Practical Intraoperative Stereo Camera Calibration

Philip Pratt, Christos Bergeles, Ara Darzi, and Guang-Zhong Yang

Hamlyn Centre for Robotic Surgery,
Imperial College of Science, Technology and Medicine,
London SW7 2AZ, UK
{p.pratt,c.bergeles,a.darzi,g.z.yang}@imperial.ac.uk

Abstract. Many of the currently available stereo endoscopes employed during minimally invasive surgical procedures have shallow depths of field. Consequently, focus settings are adjusted from time to time in order to achieve the best view of the operative workspace. Invalidating any prior calibration procedure, this presents a significant problem for image guidance applications as they typically rely on the calibrated camera parameters for a variety of geometric tasks, including triangulation, registration and scene reconstruction. While recalibration can be performed intraoperatively, this invariably results in a major disruption to workflow, and can be seen to represent a genuine barrier to the widespread adoption of image guidance technologies. The novel solution described herein constructs a model of the stereo endoscope across the continuum of focus settings, thereby reducing the number of degrees of freedom to one, such that a single view of reference geometry will determine the calibration uniquely. No special hardware or access to proprietary interfaces is required, and the method is ready for evaluation during human cases. A thorough quantitative analysis indicates that the resulting intrinsic and extrinsic parameters lead to calibrations as accurate as those derived from multiple pattern views.

1 Introduction

The provision of stereo camera systems during minimally invasive procedures affords the surgeon with a visual experience and appreciation of depth close to traditional open surgery. Through the calibration of an appropriate camera model, this also facilitates image guidance applications, comprising visualisation of preoperative and intraoperative images, physical measurement via triangulation of point-wise correspondences, spatial registration and geometrical reconstruction of the surgical scene. Although 3D systems with distally mounted sensors and large depths of field are on the market, many of the stereo systems in active clinical use fall into the category of rod-lens endoscopes attached to twin distal cameras. Depths of field are typically shallow, and actuated focusing mechanisms are usually provided. During interventions, the surgeon adjusts focus in accordance with camera motion and proximity to anatomical targets. However, in the general case, such a change to the camera optical characteristics renderers any prior set of calibrated intrinsic and extrinsic parameters invalid, and thus any image guidance application ineffective or, at least, inaccurate.

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 667–675, 2014.

The notion of being able to perform calibration 'before the procedure' is somewhat unrealistic, in the sense that the surgeon must first insert the endoscope into the patient in order to determine the optimal initial focus setting. Even with a considerable degree of experience in the operating theatre, the ubiquitous calibration technique employing the capture of multiple views of a calibration pattern with known geometry can then take several minutes to complete. Unless camera calibration is near instantaneous and can be performed with the endoscope in position, a genuine barrier to rapid clinical translation and widespread adoption of a large class of image guidance applications exists. Ultimately, manufacturers will provide built-in means to do this through the lookup of pre-calibrated parameter settings based on interfaces capable of reading actuated focus positions. Until such time, a practical and easily reproducible interim solution is required, based on standard software components. Robustness and repeatability of results are of paramount importance.

Several schemes have been proposed to date, although none of them fulfills all of these criteria adequately. In the context of stereo image-guided microscope-assisted interventions, Edwards *et al.* [1] expressed the variation of the parameters of a distortion-free pinhole camera model as bivariate polynomial functions of zoom and focus, whose input values were provided by rotary encoders. A custom-made reference object was used to provide point correspondences at a large number of settings over the applicable 2D zoom and focus range. Alternatively, self-calibration approaches, including those of Schmidt *et al.* [2], Stoyanov *et al.* [3] and Dang *et al.* [4], seek to refine subsets of the camera parameters in a continuous manner by tracking features in the observed scene, and where applicable, identifying correspondences and imposing additional geometric constraints. However, these methods fail to accommodate variation in the magnitude of lens distortion typically caused by focus adjustment. Barreto *et al.* [5] describe an algorithm requiring only a single image of a planar grid pattern, showing that a camera using the 'division model' for radial lens distortion can be calibrated from just one pattern-to-image homography. However, the dependency on such a relatively small number of input data points overall must account for the apparently large observed parameter standard deviations.

Most optical systems are subject to varying levels of aberration and distortion in response to physical changes in their configuration (figure 1), and the novel calibration method described herein seeks to exploit such imperfections. Each stereo camera parameter is modeled as a simple function of a single 'focus position' variable, fitted to a sequence of accurate, pre-calibrated values across the desired focus range, in the expectation that a single presentation of a reference pattern is sufficient to manifest the degree of aberration and perform the 1D search for the prevailing focus position.

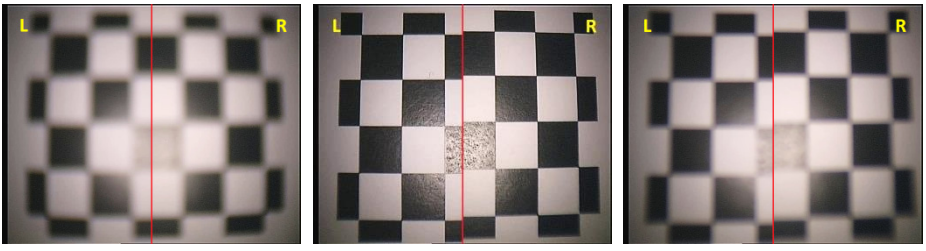

Fig. 1. Combined left and right pattern views across the range of focus positions

2 Materials and Methods

2.1 Stereo Camera Model

The model and reference calibration procedure follow Zhang [6] and Bouguet [7]. They include both radial and tangential distortion, using the second and fourth radius powers. The focal length parameters f_x and f_y in the left and right cameras are allowed to vary independently. For each camera, points $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ are projected to screen coordinates (u, v) using (1), where $\mathbf{R}$ and $\mathbf{t}$ are the extrinsic rotation and translation, accounting for 6 degrees of freedom. The coefficients k_1, k_2 and p_1, p_2 determine the levels of radial and tangential distortion, respectively. Including the principal points (c_x, c_y), a total of 8 + 8 + 6 = 22 separate parameters is optimised. The canonical OpenCV implementation [8] is employed when performing calibrations using multiple pattern images, wherein the Levenberg-Marquardt algorithm is used to refine initial parameter estimates and minimise total reprojection error globally.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{R} \begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{bmatrix} + \mathbf{t}$$

$$x' = x/z$$

$$y' = y/z$$

$$x'' = x'(1 + k_1 r^2 + k_2 r^4) + 2p_1 x' y' + p_2(r^2 + 2x'^2) \quad (1)$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2y'^2) + 2p_2 x' y'$$

$$r^2 = x'^2 + y'^2$$

$$u = f_x\, x'' + c_x$$

$$v = f_y\, y'' + c_y$$

2.2 Correlation and Sensitivity Analysis

In this study, two separate endoscope and camera head assemblies were used: a clinically active *da Vinci* Si Surgical System (Intuitive Surgical, Sunnyvale, CA, USA), supporting HD (1920x1080) video feeds; and an original first-generation system with PAL (720x576) resolution. Careful observation of reference patterns with adjacent left and right views over the range of focus positions revealed definite trends in camera behaviour. The HD system is characterised principally by varying focal lengths and horizontal disparity, whereas the SD system exhibits noticeable changes in the amount of radial distortion and both horizontal and vertical disparity. Despite achieving relatively low mean reprojection errors, typical calibration procedures [8] exhibit a reasonable degree of instability on account of the very high number of dimensions in which the optimisations are performed. Therefore, in order to verify that observed

camera behaviour can be modelled in the presence of such noise as a function of focus position, a detailed correlation and sensitivity analysis was performed.

For each camera and at 6 monotonically increasing and approximately equally-spaced focus positions, 25 calibration repetitions were performed using 50 pattern image pairs drawn uniformly at random from collections of between 72 and 90 pairs overall. Inconsequentially, the number of available pairs varied with footage length. Linear least-squares regressions between each of the camera parameters and the dimensionless focus position, denoted γ, were performed. Simple linear relationships were chosen in the knowledge that the aberrations induced by changes in focus are relatively subtle. The R^2 values of the Pearson correlation coefficient were calculated. Furthermore, to estimate the sensitivity of the camera model to indicative changes in input parameters, an illustrative 21x21 lattice uniformly covering the (X, Y) range [-20mm, 20mm] at a view-filling distance $\mathrm{Z} = 42.5\mathrm{mm}$ was projected to the (u, v) frame. For each parameter in turn, two-sided perturbations were applied to the third, i.e. centremost, focus position in accordance with the regressed values at the extremal positions. Explanatory parameters must have both high correlation and sensitivity.

2.3 Dimensionality Reduction

Expressing the camera parameters as linear functions of the single focus position variable γ reduces the dimensionality of the calibration process to one. Using the error metric illustrated in figure 2, a single pattern presentation can suffice. Firstly, the left and right poses $\mathbf{E}_L$ and $\mathbf{E}_R$ are estimated [8] in accordance with their respective intrinsic matrices $\mathbf{I}$ and distortion coefficients $\mathbf{d}$, themselves functions of γ. Secondly, the extrinsic transformation between left and right cameras, when salient, is determined by direct interpolation and orthonormalisation of matrix entries, as justified for limited rotations.

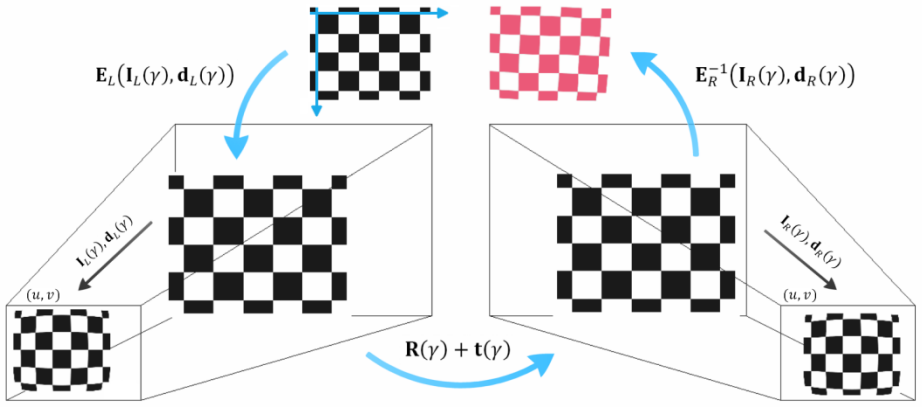

Fig. 2. Round trip of pattern corners back to pattern frame (target highlighted in red)

The N points $\mathbf{p}_i$ comprising the reference pattern are transformed from its own frame into the left camera frame, then back to the original pattern frame via the right camera. Thus, as shown in equation (2), the mean round trip error M over all points will depend on focus position γ, and it is this function that is minimised in order to determine the prevailing calibration settings. The function is very quick to evaluate and a trivial 1D incremental search is perfectly adequate.

$$M(\gamma) = \frac{1}{N}\sum_{i=1}^{N} \left\| \mathbf{E}_R^{-1}(\gamma)\,[\mathbf{R}(\gamma) + \mathbf{t}(\gamma)]\mathbf{E}_L(\gamma).\mathbf{p}_i - \mathbf{p}_i \right\| \tag{2}$$

2.4 Validation

Calibration quality was assessed by comparing single view results to those from traditional multiple view calibrations, using 22-25 image pairs, at a number of arbitrary focus positions within the modelled range. Rather than compare camera parameters directly, which can prove difficult to interpret, this study considers derived physical quantities and related algorithms dependent on the former. Firstly, point triangulation was used to measure lengths between gradations on the shaft of a pair of calipers viewed in a number of arbitrary positions. The points were picked manually in the left and right views. Secondly, the ability of the calibration to produce disparity maps in a robust manner was tested using a semi-global block matching algorithm [9]. Bouguet's method [8] for stereo rectification was employed. This test effectively measures the degree to which correspondences are made to lie on the same scanline. Examining the vicinity of the pattern grids, the pixel-wise densities of valid disparity regions were calculated. Since the new calibration scheme is effectively a one-time fit over the whole focus range, validation was performed deliberately both before and after the endoscope had been disconnected and re-attached to its camera head with the bayonet mount (figure 3). Their relative positioning must prove to remain the same.

2.5 Deployment Scenarios

Two clinical usage scenarios were considered. For procedures where the endoscope enters the body through a cannula, sterilisable *KeyDot*® markers (KeySurgical Inc., Eden Prairie, MN, USA) can be attached to conventional surgical instruments and probes [10]. Figure 3 illustrates the chessboard pattern laser-engraved onto 6.35mm diameter acrylate discs, and the typical view afforded during ultrasound-guided robotic partial nephrectomy [11]. In cases where camera access is more superficial and made via a wide orifice, e.g. transoral robotic surgery (TORS), a perishable non-sterile laser-printed pattern will suffice. In both situations, to avoid harsh reflections and sensor saturation, it is necessary to reduce the intensity of illumination by curtailing the light fountain, but this is immediate and does not impede surgical workflow.

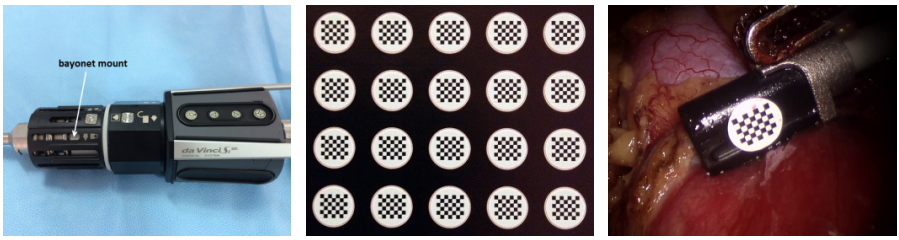

Fig. 3. Endoscope bayonet mount; KeyDot patterns; attachment to existing surgical probe

3 Results

3.1 Correlation and Sensitivity Analysis

For each of the calibrated intrinsic parameters in the stereo camera model, table 1 shows the squared coefficient of correlation with respect to focus position, and the total pixel displacement Δ resulting from the two-sided perturbations described in section 2.2. Results are tabulated for both HD and SD camera heads, and their respective left and right camera models. Entries with the most prominent correlations and sensitivities are highlighted. These results proved to be consistent with observation, and established that linear regression and a 1D error minimisation approach are plausible means for performing accurate single view calibrations.

Table 1. Squared correlation coefficients and total pixel displacements

	HD (left) R^2	HD (left) Δ	HD (right) R^2	HD (right) Δ		SD (left) R^2	SD (left) Δ	SD (right) R^2	SD (right) Δ
f_x	**0.8463**	**1321.42**	**0.8723**	**1487.56**		0.0168	49.13	0.0003	6.16
c_x	**0.7653**	**11443.95**	**0.5956**	**7613.20**		**0.6189**	**4463.80**	0.0189	472.53
f_y	**0.8154**	**1296.45**	**0.8331**	**1314.96**		0.0084	40.89	0.0029	21.43
c_y	**0.6816**	**4324.67**	**0.7776**	**4885.84**		**0.3951**	**2477.54**	0.0603	853.56
k_1	0.5563	638.67	0.2053	303.73		**0.9194**	**1135.91**	**0.8907**	**1874.32**
k_2	0.0285	131.92	0.0301	223.62		0.0224	96.64	0.0215	201.38
p_1	0.2608	126.59	0.6522	332.48		0.5149	241.56	0.3042	182.72
p_2	0.0003	3.16	0.0114	21.49		0.5821	259.29	0.8460	372.36

As anticipated, the extrinsic parameters were found to be essentially constant over the focus range. In the HD case, the largest Frobenius norm of the rotation with respect to the identity matrix was 0.017, and the (x, y, z) translation standard deviations were found to be 0.008mm, 0.002mm and 0.048mm, respectively. Likewise, for the SD case, the largest rotation norm was 0.047, with translation deviations of 0.033mm, 0.026mm and 0.064mm, respectively.

3.2 Validation

For each discrete position at which focus data was collected, and using a single image pair from the respective video footage used to generate the interpolated sequence of calibrations, figure 4 shows plots of the mean round trip error against focus position for the HD case. These represent the initial verification of the single-view calibration procedure, where the minima of the round trip error graphs are expected to be coincident with their respective ground truth focus positions. This is a test of the method's ability to recover the input data from the fitted model, which provides interpolated calibrations across the continuum of focus positions. The recovered abscissae were found to be 1.1, 2.2, 3.2, 3.9, 4.9 and 6.0. Considering that a fast, yet precise search increment of 0.1 was used, all are in close proximity to the expected integer values. These figures have a limited physical interpretation, although they constitute a necessary but not sufficient test of the accuracy of interpolated calibrations in practice.

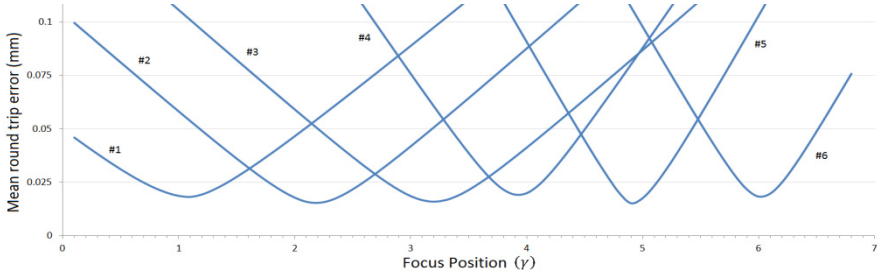

Fig. 4. Mean round-trip error for each discrete focus setting

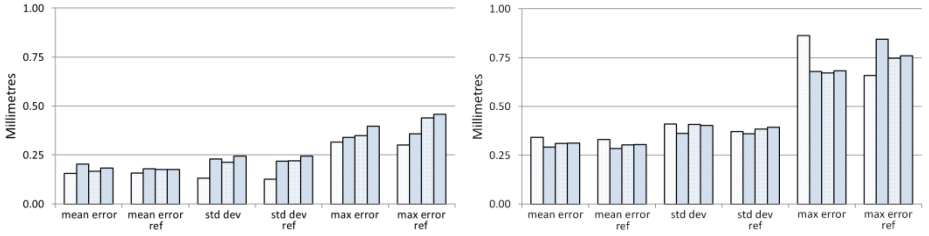

Fig. 5. Measurement errors with reference levels for HD (left) and SD (right) calibrations

Figure 5 shows the results of triangulated caliper shaft measurements. In each of four repeated runs for the HD and SD calibrations, the mean error, standard deviation and maximum error are shown for the single view and reference multiple view cases, calculated over 10 separate measurements of gradations set 20mm apart (figure 6). The HD calibrations were performed at different arbitrary focus positions using a laser-printed pattern, where the latter three runs were made after endoscope reattachment. In contrast, the SD calibrations considered a single focus position following reattachment, but with four different views of the same *KeyDot*® marker pattern.

With examples illustrated in figure 6, the four separate SD single view calibrations were able to produce reconstructed geometry with pixel densities of 99.82%, 99.84%, 100.00% and 98.56%, as accumulated over the homogeneous region comprising the chessboard grid. In comparison, the poorest reconstructions, generated from multiple view calibrations performed at the extremal focus positions, were calculated to be 74.31%, 46.53%, 10.02% and 49.83%, respectively. This shows the dramatic effect that inaccurate stereo calibrations can have on dense reconstruction performance.

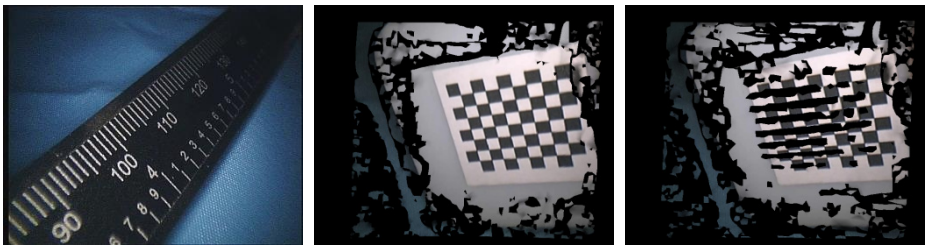

Fig. 6. Caliper shaft (left); reconstructed pattern: single view (centre) vs. extremal (right)

4 Discussion and Conclusion

In combination with stereo reconstruction performance, the comparison of single pattern view and reference measurement results indicate that the novel calibration method described herein is suitable to undergo trials in a live clinical setting. It fulfills the criteria set out in section 1, namely that the implementation is straightforward to reproduce, and that it generates accurate calibrations in a robust and repeatable manner; this is due to the camera model having been fitted to a range of focus positions simultaneously, using a large number (between 72 and 90) of input image pairs at each position. Furthermore, the round-trip error optimisation process is nearly instantaneous. This and the requirement for a solitary pattern view result in very little disruption to surgical workflow. Indeed, future work will look to monitor the quality of calibration continuously and track changes in focus position automatically. In addition, an extended study will determine whether the fitted model is subject to drift over time and if necessary, set an appropriate frequency for model re-fitting. Meanwhile, the method represents an important step in breaking down the barriers to widespread adoption of many recent developments in image guidance technology.

References

1. Edwards, P., King, A., Maurer, C., De Cunha, D., Hawkes, D., Hill, D., Gaston, R., Fenlon, M., Jusczyzck, A., Strong, A., Chandler, C., Gleeson, M.: Design and evaluation of a system for microscope-assisted guided interventions (MAGI). IEEE Transactions on Medical Imaging 19(11), 1082–1093 (2000)

2. Schmidt, J., Vogt, F., Niemann, H.: Nonlinear Refinement of Camera Parameters using an Endoscopic Surgery Robot. In: Ikeuchi, K. (ed.) Proceedings of IAPR Conference on Machine Vision Applications, pp. 40–43 (2002)
3. Stoyanov, D., Darzi, A., Yang, G.Z.: Laparoscope Self-Calibration for Robotic Assisted Minimally Invasive Surgery. In: Duncan, J., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 114–121. Springer, Heidelberg (2005)
4. Dang, T., Hoffmann, C., Stiller, C.: Continuous Stereo Self-Calibration by Camera Parameter Tracking. IEEE Transactions on Image Processing 18(7), 1536–1550 (2009)
5. Barreto, J., Roquette, J., Sturm, P., Fonesca, F.: Automatic Camera Calibration Applied to Medical Endoscopy. In: Proceedings of the British Machine Vision Conference, pp. 1–10 (2009)
6. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)
7. Bouguet, J.-Y.: Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc
8. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Inc. (2008)
9. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(2), 328–341 (2008)
10. Pratt, P., Di Marco, A., Payne, C., Darzi, A., Yang, G.-Z.: Intraoperative Ultrasound Guidance for Transanal Endoscopic Microsurgery. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 463–470. Springer, Heidelberg (2012)
11. Hughes-Hallett, A., Pratt, P., Mayer, E., Di Marco, A., Yang, G.-Z., Vale, J., Darzi, A.: Intraoperative Ultrasound Overlay in Robot-assisted Partial Nephrectomy: First Clinical Experience. European Urology (2013), http://dx.doi.org/10.1016/j.eururo.2013.11.001