

Laparoscope Self-calibration for Robotic Assisted Minimally Invasive Surgery

Danail Stoyanov¹, Ara Darzi², and Guang-Zhong Yang^{1,2}

¹ Royal Society/Wolfson Foundation Medical Image Computing Laboratory

² Department of Surgical Oncology and Technology

Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK

{danail.stoyanov, a.darzi, g.z.yang}@imperial.ac.uk

<http://vip.doc.ic.ac.uk>

Abstract. For robotic assisted minimal access surgery, recovering 3D soft tissue deformation is important for intra-operative surgical guidance, motion compensation, and prescribing active constraints. We propose in this paper a method for determining varying focal lengths of stereo laparoscope cameras during robotic surgery. Laparoscopic images typically feature dynamic scenes of soft-tissue deformation and self-calibration is difficult with existing approaches due to the lack of rigid temporal constraints. The proposed method is based on the direct derivation of the focal lengths from the fundamental matrix of the stereo cameras with known extrinsic parameters. This solves a restricted self-calibration problem, and the introduction of the additional constraints improves the inherent accuracy of the algorithm. The practical value of the method is demonstrated with analysis of results from both synthetic and *in vivo* data sets.

1 Introduction

With the maturity of master-slave robotic manipulators, the clinical applications of minimally invasive surgery (MIS) are rapidly advancing. Robotic assistance provides enhanced instrumental control and intuitive 3D manipulation of the operating field, and advanced systems can include image guidance, motion compensation and active constraints. For MIS involving large soft tissue deformation, intraoperative guidance with patient specific anatomy presents a significant challenge and reliable 3D reconstruction of soft-tissue deformation *in situ* is essential [1]. The prerequisite of this for optical methods, however, is accurate and robust camera calibration. Existing research has shown that after calibration and feature correspondence, it is possible to perform real-time 3D reconstruction of soft tissue deformation with stereo laparoscope cameras [2].

Thus far, preoperative calibration for robotic assisted MIS is generally achieved with a calibration object [3] based on the assumption that during the operation the intrinsic parameters of the laparoscope remain fixed. This, however, is not true for complex procedures as for example in thoracoscopic surgery, where parameters such as the focal length may change in order to optimize the surgical field-of-view. In such cases, frequent recalibration of the laparoscope is not feasible and *in situ* self-calibration is the only practical way forward. Hitherto, self-calibration has received

extensive interest in computer vision and most existing techniques assume a rigid world, which can be viewed from a number of different positions [4]. For continuously deforming scenes, however, such temporal constraints are invalid and only inter-stereo epipolar geometry can be used to derive camera parameters. Hartley [5] has shown that the focal lengths of the stereo camera can be extracted from the fundamental matrix by using Singular Value Decomposition (SVD) and a number of other techniques based on the same constraint have also been developed [6,7]. Brooks *et al.* [8] demonstrated that in a typical stereo configuration, the problem is degenerate for cameras with different focal lengths [9]. Strum [10] subsequently provided a detailed analysis for deriving matching focal lengths of the cameras based on epipolar geometry, and outlined the potential singularities when variable focal lengths were used [11].

It has been recognized that the above methods are susceptible to noise in the proximity of singularities. Frahm *et al* [12] proposed a method based on known rotations between views and using one epipole for embedding the unknown translations. Earlier work by Stein [13] and McLaughlin *et al* [14] also used partial knowledge of rotation to restrict the problem. These techniques, however, are not specifically designed for stereo camera setups and can result in complex formulations with additional degenerate configurations. Whilst for many vision systems critical motion can be avoided through appropriate setups, this is not the case for stereo laparoscopes where the camera arrangement is on a much smaller scale to assist the 3D perception of the surgeon. The purpose of this paper is to develop a parameterization scheme for the intrinsic parameters of the stereo laparoscope based on the epipolar constraint with known intrinsic parameters. This simplifies the traditional self-calibration to an over determined problem with respect to focal lengths. We demonstrate that the problem is solvable with a minimum of two correspondences and the stability of the algorithm to different noise levels can be further improved by using a robust estimator. The practical value of the technique is demonstrated with both numerical simulation and *in vivo* robotic assisted MIS data.

2 Methods

2.1 Stereoscopic Laparoscope Model

It is assumed in this study that the stereo laparoscope cameras follow the standard pinhole model, with which the mapping of a point $\mathbf{M} = [X \ Y \ Z \ W]^T$ in projective 3D space onto the corresponding image point $\mathbf{m} = [x \ y \ w]^T$ can be described up to a scale by a matrix multiplication with homogeneous coordinates:

$$\mathbf{m} \sim \mathbf{P}\mathbf{M} \quad (1)$$

where \sim denotes equality up to scale. The matrix \mathbf{P} is the camera's projection matrix and may be decomposed into the intrinsic and extrinsic camera parameters:

$$\mathbf{P} = \mathbf{K}[\mathbf{R} | -\mathbf{R}\mathbf{C}] \quad (2)$$

The camera orientation and position with respect to a reference coordinate system are expressed by a rotation matrix \mathbf{R} and a translation vector $\mathbf{t} = -\mathbf{R}\mathbf{C}$, where \mathbf{C} is the

location of the optical center. The upper triangular matrix \mathbf{K} encompasses the focal length, f , skew, s , aspect ratio, μ , and the image coordinates of the principal point, $[u, v]$:

$$\mathbf{K} = \begin{bmatrix} f & s & u \\ & \mu f & v \\ & & 1 \end{bmatrix} \tag{3}$$

Without a loss of generality, the camera matrices for the stereoscopic laparoscope can be represented with the following equation by taking the left camera as the reference coordinate system:

$$\mathbf{P} = \mathbf{K}[\mathbf{I} | \mathbf{0}] \text{ and } \mathbf{P}' = \mathbf{K}'[\mathbf{R} | \mathbf{t}] \tag{4}$$

The inherent relationship between the left and right camera frames is described by the epipolar geometry, which is algebraically encapsulated by the fundamental matrix, \mathbf{F} [15]. The fundamental matrix is a rank two matrix, defined up to scale and thus it has seven degrees of freedom. Its dependency on the camera parameters can be expressed as:

$$\mathbf{F} \sim \mathbf{K}'^{-T} \left(\begin{bmatrix} \mathbf{t} \\ \times \end{bmatrix} \mathbf{R} \right) \mathbf{K}^{-1} \sim \mathbf{K}'^{-T} \mathbf{E} \mathbf{K}^{-1} \tag{5}$$

where $[\cdot]_{\times}$ is used to denote a skew symmetric matrix and \mathbf{E} is the essential matrix defined by the rotation and translation between the cameras [15]. For corresponding image points (points which are projections of the same 3D world point), \mathbf{F} can be used to express the constraint:

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0 \tag{6}$$

Given sufficient corresponding points (eight or more for a unique solution) the fundamental matrix can be determined from Eq. (6) and this is the basic knowledge about the camera geometry, which can be directly obtained from the images.

2.2 Self-calibration with Varying Focal Lengths

Without prior knowledge about the cameras, Eq. (5) is governed by 15 unknowns, and the seven degrees of freedom of the fundamental matrix are insufficient to provide a solution for the camera parameters from a single pair of stereo images. In robotically assisted MIS, however, it is feasible to calibrate the stereoscopic laparoscope before the procedure. Therefore, the knowledge about the initial extrinsic and intrinsic parameters of the cameras can be used to simplify Eq. (5) for deriving a solution for the varying focal length problem. By assuming all parameters except for the focal lengths are known, Eq. (5) can be written as:

$$\tilde{\mathbf{F}} \sim \text{diag}(1, 1, f') \mathbf{E} \text{diag}(1, 1, f) \sim \begin{bmatrix} e_{11} & e_{12} & e_{13}f \\ e_{21} & e_{21} & e_{13}f \\ e_{31}f' & e_{31}f' & e_{13}ff' \end{bmatrix} \tag{7}$$

For computing $\tilde{\mathbf{F}}$ based on Eq. (6), the coordinates of the corresponding image points are normalized by the known intrinsic parameters to $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{m}}'$ [15]. Since the knowledge of the essential matrix is also available from the initial calibration, Eq. (7) yields a bilinear equation with respect to the focal length for each point correspondence:

$$\begin{aligned} (xy'e_{21} + xw'e_{31})f + (wx'e_{13} + wy'e_{23})f' + ww'e_{33}ff' = \\ xx'e_{11} + xy'e_{21} + yx'e_{12} + yy'e_{22} \end{aligned} \quad (8)$$

It is evident that the above equation can be solved from a minimum of two correspondences. But in practice, the problem is likely to be over determined and may be solved in a least-squares sense. The minimum solution required is important, however, to robust estimation schemes such as the random sampling consensus (RANSAC) approach [15], which outperform linear methods in the presence of noise. We have therefore used an approach analogous to the RANSAC estimation of the fundamental matrix using the minimum solution obtained from Eq. (8). With the proposed technique, inliers are determined using the first-order approximation to the geometric re-projection error:

$$\sum_i \frac{(\tilde{\mathbf{m}}_i^\top \tilde{\mathbf{F}} \tilde{\mathbf{m}}_i)^2}{(\tilde{\mathbf{F}} \tilde{\mathbf{m}}_i)_1^2 + (\tilde{\mathbf{F}} \tilde{\mathbf{m}}_i)_2^2 + (\tilde{\mathbf{F}}^\top \tilde{\mathbf{m}}'_i)_1^2 + (\tilde{\mathbf{F}}^\top \tilde{\mathbf{m}}'_i)_2^2} \quad (9)$$

The number of samples used in the robust estimation is determined adaptively and once the desired solution is determined, Eq. (9) is used as a cost function to refine the estimated focal lengths by using the Levenberg-Marquardt algorithm. With the proposed framework, incremental changes of the known extrinsic parameters, as well as shifts of the principal points, can also be incorporated in the minimization procedure.

It can be shown that the practical critical motion (extrinsic configuration of the stereo cameras for which the solution is degenerate) for the proposed method is when the optical axis are parallel. This configuration is a generic singularity for most existing methods, which results in a reduced version of the fundamental matrix so that Eq. (8) cannot be solved. In practice, however, the stereo laparoscope is always setup with a small vergence angle to assist stereoscopic fusion of the surgeon with the left and right visual channels. Although this naturally avoids the singularity problem, it is important to note that the cameras can be in proximity of the degeneracy.

2.3 Numerical Validation and *In vivo* Experiment

To assess the general performance of the proposed method, particularly in the vicinity of the singular configuration, a synthetic test data set was used. In this experiment, the left and right cameras started from a degenerate position with parallel image planes with optical axes as shown in **Fig. 1**. We varied the vergence angle, θ , from zero with increasing steps up to ten degrees. In this configuration, the problem can be unstable [8] as the e_{33} element of the essential matrix is zero, and therefore there is a single

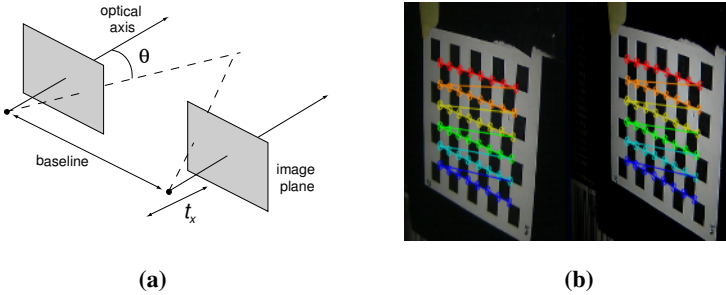


Fig. 1. (a) A schematic illustration of the setup used for synthetic simulation. (b) An example stereo view of the calibration grid used for deriving the ground truth data.

linear solution for Eq. (8) when only two correspondences are available. For numerical validation, a total of 100 random 3D points in the viewing volume in front of the cameras were used, and the projection of each point in the image planes was corrupted by additive zero mean Gaussian noise with standard deviation varied from zero to two pixels. The effect of error with known extrinsic parameters was analyzed also by corrupting the known vergence angle and baseline with zero mean Gaussian noise with standard deviation varying from zero to two in units of degrees and millimeters, respectively. We performed a total of 100 trials for each noise level and the mean estimation was used for final analysis.

For *in vivo* validation of the proposed technique, a surgical procedure with the daVinci™ surgical robot was used. The ground truth of the calibration data was obtained by using a planar grid method [17] before the insertion of the laparoscope into the patient. The proposed algorithm was used to intraoperatively compute the focal length of the stereo system. Image correspondences were obtained by using a variant of the stereo feature matching algorithm described by Pilu [18]. With this experiment, the focal lengths of the cameras were altered during the procedure in order to validate the proposed method. The ground truth parameters were recalculated immediately after the change of parameters by removing the stereo laparoscope out of the patient for recalibration.

3 Results

In **Fig. 2**, we demonstrate the overall performance of the proposed method as the cameras rotate about their own vertical axis forming a vergence angle. **Fig. 2 (a)** shows the results of solving Eq. (8) by using a least-squares-method (LSM), whereas **Fig. 2 (b)** shows the corresponding results by using the robust algorithm. It is evident that in the presence of noise, the robust algorithm clearly out-performs LSM and it retains a good accuracy in the vicinity of degeneracy.

In **Fig. 3**, the effects of noise in the known extrinsic parameters are analyzed for the proposed calibration method by varying stereo vergence angle. It is evident that the algorithm performed well in the presence of significant errors in the baseline and rotation angles. Similarly to **Fig. 2**, the algorithm is also relatively robust in the neighborhood of the singular camera configuration.

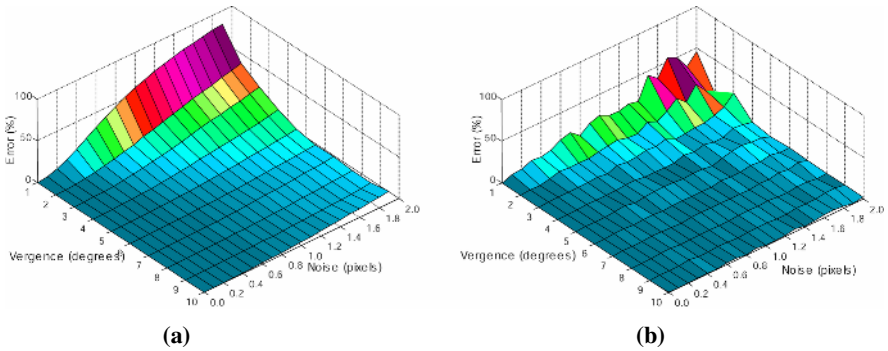


Fig. 2. Average percentage errors in focal length estimation by varying vergence angle near the singularity point with noise corrupted synthetic data. **(a)** Results with least-squares solution using SVD, and **(b)** results with the proposed robust algorithm.

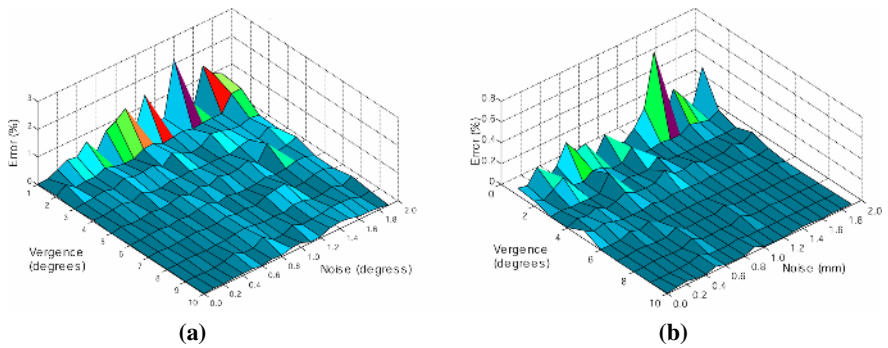


Fig. 3. Percentage errors in focal length estimation w.r.t. noise in the known extrinsic parameters of **(a)** vergence angle, **(b)** baseline

The results obtained from *in vivo* experiment are listed in **Fig. 4**. For the calculation of the ground truth calibration data, the intrinsic accuracy of the method had a pixel reprojection error of 0.12 pixels, and a reconstruction error of 0.78 mm. **Fig. 4 (b)** demonstrates the estimated focal lengths over the acquired video sequence, where the solid lines correspond to the ground truth data. It is clear from **Fig. 4** that after the change in camera focal length, the algorithm is able to reliably quantify the shift involved. For providing a detailed statistical analysis of data, **Fig. 5** illustrates the standard deviation of focal length estimation for the left laparoscope camera before and after the change in focal length. The shaded region outlines the standard deviation of the derived focal length value in every 0.4s window of the video sequence. The variance in the estimation is dominated by the accuracy of feature matching in the stereo pair, which can vary for images with many specular highlights or significant motion blurring from caused instrument or tissue movement.

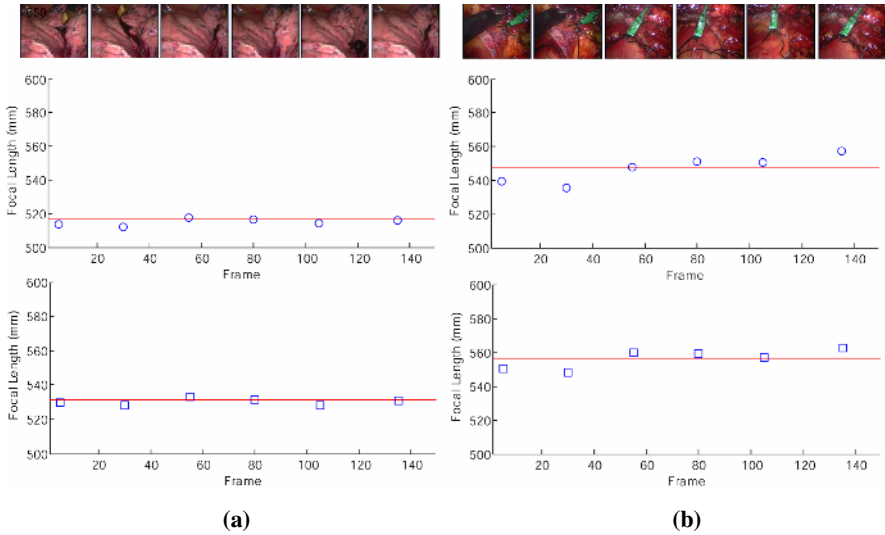


Fig. 4. *In situ* calibration results for a robotic assisted MIS video where the ground truth is shown as a solid red line. Columns (a) and (b) show the changes in focal length during the operation. The circular and square markers represent the estimations of the focal lengths of the left and right cameras, respectively.

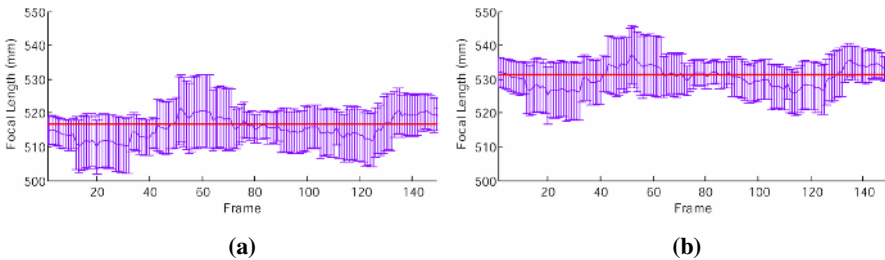


Fig. 5. The standard deviation of focal length estimation for the left stereo laparoscope camera before (a) and after (b) focal length change (the estimation window was set to be 0.4s)

4 Discussion and Conclusions

In this paper, we have proposed a practical method for intra-operatively determining the varying focal length of stereo laparoscopes. The validity of the method has been demonstrated on both synthetic and *in vivo* data. The results indicate the high accuracy obtained despite the near singular arrangement of the cameras. The performance of the algorithm against noise of the known extrinsic parameters, suggests that it may be possible to fully calibrate the system in different surgical procedures by using a prior estimate of the camera parameters. Thus far, approaches to self-calibration typically involve temporal constraints over rigid multi-view geometry. Due to extensive deformation of the soft tissue, constraints can only be enforced across two-view inter-stereo epipolar geometry for robotic MIS procedures. The proposed

method represents a first step towards active calibration of changing camera parameters during surgery, and the results derived have shown the robustness of the technique in proximity of the generic degeneracy and against noise influence.

References

1. Taylor, R. H., Stoianovici, D.: Medical Robotics in Computer-Integrated Surgery. *IEEE Transactions on Robotics and Automation*, (19):765-781, 2003.
2. Stoyanov, D., Darzi, A., Yang, G.-Z.: Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. In: *Proc. MICCAI*, 41-48, 2004.
3. Morgues, F., Coste-Manière, È.: Flexible calibration of actuated stereoscopic endoscope for overlay in robot assisted surgery. In: *Proc. MICCAI*, 2488, 24-34, 2002.
4. Pollefeys, M.: Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences. PhD Thesis, Katholieke Universiteit Leuven, Belgium, 1999.
5. Hartley, R. I.: Estimation of Relative Camera Positions for Uncalibrated Cameras. In: *Proc. ECCV*, 579-587, 1992.
6. Bougnoux, S.: From Projective to Euclidean Space under Any Practical Situation, A Criticism of Self-Calibration. In: *Proc. ICCV*, 790-796, 1998.
7. Kanatani, K., Matsunaga, C.: Closed Form Expressions for Focal Lengths from the Fundamental Matrix. In: *Proc. ACCV*, 128-133, 2000.
8. Brooks, M.J., De Agapito, L., Huynh, D.Q., Baumela, L.: Towards Robust Metric Reconstruction Via A Dynamic Uncalibrated Stereo Head. *Image and Vision Computing*, (16):989-1002, 1998.
9. De Agapito, L., Huynh, D. Q., Brooks, M. J.: Self Calibrating a Stereo Head: An Error Analysis in The Neighbourhood of Degenerate Configurations. In: *Proc. ICCV*, 747-753, 1998
10. Sturm, P.: On Focal Length Calibration from Two Views. In: *Proc. CVPR*, 145-150, 2000.
11. Sturm, P.: Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length. *Image and Vision Computing*, (20):415-426, 2002.
12. Frahm, F., Koch, R.: Camera Calibration with known Rotation. In: *Proc. ICCV*, 1418-1425, 2003.
13. Stein, F.: Accurate Internal Camera Calibration Using Rotation with Analysis of Sources of Error. In: *Proc. ICCV*, 230 - 236, 1995.
14. McLauchlan, P. F., Murray, D. W.: Active camera calibration for a head-eye platform using the variable state-dimension filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (18):15-22, 1996.
15. Hartley, R., Zisserman, A.: *Multiple view geometry in computer vision*. Cambridge University Press. 2000.
16. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (22):1330-1334, 2000.
17. Pilu, M.: A Direct Method for Stereo Correspondence based on Singular Value Decomposition. In: *Proc. CVPR*, 261-266, 1997.