require knowledge of the 3D shape of the marine growth to be computed.

The calibration procedures under consideration are evaluated on controlled shapes representative of marine growth, as well as on synthetic data that has been designed to reflect a realistic underwater scene whilst maintaining full control over the camera parameters and the 3D structure.

3.2 Real Data

The specimens were designed to represent the size and shape of typical hard fouling organisms such as barnacles and semi-rigid fouling organisms such as sea anemones. They are comprised of standard geometric shapes. The barnacles are represented by cones of height 25 mm, while the sea anemones are represented by hemi-spheres/protracted domes that come in three heights - ranging from 25 mm to 40 mm.

The stereo system consisted of two Canon EOS 600D DSLR cameras. Both cameras shared identical settings (e.g. aperture, ISO, shutter speed) and their focal lengths were 32 mm. The image resolution was 2282 x 1167 pixels. An example of an image taken by the left camera of the simulated barnacles and sea anemones is shown in Figure 2.

3.3 Synthetic data

This study used synthetic data as it provided a way of imitating the texture and irregular shape of natural marine growth in a typical underwater inspection setting, whilst retaining complete control of the scene and the cameras. As a reference, a photograph of natural marine growth on a real world structure is presented in Figure 3. The synthetic data consisted of a virtual 3D model of a jacket-type platform that was colonised by marine growth, as shown in Figure 4. The 3D model was set in a physically accurate underwater environment, where the common issues that affect underwater optical imaging systems were present, such as scattering and colour absorption. Other factors that occur in real world scenes, such as luminous complexities (e.g. bright spots caused by the use artificial lighting, surface reflections and caustics) and marine snow (turbidity) were also accounted for.

![Figure 2. Simulated marine growth species taken by the left stereo camera.](image)

![Figure 3. Real world structure affected by marine growth.](image)

![Figure 4. Rendered image of marine growth colonised structure taken by the left camera of the virtual stereo system.](image)

The checkerboard based calibration for the virtual stereo system was carried out in the same manner as the calibration for the real stereo system. It entailed capturing 14 images of a checkerboard pattern, held at various random locations and orientations, in the underwater scene as the calibration phase must be done under the same conditions as that of the main image acquisition phase. An example of an image taken by the left virtual camera of the checkerboard is shown in Figure 5.
4 RESULTS

This section presents the estimated camera calibration matrices based on the checkerboard calibration and self-calibration procedures. The results of the 3D shape recovery are also provided.

4.1 Results for Real Data

The calibration matrices for the left camera, $K$, as estimated using checkerboard calibration and self-calibration are as follows:

$$
K_{	ext{Checkerboard}} = \begin{bmatrix}
3686 & 0 & 1221 \\
0 & 3717 & 817 \\
0 & 0 & 1
\end{bmatrix}
$$

$$
K_{\text{Self-calibration}} = \begin{bmatrix}
3649 & 0 & 1296 \\
0 & 3624 & 864 \\
0 & 0 & 1
\end{bmatrix}
$$

The calibration matrices for the right cameras are similar; they are omitted due to space constraints. Using these calibration matrices, the scene was reconstructed and the height of the marine growth species was estimated. A view of the reconstructed scene using self-calibration is shown in Figure 6. The estimated height values corresponding to each calibration approach are presented in Table 1.

Table 1. Reconstruction Accuracy

| Marine Growth Species | No. of cases | Actual Height | Root-Mean-Square Error | \begin{tabular}{ll} 
Checkerboard calibration \ & Self-calibration \ & \end{tabular} |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea Anemones (small)</td>
<td>7</td>
<td>25 mm</td>
<td>3.9 mm</td>
<td>8.6 mm</td>
</tr>
<tr>
<td>Barnacles</td>
<td>7</td>
<td>25 mm</td>
<td>8.2 mm</td>
<td>7.7 mm</td>
</tr>
<tr>
<td>Sea Anemones (medium)</td>
<td>2</td>
<td>35 mm</td>
<td>4.4 mm</td>
<td>4.7 mm</td>
</tr>
<tr>
<td>Sea Anemones (large)</td>
<td>3</td>
<td>40 mm</td>
<td>3.4 mm</td>
<td>6.3 mm</td>
</tr>
</tbody>
</table>

4.2 Results for Synthetic Data

The calibration matrices for the left camera, $K$, as estimated using checkerboard calibration and self-calibration, along with the ground truth matrix (which is known and can be controlled since virtual cameras are being used), are as follows:

$$
K_{\text{Checkerboard}} = \begin{bmatrix}
1334 & 0 & 801 \\
0 & 1334 & 601 \\
0 & 0 & 1
\end{bmatrix}
$$

$$
K_{\text{Self-calibration}} = \begin{bmatrix}
1340 & 0 & 804 \\
0 & 1324 & 600 \\
0 & 0 & 1
\end{bmatrix}
$$

$$
K_{\text{Ground Truth}} = \begin{bmatrix}
1333 & 0 & 800 \\
0 & 1333 & 600 \\
0 & 0 & 1
\end{bmatrix}
$$

Depth maps were generated using camera matrices based on the checkerboard calibration and self-calibration, as shown in Figure 7a and 7b respectively. Comparison with the ground truth depth map, which is illustrated in Figure 7c, reveals that the Root-Mean-Square Error (RMSE) for the checkerboard calibration was 22.9 cm, while the RMSE for the self-calibration was almost identical at 23 cm. These depth maps represent the actual distance of points in the scene from the camera. Only visible points on the structure were considered for the purpose of assessing the reconstruction error more effectively.
4.3 Discussion

For the real cameras, the focal lengths estimated through self-calibration and checkerboard calibration showed good agreement with each other. While there was a more notable difference between the estimated principal point values, this did not translate into a significantly errant reconstruction as both calibration methods produced good reconstructions, as evident from the generally low RMSE reported in Table 1. It is apparent though that the checkerboard calibration offers some accuracy improvements over self-calibration.

For the virtual cameras, the estimated camera matrices were very close to the ground truth values, although once again, the checkerboard calibration produced marginally better results. These findings suggest that checkerboard calibration should be undertaken to obtain the best results, however, self-calibration can be a good alternative if checkerboard calibration is not possible.

CONCLUSION

Cameras are a versatile tool capable of making visual data a part of quantitative assessment. In order to extract quantitative information from the visual data, it is necessary to perform camera calibration. This task is often overlooked despite being a crucial part of the imaging pipeline. With this in mind, it is of great practical importance that inspectors are aware of the various camera calibration approaches available and have knowledge about their respective advantages and limitations so that they can fully capitalise on the power of cameras as a quantitative assessment tool and integrate them into the inspection framework.

This paper presents a comparison between two of the most prominent classes of camera calibration techniques, namely, the classical checkerboard procedure and self-calibration. These calibration procedures are evaluated on controlled shapes that are representative of macro-scale marine growth, as well as on synthetic data which has been specifically designed to reflect some of the deleterious effects and phenomena encountered in underwater environments, such as luminous complexities, poor visibility, light attenuation and backscatter which diminish the ability of the camera, and calibration procedures, to effectively identify and quantify instances of deterioration. Results indicate that checkerboard calibration is noticeably more accurate than self-calibration; however, the practical advantages of using self-calibration may outweigh this reduction in accuracy.

ACKNOWLEDGMENTS

The authors wish to thank Science Foundation Ireland and Capacités SAS/IXEAD society for supporting this research.

REFERENCES


