On Pattern Selection for Laparoscope Calibration

Stephen Thompson[1], Yannic Meurer[1], Eddie Edwards[1], João Ramalhinho[1],
Maria Ruxandra Robu[1], Danail Stoyanov[1], Sébastien Ourselin[1], Brian Davidson[2],
David Hawkes[1], Matthew J. Clarkson[1]

[1]Centre for Medical Image Computing, University College London, United Kingdom
[2]Division of Surgery and Interventional Science, UCL Medical School, United Kingdom

ABSTRACT

Camera calibration is a key requirement for augmented reality in surgery. Calibration of laparoscopes provides two challenges that are not sufficiently addressed in the literature. In the case of stereo laparoscopes the small distance (less than 5mm) between the channels means that the calibration pattern is an order of magnitude more distant than the stereo separation. For laparoscopes in general, if an external tracking system is used, hand-eye calibration is difficult due to the long length of the laparoscope.

Laparoscope intrinsic, stereo and hand-eye calibration all rely on accurate feature point selection and accurate estimation of the camera pose with respect to a calibration pattern. We compare 3 calibration patterns, chessboard, rings, and AprilTags. We measure the error in estimating the camera intrinsic parameters and the camera poses. Accuracy of camera pose estimation will determine the accuracy with which subsequent stereo or hand-eye calibration can be done. We compare the results of repeated real calibrations and simulations using idealised noise, to determine the expected accuracy of different methods and the sources of error. The results do indicate that feature detection based on rings is more accurate than a chessboard, however this doesn't necessarily lead to a better calibration. Using a grid with identifiable tags enables detection of features nearer the image boundary, which may improve calibration.

1. INTRODUCTION

Accurate calibration of laparoscope optics is essential to the ongoing development of image guidance systems for laparoscopic surgery such as in.[1–7] Intrinsic calibration determines the camera's intrinsic parameters, focal length, principal point and distortions coefficients. If the laparoscope is to be tracked using an external tracker it is also necessary to determine the location of the lens relative to the attached tracker, known as hand-eye calibration.[8–13] Similarly, for a stereo laparoscope calibration can be used to determine the position of one lens relative to the other.

Camera calibration is a widely studied problem in the field of computer vision. Zhang[14] outlined the basic approach to intrinsic calibration used here. The chessboard is widely used in camera calibration, and the algorithm to determine the chessboard corners is a well established part of the OpenCV library*. However as pointed out by Datta[15] the corner detection algorithm is susceptible to errors when the grid pattern is not orthogonal to the camera axes. The pattern of rings was proposed to avoid some of the shortcomings of the chessboard. Datta's template matching algorithm is also used to determine the centre of each ring. Both the ring and chessboard pattern suffer from the drawback that the entire pattern must be visible to enable point to point correspondence to be determined, and the orientation of the pattern cannot be determined. One way round this is to use a calibration grid of uniquely identifiable patterns,[16,17] allowing point to point correspondence to be determined for incomplete views. One such set of patterns and the algorithm to detect them is contained in the AprilTags library.[18]

Whilst there is a large body of published work on both camera calibration and hand-eye calibration, it often does not address the unique challenges of laparoscope calibration. Typically the range of views of the calibration pattern that can be gathered is limited by the geometry of the laparoscope and/or the tracking system. The relatively high radial distortion in the laparoscope optics, means that it can be difficult to detect features near the image edges reliably. Similarly the calibration images can be poorly illuminated. In the case of stereo
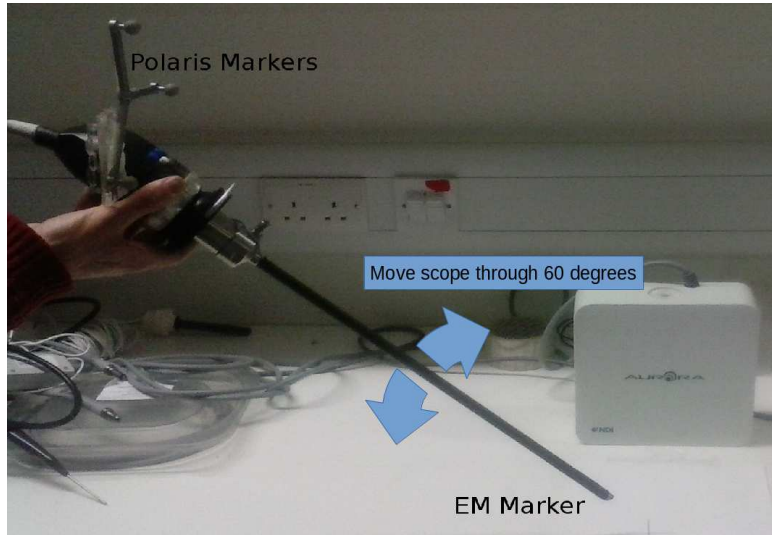
---

*www.opencv.org

Figure 1. The experiment setup for recording calibration sequences. Each calibration pattern is placed flat on the table, then the scope is moved systematically through a 60 degree cone, acquiring images of each calibration pattern from a range of views.

laparoscopes the separation between channels is small in comparison to the size of calibration grids that can be reliably imaged. This makes it difficult to accurately determine the relative position of the two lenses. The requirement to maintain sterility of the laparoscope also makes the attachment of active or passive control arms difficult, so in contrast to calibration of robot cameras, the laparoscope must be moved by hand, which makes strict acquisition protocols difficult to implement.

The work in this paper forms part of our ongoing work to develop robust and user friendly methods for the calibration of tracked, stereo laparoscopes for image guided surgery. A starting step in all existing camera calibration processes is the ability to accurately measure features in recorded images, for example, determining the position of chessboard corners. There are several published calibration patterns that provide easy to detect features. In this paper we attempt a systematic and unbiased analysis of the effect of the feature detection accuracy on the accuracy and robustness of the camera calibration. We analyse the inaccuracy in estimating the cameras intrinsic parameters (focal length, principal point, distortion) and extrinsic parameters (pose). Inaccuracy in pose estimation will directly effect subsequent stereo or hand-eye calibration.

2. METHODS

2.1 Data Acquisition

Figure 1 shows our laboratory set up. In this experiment we used a Viking stereo laparoscope[†] to acquire several thousand images of 4 different calibration patterns. The output of the laparoscope is captured via an NVidia SDI frame grabber, and saved using the NifTK software package.[19] The output from the capture card is HD video (1920x1080 pixels) at 25 frames per second. Each channel (left and right) of the stereo laparoscope video gives a 1920 x 540 pixels image. This results in an apparent vertical squashing of the saved images, which is important when using calibration patterns that assume certain pattern geometry.

For each of four calibration patterns we acquired several thousand frames of video images. As mentioned in the introduction, attaching arms to the laparoscope to control motion is not practical clinically, so to replicate clinically realistic conditions each acquisition was done freehand. Care was taken to follow the same acquisition protocol for each pattern, moving the distal end of the scope through a 60 degree cone, while keeping the calibration grid in frame, over a period of approximately 4 minutes.
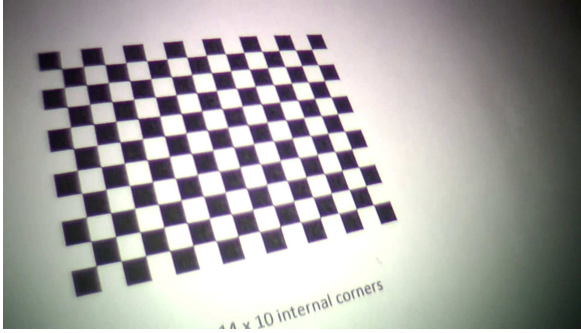

[†]www.conmed.com

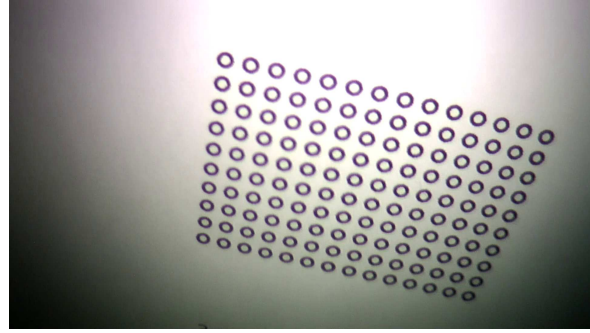Figure 2. The chessboard calibration pattern.

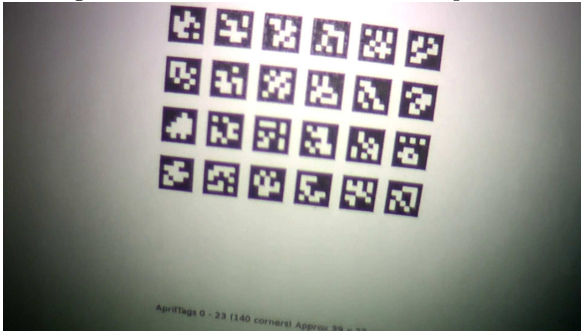Figure 3. The ring calibration pattern.

Figure 4. The 24 tag AprilTag calibration pattern.

Figure 5. The 216 tag AprilTag calibration pattern.

2.2 Calibration Patterns

Four calibration patterns were used. The first of these was a chessboard pattern, see figure 2, with 14 x 10 internal corners, with a square size of 3 mm. Such patterns are widely used in the literature and in our experience provide reliable calibration. However, due to distortions near the edges of laparoscopic images feature detection of straight lines can be difficult. One method to combat this is to use round features.[15] We used a grid of rings, with geometry matching the chessboard, (14 x 10, with a pitch of 3mm), see figure 3.

Both of these patterns suffer the drawback that as each detected feature is not uniquely identifiable, only frames where all features (140) have been detected can be used for calibration. One way around this is to use a grid of identifiable tags, enabling any frame with detected corners to be used. We used two patterns based on AprilTags.[18] The first pattern consisted of 24 tags, arranged in a 6x4 grid, resulting in 120 uniquely identifiable features (4 corners per tag + the tag centre), see figure 4. The grid was arranged in such a way as to approximate as closely as possible the geometry of the chessboard grid. Each tag was 5.25 x 5.25 square with a horizontal pitch of 6.73 mm and a vertical pitch of 7.26 mm. To take advantage of the ability to use partial views of the grid we also used a larger AprilTag grid, using the same tag size and pitch, but with a grid of 18 x 12 tags, see figure 5. Both grids used AprilTag's 25h7 tag family. Calibration grids were generated using ImageMagick [‡] and printed at 600 DPI using a Canon Image Runner Advance 4251.

2.3 Feature Detection

For each frame of saved video images, feature were detected as follows. For the chessboard pattern the chessboard corner detectors from the OpenCV library were used [§]. First cv::findChessboardCorners, then cv::cornerSubPix was used to refine the detection. Only frames where 140 features were detected were accepted. For the rings patterns two detectors were compared, similarly to Datta el al.[15] In the first case, an OpenCV simple blob detector was used with adaptive thresholding and cv::findCirclesGrid to find both the inner and outer circles of the rings. The ring centres were then used as feature locations. In the second case we implemented template

---

[‡]www.imagemagick.org

[§]www.opencv.org

Description	Notation (left/right)	Units
Horizontal focal length	f_xL / f_xR	pixels
Vertical focal length	f_yL / f_yR	pixels
Horizontal Principal Point	c_xL / c_xR	pixels
Vertical Principal Point	c_yL / c_yR	pixels
1st Order Radial Distortion Coeff.	k_1L / k_1R	-
2nd Order Radial Distortion Coeff.	k_2L / k_2R	-
1st Order Tangential Distortion Coeff.	p_1L / p_1R	-
2nd Order Tangential Distortion Coeff.	p_2L / p_2R	-
3rd Order Radial Distortion Coeff.	k_3L / k_3R	-
Right to left lens rotations	$r2l_{rx}$, $r2l_{ry}$, $r2l_{rz}$	radians
Right to left lens translations	$r2l_{tx}$, $r2l_{ty}$, $r2l_{tz}$	mm

Table 1. The camera parameters for the stereo pinhole model used.

matching as described by Datta et al.[15] to refine the estimated feature location. We found that OpenCV's ring circle detector often returns incorrectly ordered features when using symmetric grid patterns. This was corrected by post processing of the resulting text files. Lastly for the AprilTag grids we used the AprilTag's library TagDetector.

Each of the feature detectors assume to some extent that the imaged features are approximately square/round. In our case the images are squashed vertically due to the acquisition protocol. So the rings are somewhat elliptical and the chessboard is oblong. To test the effect of this squashing, feature extraction was run on the original unscaled images, and on the same images after they had been rescaled vertically to regain the correct aspect ratio. Rescaling was done using OpenCV's resize function, with linear interpolation.

2.4 Calibration

The extracted features were stored in separate text files for each frame of video. Random samples of extracted feature files were then used to perform stereo camera calibration using OpenCV's stereoCalibrate function. This function attempts to determine the intrinsic parameters for each video channel together with the stereo camera geometry. The estimated camera parameters are shown in table 1. For each laparoscope lens we adopt the standard pin hole camera model with 3 radial and 2 tangential distortion coefficients. We parameterise the relative positions of the two lenses as 3 rotations and 3 translations. This gives a total of 26 camera parameters to be calibrated.

2.5 Estimation of Calibration Accuracy due to Feature Detection Performance.

For these calibrations the real optical properties remain unknown. Therefore we use the variance in the estimates of the laparoscope's intrinsic parameters across 1000 repeated calibrations as a proxy measure of calibration error for each method. Similarly we use the variance in the estimated stereo separation as a proxy measure of the error in estimating the camera poses. For each feature detector we repeated calibration using random samples of input feature frames with increasing numbers of frames used, from 2 per channel to 30 per channel. In the case of the AprilTag detector, only frames with greater than 100 features (20 tags) were used for calibration.

The results are reported as plots of mean and standard deviation for each method and each number of frames used. Four intrinsic parameters for the left hand camera are reported, the horizontal focal length, the principal point, and the 1st order radial distortion coefficient. We found that the vertical focal length followed very similar trends to the horizontal focal length, so did not add to the analysis. We also found that the remaining distortion coefficients were very small. The data for the right hand channel intrinsic parameters provided very similar results to that for the left hand channel. We also report the horizontal stereo separation. The remaining stereo parameters were small and did not add significant information to the results. Finally we report the RMS residual projection error (e), in pixels. Whilst the value of the residual error cannot be used as a measure of calibration accuracy, it does depend largely on the feature detection noise, so can be used as a good proxy measure of feature detector performance.

Method	Number of Video Frames	Number of Successful Extractions	Number of Successful Extractions with Rescaled Images
OpenCV Chessboard	14266	5524	8423
OpenCV Rings	14720	4559	6256
OpenCV AprilTags 24	14644	13904	13337
OpenCV AprilTags 216	13880	13400	13151

Table 2. Number of successfully extracted frames for each feature detection method. More chessboards are successfully detected than rings. In each case rescaling the image before feature detection gives a significant increase in detection success. AprilTags provides the most successful detections, however this does not take into account the number of detected tags per frame.
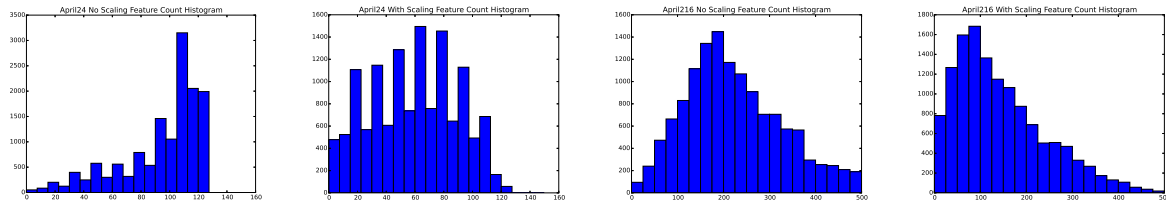

Figure 6. Histograms of the number of detected features (5 per tag) per frame for the AprilTag data sets. As expected there are many more features per frame for the larger grid (the right most two figures). For both the small and large grid, rescaling the image with linear interpolation causes a large drop off in detector performance. The same drop off was observed using linear and cubic interpolation to rescale the images.

To check the calibration performance when the gold standard is known the calibration process was repeated using simulated data. Calibration was performed using a random subset of previously recorded data, the estimated grid position is then used as a perfect gold standard which is projected from 3D to 2D using previously measured calibration parameters. Zero mean, isotropic, independent, normally distributed noise is added to the extracted feature locations and the calibration process repeated with increasing noise standard deviation. The resulting calibration results can then be compared with the results for recorded data.

The algorithms used are implemented in the NiftyCal library, which forms part of the NifTK software package.[19] This software is under active development with a view to making parts of the library more widely available.

3. RESULTS

3.1 Feature Extraction Statistics

Table 2 shows the total number of successfully extracted frames for each method. Rescaling the images prior to feature detection improves performance of the chessboard and ring detector. Figure 6 presents histograms of the number of features detected per frame for the two AprilTag data sets. It is clear that whilst the AprilTag detector works quite well on the unscaled images, rescaling the images has significantly reduced the performance of the AprilTag detector. In these cases we used linear interpolation to rescale the images. We repeated the experiment for the AprilTag detector using cubic interpolation, but saw no significant change in results.

Figure 7 presents histograms showing the pixel location of detected features for each calibration pattern. Figure 8 presents equivalent histograms for the detectors after image scaling. The ring detector appears to have a slightly more uniform coverage with better ability to detect features near the image edge. The large AprilTag grid provides the best coverage.

3.2 Calibration Performance for Standard Size Calibration Grids

Figure 9 compares the calibration performance of the three standard size calibration patterns. No single detector appears to outperform the others across all parameters values. The plot of RMS residual error indicates that the AprilTag detector is significantly noisier than either of the other two, but the difference in detector noise does
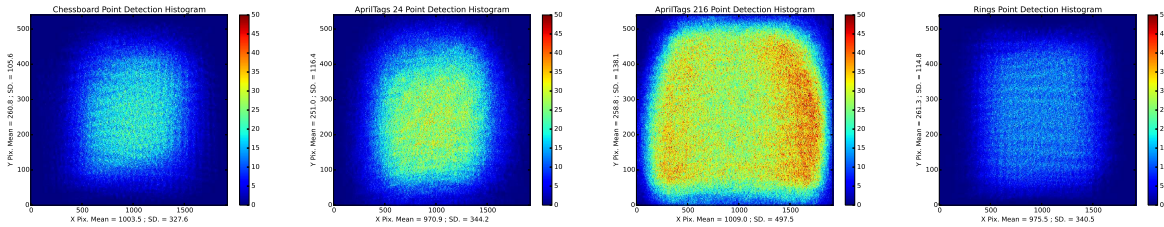
Figure 7. Histograms of the detected point pixel locations for chessboard, AprilTag 24, AprilTag 216 , and ring patterns. The ring pattern appears to provide a more even spread of detected features over the image than the small AprilTag pattern or the chessboard. The large AprilTag pattern provides the best coverage.
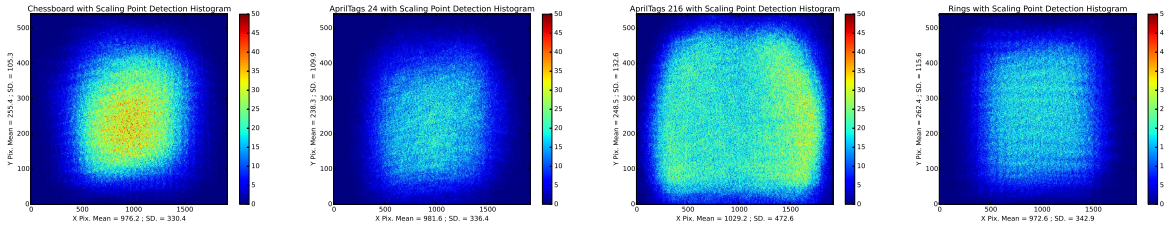

Figure 8. Histograms of the detected point pixel locations for chessboard, AprilTag 24, AprilTag 216 , and ring patterns. Feature detection was performed after rescaling the image. Note that the spread of detected points is increased for both the chessboard and ring pattern, but significantly reduced for the AprilTag patterns.

not necessarily result in a more accurate calibration. Figure 10 presents the same results when image scaling is used before feature detection. Scaling the images does not appear to significantly effect calibration performance.

3.3 Calibration Performance using Template Matching

Figure 11 shows the results of repeated calibrations after feature detection is refined using template matching. With the exception of a change in the estimated stereo separation the template matching appears to have little effect on the overall accuracy. The RMS residual errors indicate that template matching has reduced feature detection error, and avoiding image rescaling provides the best results. However the reduction in feature detection error does not noticeably effect calibration accuracy.

3.4 Calibration Performance using only AprilTag Centres

Figure 12 compares calibration using all the detected AprilTag features (corners and centres) with only using the centre features. In terms of calibration parameters there is no clear benefit to either. There is however a large drop in the RMS residual error. The error is now less than that seen for the ring detector, however this may be due to significantly smaller number of features per frame (≈ 22 for AprilTags vs 140 for rings). Calibration with centres only should yield faster computation due to the shorter feature vectors used.

3.5 Calibration Performance using a Larger Grid

Figure 13 presents calibration results when using the centre points of a larger grid of AprilTags. The standard deviations of the resulting calibrations have significantly increased. To investigate why this was so we looked at histograms of some of the estimated parameter values, see figure 14. What we observed was that calibration parameters for the other detectors were approximately normally distributed over repeated runs. The large grid of AprilTags however resulted in clearly non normally distributed parameter estimates. Typically with a high central peak and long tails. Potential causes of this behaviour are discussed later.
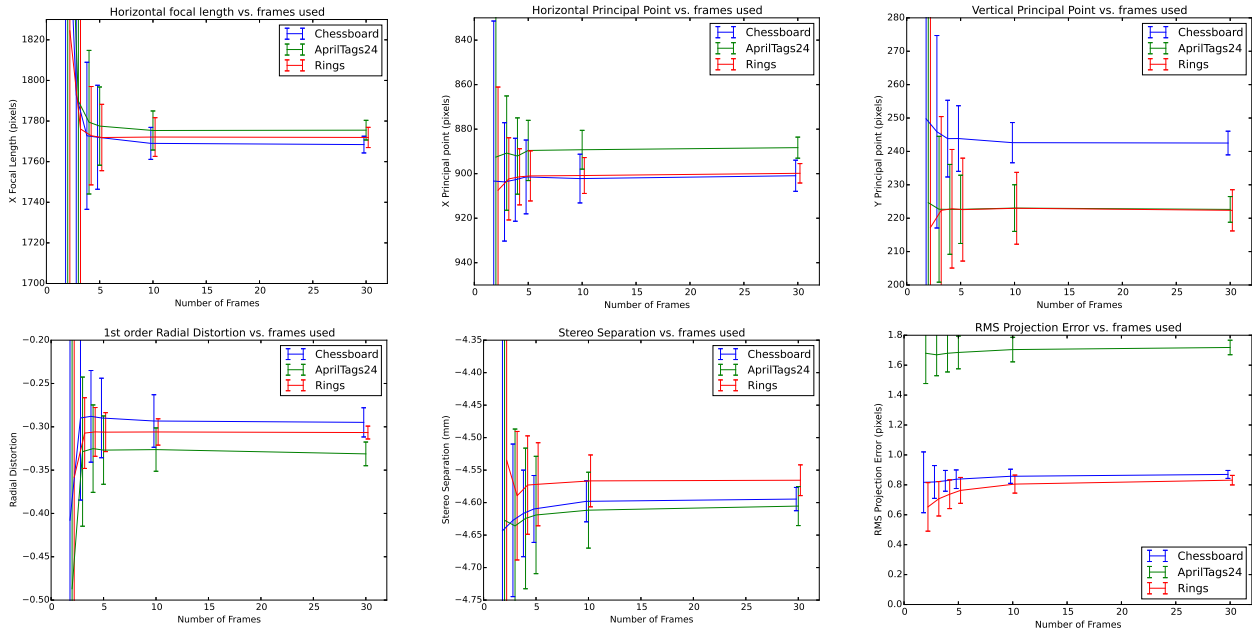
Figure 9. Mean and standard deviation of 5 camera parameters and the residual error for 1000 repeated calibrations using different number of calibration frames and three different feature detectors. Based on the observed standard deviations, the choice of feature detector does not have a significant impact on the accuracy of estimating the focal length, though the ring detector may work better when 5 or fewer frames are used. The AprilTags detector appears to be the best for estimating the principal points, through there is significant disagreement between the mean values. The ring detector appears to give the best estimate of radial distortion. The chessboard gives the best estimate of stereo separation. Looking at the last plot of RMS error, it seems that the AprilTags feature detector is significantly noisier than either of the other two, whilst the ring feature detector is slightly more accurate than the chessboard.

3.6 Calibration Performance using Simulated Data

Figure 15 shows the results of repeated simulation using noisy data. The results suggest that as the standard deviation of the calibration parameters increases with noise, this can be used as a proxy measure of feature detector performance. It can also be seen that the RMS residual projection error provides a good measure of the input feature noise. Interestingly the calibrations fail to converge to the correct parameter values for principal points and distortion. Given the simple noise model used this is surprising. However it is know that the OpenCV stereoCalibrate function can find it difficult to accurately estimate all parameters at once, and it is recommended to perform single channel calibration first.

4. DISCUSSION

From the results it is clear that selection of features to use for calibration is not as straightforward as we had hoped. In general, given the same grid geometry, there is little to choose between the three feature detectors once the results have been passed through a full stereo calibration. The number of frames used for calibration has a much more significant impact on calibration performance than choice of feature detection. There is some evidence that in terms of detection coverage and calibration accuracy that the ring detector performs better than the others, but it is marginal.

Rescaling the images prior to feature detection improved feature detection rate for both the chessboard and rings detectors at a small cost in accuracy. However this loss of accuracy did not effect calibration results. It appears that image rescaling with the AprilTag detector is best avoided at present.

The most promising solution seems to be to use a larger grid of uniquely identifiable patterns, but this requires some refinement of the detector algorithm. One possible cause of the large number of outliers may be
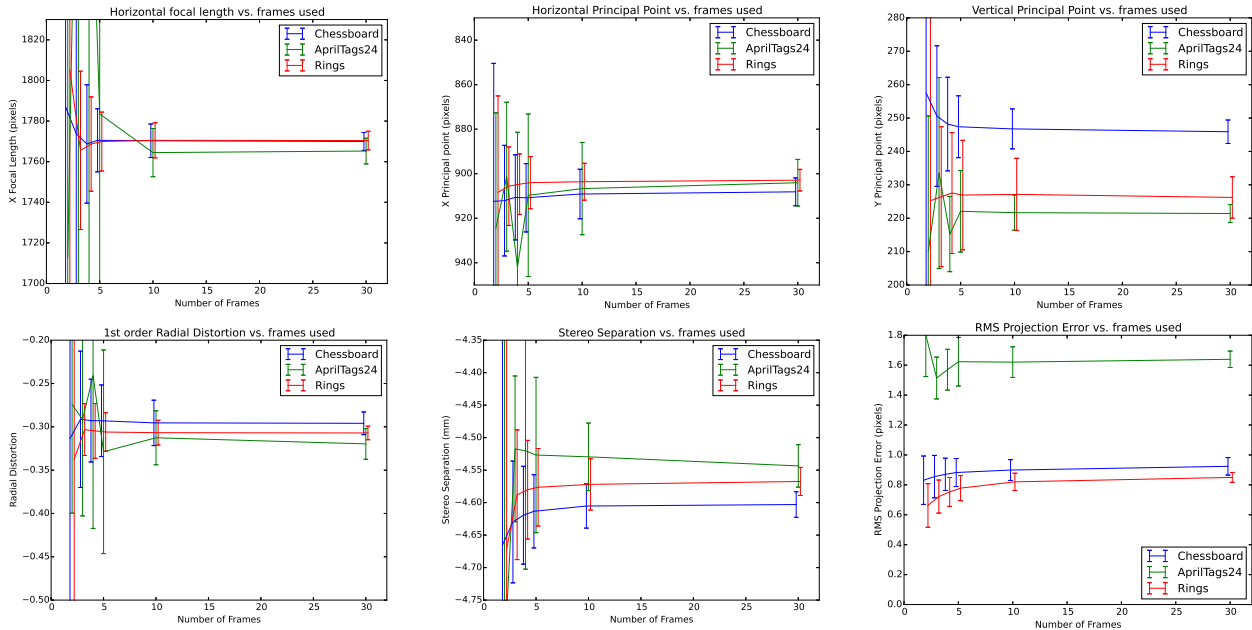
Figure 10. Mean and standard deviation of 5 camera parameters and the residual error for 1000 repeated calibrations using different number of calibration frames and three different feature detectors. Prior to feature detection each image was rescaled using OpenCV's resize function with linear interpolation. None of the results appear significantly different to for feature detection without rescaling. The most noticeable change is that in the vertical principal point estimated using the ring detector.
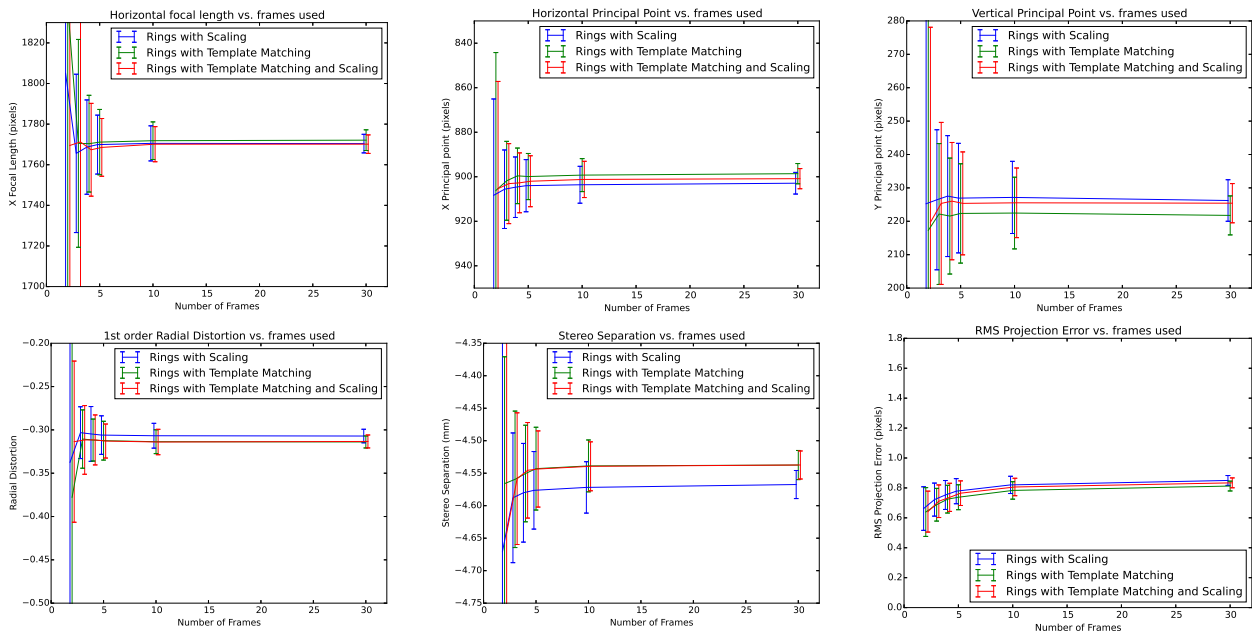

Figure 11. Mean and standard deviation of 5 camera parameters and the residual error for 1000 repeated calibrations using different number of calibration frames and the ring feature detection, with and without the additional step of template matching. These results suggest that use of template matching to refine feature detection has not had a significant impact on calibration outcomes for this data. The exception to this in the estimation of stereo separation, where the use of template matching reduces the estimate stereo separation by around 0.05 mm. There is a slight reduction in RMS residual error when using template matching, suggesting that it does reduce feature detection error.
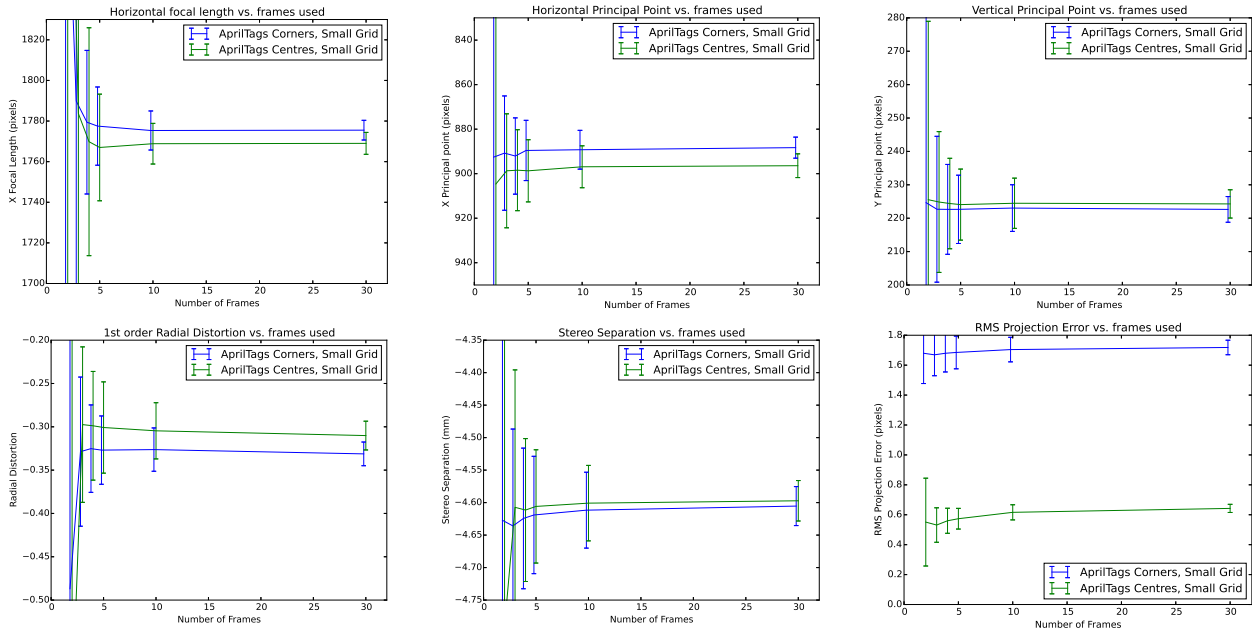
Figure 12. Mean and standard deviation of 5 camera parameters and the residual error for 1000 repeated calibrations using different number of calibration frames and AprilTags for feature detection. Results are only shown for the small grid, though the trends were similar for the large grid. Two detection methods were used, in the first all the points found by the AprilTags detector were used, i.e. 5 points per tag, the 4 corners and the centre. In second method only the detected centre were used, in theory reducing detector noise. The results indicate that only using the centres does not significantly reduce the noise of the resulting parameter estimation. The small changes in mean parameter values suggest that there may be biases introduced in the detector. Looking at the RMS residual projection errors it is clear that only using the centres gives a marked reduction in feature detection noise.

the presence of misidentified tags in the extracted feature points. We saw evidence of misidentification when running both the small and large grid AprilTag detector, as the stereo calibration throws an exception when points are found in the image data that have no corresponding point in the reference set. In these cases the detected tag will be ignored and not effect calibration. If however the misidentified tag has a corresponding tag in the reference set, it will lead to incorrect calibration. The probability of a misidentified tag having a corresponding tag in the reference set is directly related to the number of tags in the reference set. The AprilTag 25h7 tag family has 242 unique tags. For the calibration grid with 24 tags there is a $\frac{24}{242} \approx 10\%$ chance that a misidentified tag will have a corresponding tag in the reference set. For the large calibration grid there is a $\frac{216}{242} \approx 90\%$ chance that a misidentified tag will have a corresponding tag in the reference data set. This would explain why this effect is only seen in the large grid pattern.

Work is ongoing to check for tag misidentification and confirm whether this is the cause of the outliers for the large grid of AprilTags. If this can be corrected then it seems that the use of a larger grid has the potential to improve parameter estimation. The detector used did not have any knowledge of the expected grid order of tags, it is likely that implementing this to detect and remove incorrectly identified tags might produce an improved calibration result.

We observed in the result section that different parameters were estimated better with different feature types. This suggests that the requirements for a calibration pattern to estimate the camera's intrinsic parameters are different to one that would best estimate the camera pose, and hence the stereo geometry. It seems likely that the accuracy of pose estimation should increase as the camera moves closer to the calibration pattern. This can only be done practically by using patterns with identifiable tags. As we excluded images with less than 20 tags visible from our data, we probably won't have observed that here. It would be interesting to repeat this experiment with the camera closer to the AprilTag grid.
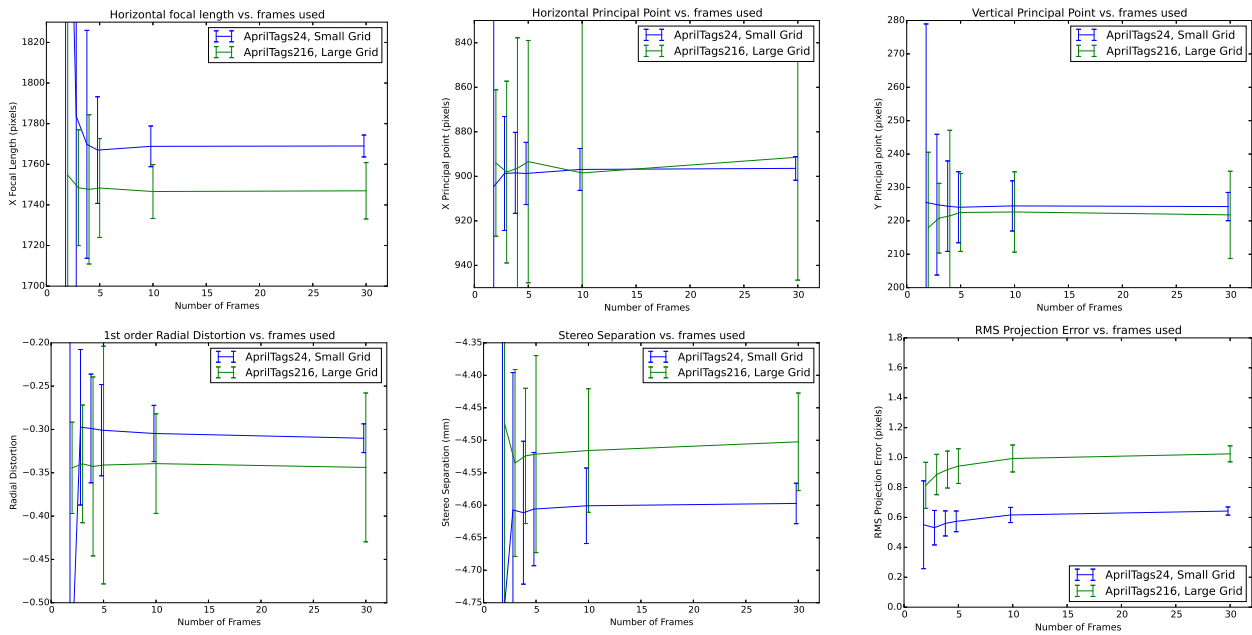
Figure 13. Mean and standard deviation of 5 camera parameters and the residual error for 1000 repeated calibrations using different number of calibration frames and AprilTags for feature detection. Two grid sizes were used, one with 24 tags to keep similar geometry to the chessboard, and one with 216 tags, allowing more complete sampling across the image. Use of the large grid has led to a large increase in the observed standard deviations, which we believe is due to outliers caused by misidentified tags. Interestingly, the larger grid leads to a significant reduction in the estimated focal lengths, and a consequent reduction in the estimated stereo separation.
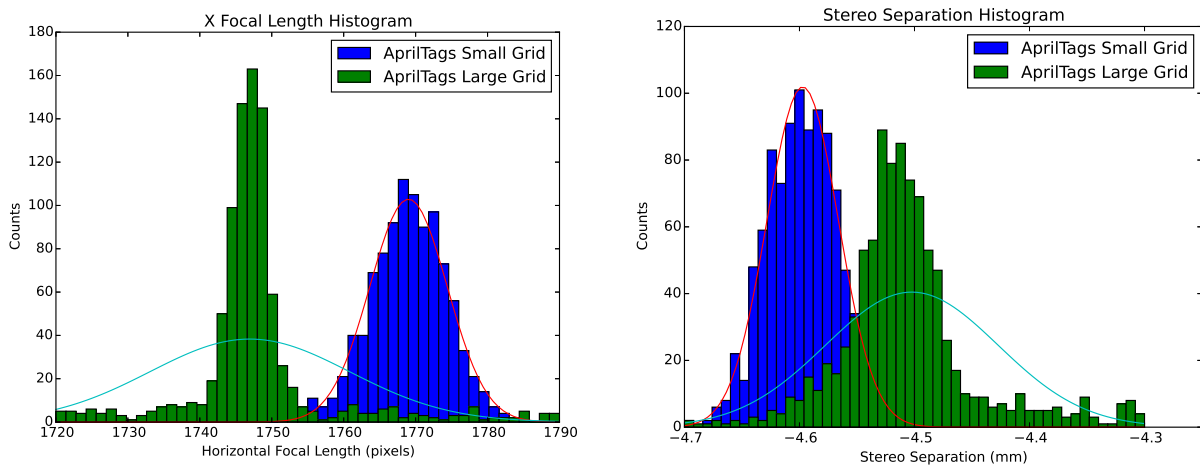

Figure 14. Histograms of the calibrated parameter values for a large and small AprilTag grid. Most of the parameters estimated using the chessboard, ring, and small grid AprilTag detectors are approximately normally distributed (normal curves overlaid). In the case of the large grid AprilTag pattern however there is a higher central peak and long tails. One explanation for this is that th larger number of AprilTags, and their location near the more poorly lit and distorted image edges, leads to an increase in the number of misidentified AprilTags.
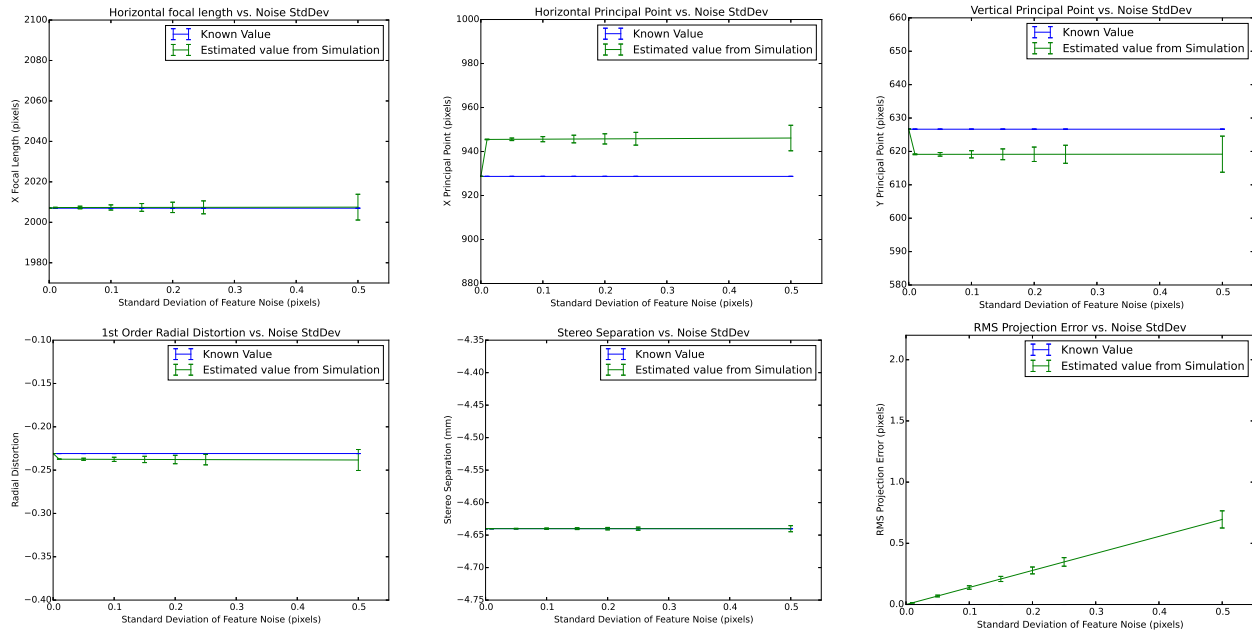
Figure 15. Results for simulated calibration with normally distributed, isotropic, independent noise added to the extracted pixel locations. In all cases the spread of parameter value estimation increases in line with noise. Interestingly the mean values for principal points and distortion do not match the ground truth value. Note that the absolute values of the calibration parameters is slightly different to those reported for the recorded data, as the data used for simulation was recorded with a different Viking stereo laparoscope.

The simulated results show that the RMS residual projection error can be used as a good estimator of the input feature detection noise. Furthermore, whilst the value of the RMS residual error for a single calibration cannot be used as a predictor of the accuracy of that calibration, it is is correlated with the the standard deviation of parameter estimates over repeated calibrations. Thus when a high RMS residual error is encountered, it might be beneficial to use a large number of calibration frames, and/or repeat the calibration multiple times.

Whilst we focused on the change of standard deviation of estimated parameter values, we also saw many cases where the choice of detector had an impact on the mean value of parameter estimates. This could be explained by small errors in the printing of the calibration grids leading to an incorrect reference geometry. However as was also observed this effect in the simulated data it seems likely that the choice of calibration optimiser is not ideal. There are many examples of improved calibration optimisers in the literature and it would be instructive to apply some of these to this data set.

We are currently working on extending this work to examine the best way to perform hand-eye calibration for optically and electromagnetically tracked laparoscopes. The intent with this paper was to provide some estimate of the errors due to feature detection noise, that when coupled with past work on tracking errors,[20, 21] will enable us to accurately model the calibration process.

5. CONCLUSION

The results we have shown do indicate that some feature detectors are more accurate in this application than others. The most accurate feature detector appears to be rings, with template matching refinement, on an unscaled image. However once the extracted features have been used for stereo calibration the benefits of improved feature detection are negligible. We hope that the results presented here provide a useful resource for those working on calibration of laparoscopes. We will happily share the data used on request.

ACKNOWLEDGMENTS

REFERENCES

1. S. Thompson, J. Totz, Y. Song, S. Johnsen, D. Stoyanov, S. Ourselin, K. Gurusamy, C. Schneider, B. Davidson, D. Hawkes, and M. J. Clarkson, "Accuracy validation of an image guided laparoscopy system for liver resection," *Proc. SPIE* **9415**, pp. 941509–941509–12, 2015.

2. X. Liu, S. Kang, E. Wilson, C. Peters, T. Kane, and R. Shekhar, "Evaluation of electromagnetic tracking for stereoscopic augmented reality laparoscopic visualization," in *Clinical Image-Based Procedures. Translational Research in Medical Imaging*, M. G. Linguraru, C. Oyarzun Laura, R. Shekhar, S. Wesarg, M. . Gonzlez Ballester, K. Drechsler, Y. Sato, and M. Erdt, eds., *Lecture Notes in Computer Science* **8680**, pp. 84–91, Springer International Publishing, 2014.

3. R. Shahidi, M. Bax, J. Maurer, C.R., J. Johnson, E. Wilkinson, B. Wang, J. West, M. Citardi, K. Manwaring, and R. Khadem, "Implementation, calibration and accuracy testing of an image-enhanced endoscopy system," *Medical Imaging, IEEE Transactions on* **21**, pp. 1524–1535, Dec 2002.

4. S. Thompson, G. Penney, M. Billia, B. Challacombe, D. Hawkes, and P. Dasgupta, "Design and evaluation of an image-guidance system for robot-assisted radical prostatectomy," *BJU International* **111**(7), pp. 1081–1090, 2013.

5. J. Bano, S. Nicolau, A. Hostettler, C. Doignon, J. Marescaux, and L. Soler, "Registration of preoperative liver model for laparoscopic surgery from intraoperative 3d acquisition," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*, H. Liao, C. Linte, K. Masamune, T. Peters, and G. Zheng, eds., *Lecture Notes in Computer Science* **8090**, pp. 201–210, Springer Berlin Heidelberg, 2013.

6. M. Baumhauer, T. Simpfendrfer, B. Mller-Stich, D. Teber, C. Gutt, J. Rassweiler, H.-P. Meinzer, and I. Wolf, "Soft tissue navigation for laparoscopic partial nephrectomy," *International Journal of Computer Assisted Radiology and Surgery* **3**, pp. 307–314, 2008.

7. N. C. Buchs, F. Volonte, F. Pugin, C. Toso, M. Fusaglia, K. Gavaghan, P. E. Majno, M. Peterhans, S. Weber, and P. Morel, "Augmented environments for the targeting of hepatic lesions during image-guided robotic liver surgery," *Journal of Surgical Research* **184**(2), pp. 825 – 831, 2013.

8. S. Thompson, D. Stoyanov, C. Schneider, K. Gurusamy, S. Ourselin, B. Davidson, D. Hawkes, and M. J. Clarkson, "Hand–eye calibration for rigid laparoscopes using an invariant point," *International Journal of Computer Assisted Radiology and Surgery* **11**(6), pp. 1071–1080, 2016.

9. M. Shah, R. D. Eastman, and T. Hong, "An overview of robot-sensor calibration methods for evaluation of perception systems," in *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*, *PerMIS '12*, pp. 15–20, ACM, (New York, NY, USA), 2012.

10. R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *Robotics and Automation, IEEE Transactions on* **5**, pp. 345–358, Jun 1989.

11. A. Malti and J. P. Barreto, "Handeye and radial distortion calibration for rigid endoscopes," *The International Journal of Medical Robotics and Computer Assisted Surgery* **9**(4), pp. 441–454, 2013.

12. J. Heller, M. Havlena, and T. Pajdla, "Globally optimal hand-eye calibration using branch-and-bound," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38**, pp. 1027–1033, May 2016.

13. C. Wengert, M. Reeff, P. C. Cattin, and G. Szkely, "Fully automatic endoscope calibration for intraoperative use," in *Bildverarbeitung fr die Medizin*, pp. 419–23, Springer-Verlag, March 2006.

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

15. A. Datta, J.-S. Kim, and T. Kanade, "Accurate camera calibration using iterative refinement of control points," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pp. 1201–1208, Sept 2009.

16. X. Liu, W. Plishker, G. Zaki, S. Kang, T. D. Kane, and R. Shekhar, "On-demand calibration and evaluation for electromagnetically tracked laparoscope in augmented reality visualization," *International Journal of Computer Assisted Radiology and Surgery* **11**(6), pp. 1163–1171, 2016.

17. A. Richardson, J. Strom, and E. Olson, "Aprilcal: Assisted and repeatable camera calibration," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pp. 1814–1821, Nov 2013.

18. E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3400–3407, IEEE, May 2011.

19. M. Clarkson, G. Zombori, S. Thompson, J. Totz, Y. Song, M. Espak, S. Johnsen, D. Hawkes, and S. Ourselin, "The NifTK software platform for image-guided interventions: platform overview and NiftyLink messaging," *International Journal of Computer Assisted Radiology and Surgery* **10**(3), pp. 301–316, 2015.

20. S. Thompson, G. Penney, P. Dasgupta, and D. Hawkes, "Improved modelling of tool tracking errors by modelling dependent marker errors," *IEEE Transactions on Medical Imaging* **32**, pp. 165–177, Feb 2013.

21. E. Lugez, H. Sadjadi, D. R. Pichora, R. E. Ellis, S. G. Akl, and G. Fichtinger, "Electromagnetic tracking in surgical and interventional environments: usability study," *International Journal of Computer Assisted Radiology and Surgery* **10**(3), pp. 253–262, 2015.