3-22

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

# A Comparison of New Generic Camera Calibration with the Standard Parametric Approach

Aubrey K. Dunne John Mallon Paul F. Whelan
Vision Systems Group
Dublin City University
Dublin
aubrey.dunne@eeng.dcu.ie

## Abstract

*This paper deals with a recently proposed non-parametric approach to camera calibration, which is applicable to any type of sensor design. Currently, no relative quantitative performance data is available for this method. This paper addresses this issue, by providing a comprehensive evaluation with respect to the standard planar calibration technique in the literature. Experiments are conducted on simulated and real data, with the firm conclusion that the generic calibration method has the capability to outperform the standard parametric approach for imaging systems with significant distortion. The results provide important practical information for the vision community at large.*

## 1 Introduction

For many applications, from mobile robot localisation to security, there is an increasing trend towards using non-traditional imaging modalities. This has seen a more widespread use of cameras which are often collectively referred to as omnidirectional. In comparison with regular pinhole cameras, they offer the primary advantage of a greater field of view. This has raised the issue of calibrating such imaging devices. Calibration essentially refers to discovering the relation of the 3D Euclidean world to the 2D image space [1]. This paper evaluates a powerful new generic non-parametric calibration strategy that is applicable to any type of sensor. The goal of this paper is to inform the practitioner on the precision of this new technique, with experiments being conducted with reference to the standard pinhole camera calibration (including lens distortion).

Calibration of a perspective camera amounts to finding a linear mapping in projective space encoding the internal camera parameters such as focal length and principal point. However, many real camera systems cannot be fully described by this linear or pinhole model. For example, all conventional lenses exhibit some degree of non-linear lens distortion. For accurate calibration, this non-linearity must also be incorporated into the modelling. There are many well established techniques that achieve this [2, 3, 4]. However, as this lens distortion increases (as the angle of view increases) the distortion modelling in the above techniques begins to fail. This in turn has seen the emergence of many higher order parametric modelling techniques [5, 6], which are suitable for use up to fish-eye distortion levels.

There are also some central projection systems whose geometry cannot be described using the conventional pin-hole model. Catadioptric systems [7] are an example, where a wide field of view and a unique projection center are attained by combining a hyperbolic mirror with a perspective camera. The modelling of such a system is proposed in [8], as a mapping to a sphere and then to a plane. This is further extended in [9] to include radial distortions. Also, circular field of view sensors such as fisheye and circular mirrors have been calibrated [10], through a generalisation of the divisional model and polynomial eigenvalue solution employed in [11]. One can note two things from these works: that they all employ parametric calibration techniques, and that for each different sensor type, a different parametric representation is required. This latter issue can become a problem for the practitioner as it raises the issue of which modeling technique is best to choose. It also constrains sensor design to a predefined set.

All of the above methods attempt to model the passage of the light rays through the camera optics with a parametric model. An alternative non-parametric camera model that associates a ray in 3-space with each pixel in the image was introduced by Grossberg [12]. This generic calibration idea is powerful as it can be applied equally well to all types of cameras, from pinholes, fish-eye and omnidirectional to non-central cameras such as stereo systems. Recently, a more general approach was developed by [13, 14, 15]. In this general method, three views of a planar grid are required but the location of planes is not required a-priori. This calibration technique has not as yet received significant attention within the vision community. Several researchers have used this calibration technique [16], but mainly for the purpose of extending the calibration technique to the motion estimation task.

This paper proposes to benchmark the performance of this generic central calibration technique [13] with respect to the well known and understood perspective camera calibration technique of Zhang [17]. Our goal is to inform the practitioner, and the broader community, of the level of precision to be expected with the generic calibration strategy. Currently no relative performance data has been reported in the literature. As the comparison is conducted with respect to a radially distorted perspective model, this is the imaging modality that is utilised. Favorably, due to the general properties of the generic technique, the same performance levels can be equally extended to other imaging modalities. The analysis is conducted with real and simulated data, and the results are compiled in terms of noise sensitivity, the removal of nonlinear distortion, and motion estimation tasks. Our results clearly show that the performance of the generic calibration technique is effectively independent of the nonlinearities in the imaging sensor, and thus it outperforms the existing standard method for sensors incorporating significant distortion.

## 2 Calibration Methods

The work of [2] and [3] was the first that fully exploited multiple views of planar grids taken from unknown viewpoints. These works describe how to obtain linear constraints on the intrinsic parameters of the camera from a minimum of two homographies between the scene and image planes. Mainly due to an executable distributed by Zhang[1] and an Open Source implementation within the OPENCV library[2], this plane-based approach has become a standard tool for calibrating cameras. Thus, as this technique is widely used, and its performance has been well characterised, we use it to benchmark the performance of the generic calibration technique. Lens distortion is included in this technique by including a polynomial model in a final full nonlinear estimation process, often referred to as bundle adjustment.

### 2.1 Generic calibration method

Generic camera calibration is a non-parametric camera calibration method that calculates the ray direction associated with each pixel point (pixel level calibration). The technique has recently been proposed by Sturm and Ramalingam[13]. By circumventing the need for parametric camera models it is rendered generally applicable, even capable of catering for discontinuous cameras (something that was not previously possible). The method can be used for both central cameras, as used in this paper, and for non-central cameras (cameras for which the pixel rays do not converge at a single point).

The generic method aims to simply determine the ray direction corresponding to each image pixel. This is achieved by determining the points seen by a pixel on each of three differently orientated grids. Once each ray direction has been calculated, it is stored in a look up table that maps to the correct image pixel. The process can be summarised as follows:

1. Take a minimum of three initial images of a calibration grid in different orientations. Additional images are also required in various orientations to completely cover the image
2. For each pixel, determine the location seen by that pixel on each grid via homographic interpolation
3. Linearly estimate the effective centre of projection of the camera, and the orientations of the calibration grids, using this data and the known constraints
4. Refine the orientations of the initial grids and the ray directions in a bundle adjustment stage
5. Estimate the orientations of the additional grids using geometric constraints followed by bundle adjustment
6. Store the ray directions (as Plucker matrices) in a look-up table

The density of the feature points on the grids is an important factor in the accuracy of the calibration. For a pixel level calibration, if the feature points are not dense enough, then the result will display systematic errors resulting from the bias of the interpolation towards the periphery of the image. In order to satisfy the requirement for a dense feature set, techniques from the domain of structured light can be employed to directly encode grid location [18]. This is done via *active grids*, whereby a series of 22 greyscale patterns (Gray coded binary and sinusoidal) are consecutively displayed on a flat screen TFT monitor. The location on the active grid seen by each camera pixel is calculated from the intensity of that pixel in the images of the patterns. A noise analysis of the locations decoded from the active grids found that the inherent noise in the method is comparable to the effective detection accuracy of standard feature detectors.

## 3 Experiments

The results are primarily compiled over simulated data. These findings are subsequently validated on real image samples. For the simulated data, comparative experiments are designed to characterise the sensitivity to noisy input data, and to assess the nonlinearity removal with respect to increasing lens distortion or field of view. With real images, three samples with increasing field of view are used. These are similarly analysed for residual distortion levels and an additional motion estimation experiment.

### 3.1 Simulated data

Using simulated data, two experiments are conducted. Firstly, the sensitivity to noise of the generic calibration technique is compared with the sensitivity to noise of the standard method. Increasing levels of noise are induced in the input data and then each calibration technique is implemented. This process is conducted 50 times for each level of noise, whereupon the statistics shown in Fig. 1 (top) are computed. Note that for the generic algorithm the noise causes millimetre error in the ray-plane intersections. This millimetre noise was computed by a backprojection of the noisy pixel data onto each plane. As with the standard method, the effect of noise in the generic method is shown to be linearly proportional.

The second experiment aims to investigate the precision of the calibration in terms of removing nonlinearities such as lens distortion. Increasing levels of lens distortion are simulated according with an equidistance projection function: $r = f\theta$. This is a common distortion function that also allows distortion to be related to the field of view. This model is manipulated to give an increasing field of view from $30^o$ to $150^o$. This model is chosen to be different from the one used in the standard calibration algorithm. The distortion residuals following calibration are compiled over 50

[1] http://research.microsoft.com/-zhang/calib
[2] http://www.intel.com/technology/computing/opencv/index

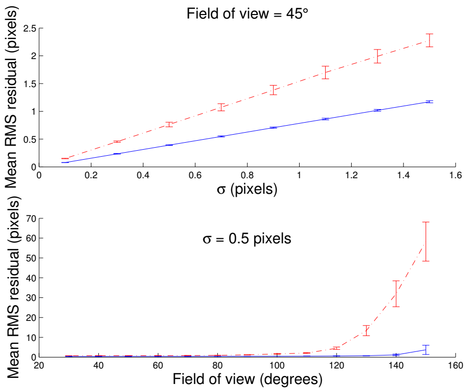


Figure 1. Mean and SD of noise sensitivity (top) and distortion residuals (bottom) for standard method (dashed) and generic method (solid).

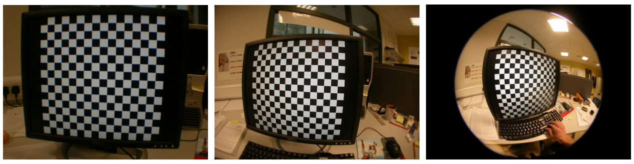

Figure 2. Three levels of distorted images used.

iterations with 1000 random points. Fig. 1 (bottom) shows the resulting statistics. As expected, for high distortion the residuals on the standard method increase. However, the generic algorithm maintains a low mean and standard deviation throughout. This level of performance can be expected regardless of the imaging modality.

Note that forming distortion corrected images is a 3-space operation within the generic calibration approach. As each pixel maps to a ray, distortion correction amounts to determining the intersections of these rays with a synthetic image plane. In our experiments, the synthetic plane is chosen as the plane perpendicular to the principal ray that passes closest to the grids used in the calibration. As the location of this plane affects the scale and location of the corrected image points, an isometry is applied to map these intersected points to their known correct positions. The isometry is linear and thus does not affect the performance of the generic method for capturing non-linear distortions.

### 3.2 Real data

As mentioned, three real images are analysed for each calibration method with respect to distortion residuals, and one image is analysed with respect to a motion estimation task. The input images are shown in Fig. 2, giving an indication of the levels of distortion present. These were taken at three different zoom levels using a Nikon CoolPix 4500 camera fitted with a FC-E8 fisheye converter (field of view $183^o$);

Following calibration with both the standard and generic techniques, the distortion residuals are measured by taking homographies between the known metric structure and the calibrated image. The resulting residuals are presented in Table 1. This shows that the accuracy of the generic calibration is slightly less than the standard technique for low distortion levels. As the level of distortion increases, the magnitude and standard deviation of the errors for the standard technique increase, while the generic method maintains its accuracy throughout. For sample number 3, a vector plot of the residuals for each method is shown in Fig. 3. This shows a classic distortion bias pattern for the standard technique, whereby the parametric model overcompensates for distortion towards the image (distortion) centre, and undercompensates towards the image periphery. In contrast, although there is a systematic error in the generic residuals, there is no distortion bias present. The systematic error is due to error in the estimate of the camera centre.

Table 1. Mean and SD of the distortion residuals after calibration for three real images.

| Sample | ID | 1 | 2 | 3 |
|---|---|---|---|---|
| Standard technique | RMS | 0.0880 | 0.1296 | 1.1983 |
| | SD | 0.0344 | 0.0635 | 0.6711 |
| Generic technique | RMS | 0.6503 | 0.7195 | 0.6811 |
| | SD | 0.3531 | 0.3906 | 0.3831 |

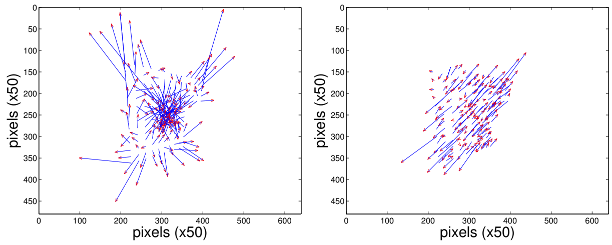


Figure 3. Vector plots of residuals for standard method (left) and generic method (right).

These results are in broad agreement with the simulated results - the error in the standard method increases as the distortion increases, whereas the generic method is not sensitive to changes in distortion. The difference in error magnitudes between the simulated data and real data results can be attributed primarily to the images used in the real experiments (Fig. 5). Only the areas covered by the grids in the images are undistorted, and thus distortion residuals for the periphery of the images, where the distortion is greatest, are not calculated. Grids do not cover complete images due to the difficulty in accurately extracting corners of severely distorted grids.

A second experiment is also conducted to assess the calibration precision. This involves calibrating with each method, and then performing a motion estimation task. Images are taken of a planar grid attached to a linear motion controller. Five images are taken with translation increments of exactly 25mm. Homographies are taken between each step and subsequently decomposed to recover the motion. For calibrated images the homography $\mathbf{H}$ can be decomposed as $\mathbf{H} = (\mathbf{R}_{3\times2}|\mathbf{Rt})$, where $\mathbf{R}$ is the relative rotation and $\mathbf{t}$ is the translation. These translation vectors are plotted end to end in Fig. 4. For visualisation purposes, the difference between each vector and the average vector is scaled by twenty. As can be seen, the generic method outperforms in the recovery of the translation component.

Lastly, samples of undistorted images resulting from each method are illustrated in Fig. 5. Again the generic method is seen to outperform the standard method around

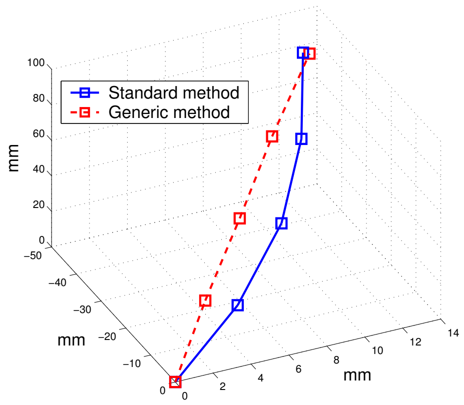


Figure 4. Translation estimation for each method. Vector errors are scaled ×20.

Figure 5. Original distorted image, standard method correction and generic method correction.

the periphery of the image.

## 4 Conclusion

This paper deals with a recent method for camera calibration that is suitable for all types of sensors. This generic property clearly makes it attractive for many applications, especially for calibrating non-perspective sensors. However, currently the expected performance of the generic method is unclear. This paper addresses this performance issue, providing a side by side comparison with the well established standard perspective camera calibration technique. Our experiments are conducted with real and simulated data, for sensitivity to noise, distortion residuals, and a motion estimation task. From these results we can conclude that the recently proposed generic calibration technique achieves good performance levels at low to mid distortions, although the standard method performs better for these distortions. For higher distortion levels the accuracy of the generic method is maintained, whilst the accuracy of the standard method significantly reduces. Overall, this performance characterisation furnishes important practical information to the vision community, clearly showing how the precision of the generic method compares with the well established standard technique. It is concluded that the generic method is particularly suited to the calibration of high distortion sensors, and should be used in place of the standard method for such sensors.

## References


[1] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge, 2 edition, 2003.

[2] P. F. Sturm and S. J. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *IEEE conference on computer vision and pattern recognition*, volume 1, pages 432–437, 1999.

[3] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[4] J. Heikkila. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, 2000.

[5] S. Shah and J. K. Aggarwal. Intrinsic parameter calibration procedure for a (high-distortion) fish-eye lens camera with distortion model and accuracy estimation. *Pattern Recognition*, 29(11):1775–1788, 1996.

[6] J . Kannala and S. S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340, 2006.

[7] S. Baker and S. Nayar. A theory of catadioptric image formation. In *Proceedings of IEEE International Conference on Computer Vision*, pages 35–42, 1998.

[8] C. Geyer and K. Daniilidis. An unifying theory for central panoramic systems and pratical implications. In *Proceedings of European Conference on Computer Vision*, pages 445–461, 2000.

[9] J. Barreto. A unifying geometric representation for central projection systems. *Computer Vision and Image Understanding*, 103(3):208–217, 2006.

[10] B. Micusik and T. Pajdla. Structure from motion with wide circular field of view cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1135–1149, 2006.

[11] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 125–132, 2001.

[12] M. D. Grossberg and S. K. Nayar. A general imaging model and a method for finding its parameters. In *Proceedings of the 8th International Conference on Computer Vision*, volume 2, pages 108–115, 2001.

[13] P. Sturm and S. Ramalingam. A generic calibration concept: Theory and algorithms. Rapport de Recherche 5058, 2003.

[14] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In *Proceedings of the European Conference on Computer Vision, Prague, Czech Republic*, volume 2, pages 1–13. Springer, 2004.

[15] S. Ramalingam, P. Sturm, and S. Lodha. Theory and experiments towards complete generic calibration. Rapport de Recherche 5562, 2005.

[16] R. Pless. Using many cameras as one. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages II: 587–593, 2003.

[17] Z. Zhang. A flexible new technique for camera calibration. Technical report, Microsoft research, 1998.

[18] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems, 2004.