

Hand-eye calibration for robotic assisted minimally invasive surgery without a calibration object

Krittin Pachtrachai, Max Allan, Vijay Pawar, Stephen Hailes and Danail Stoyanov

Abstract—In a robot mounted camera arrangement, hand-eye calibration estimates the rigid relationship between the robot and camera coordinate frames. Most hand-eye calibration techniques use a calibration object to estimate the relative transformation of the camera in several views of the calibration object and link these to the forward kinematics of the robot to compute the hand-eye transformation. Such approaches achieve good accuracy for general use but for applications such as robotic assisted minimally invasive surgery, acquiring a calibration sequence multiple times during a procedure is not practical. In this paper, we present a new approach to tackle the problem by using the robotic surgical instruments as the calibration object with well known geometry from CAD models used for manufacturing. Our approach removes the requirement of a custom sterile calibration object to be used in the operating room and it simplifies the process of acquiring calibration data when the laparoscope is constrained to move around a remote centre of motion. This is the first demonstration of the feasibility to perform hand-eye calibration using components of the robotic system itself and we show promising validation results on synthetic data as well as data acquired with the da Vinci Research Kit.

I. INTRODUCTION

Minimally invasive surgery (MIS) has significantly altered surgical practice by introducing specialised tools to access the inner anatomy with minimal trauma. Robotic systems have emerged as a powerful solution to some of the challenges in operating such specialised instrumentation by providing tele-operation of the surgical camera and tools. Robotic MIS (RMIS) systems like the da Vinci[®] (Intuitive Surgical, CA) enable surgery to be carried by the surgeon controlling articulated miniaturised instruments with dexterous master manipulators and visualizing the anatomy with stereo-endoscopic cameras inserted through keyhole ports [1], [2], [3]. The tele-operation setup additionally provides an exciting platform to realise many techniques for computer assisted interventions (CAI) such as directly overlaying intra- and pre-operative imaging onto the endoscopic video to provide multi-modal visualisation with enhanced information about the location of sub-surface tumours and blood vessels. To correctly overlay this information onto the surgical console and update it as the camera moves by using the robot kinematics, it is necessary to estimate the hand-eye

Krittin Pachtrachai, Maximilian Allan and Danail Stoyanov are with the Centre for Medical Image Computing (CMIC) and the Department of Computer Science, University College London, Gower Street, WC1E 6BT, UK krittin.pachtrachai.13@ucl.ac.uk, maximilian.allan.11@ucl.ac.uk, danail.stoyanov@ucl.ac.uk

Vijay Pawar and Stephen Hailes are with the Department of Computer Science, University College London, Gower Street, WC1E 6BT, UK V.Pawar@cs.ucl.ac.uk, S.Hailes@cs.ucl.ac.uk

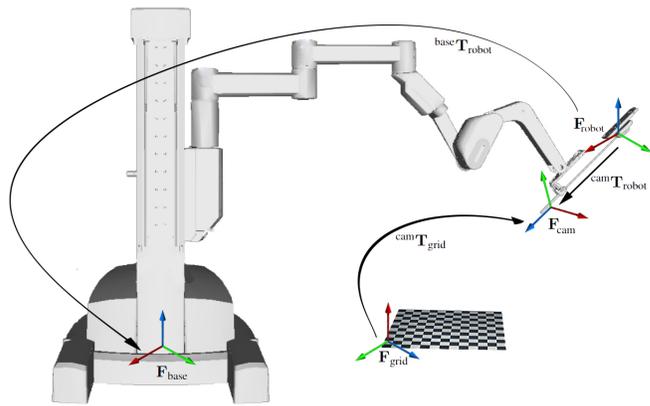


Fig. 1. Experimental set-up for the classic hand-eye calibration procedure with a da Vinci[®] (Intuitive Surgical, CA) surgical robot system. The camera is attached to the end-effector and ${}^{\text{cam}}\mathbf{T}_{\text{robot}}$ is the camera's pose relative to the end-effector [4].

transformation between the stereo-endoscopic camera and the robot coordinate system. When the hand-eye information is known, additional information exchange can happen to link information computed in the camera reference using surgical vision frame [5] to the robot kinematics for example to achieve enhance control with virtual fixtures [6] or motion compensation [7]. The challenge in hand-eye calibration during surgery is to ensure that the workflow during the procedure is not severely affected while being able to adapt to changes of the surgical camera and motion of the robot manipulator.

Conventional hand-eye calibration algorithms estimate the relative pose between the camera coordinate system and the robotic end-effector where the camera is mounted by using a calibration object as shown in Figure 1 as ${}^{\text{cam}}\mathbf{T}_{\text{robot}}$. A world coordinate frame is fixed as the origin of the calibration object \mathbf{F}_{grid} , whilst the robot's coordinate system is fixed at \mathbf{F}_{base} . Extrinsic parameter estimation during camera calibration or pose estimation are used to determine ${}^{\text{cam}}\mathbf{T}_{\text{grid}}$ and the robot's forward kinematics are used to determine ${}^{\text{base}}\mathbf{T}_{\text{robot}}$ from synchronised joint encoder data. The calibration object can have different designs, for example, [8], [9], [10], [11], [12], [13] use a planar checker board whereas [14], [15], [16] use a uniform planar grid of black circles. The principle behind estimating the hand-eye transformation using these objects is that because their physical dimensions are known in advance, it is possible to estimate the extrinsic parameters of the camera with respect to the object directly from images [17], [18]. In addition to planar calibration

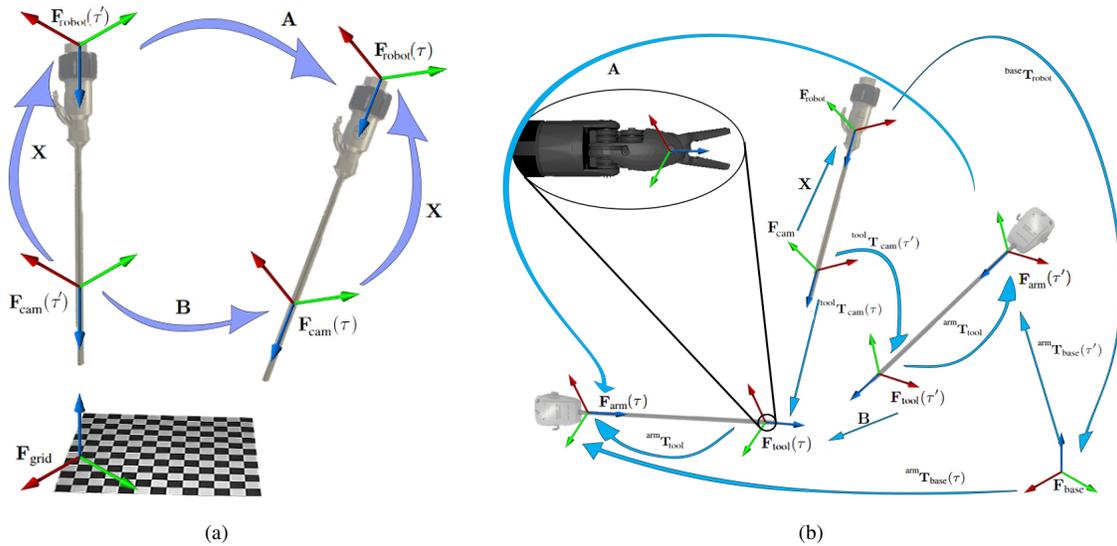


Fig. 2. (a) The schematic for general hand-eye calibration as mathematically represented by Eq. 2. The surgical camera is moved to different poses around the calibration object and images are captured at each position. (b) The schematic for the proposed hand-eye calibration incorporating a tool tracking algorithm [30] as mathematically represented by Eq. 3. No calibration grid is required and the camera does not move. Instead the instruments are moved by the robotic system.

objects, three dimensional (3D) objects can also be used but are more challenging to manufacture [19], [20]. Generally, calibration targets provide accurate estimation of the hand-eye transformation but can be practically inappropriate for many applications [21]. An alternative approach is to use structure-from-motion (SfM) to estimate the hand-eye calibration transformation directly from the camera’s surroundings [22], [23]. This type of approach is well suited to surgical applications as it does not alter the clinical workflow or require additional sterilised equipment. However, directly using SfM is not straightforward in surgery [24] because the surgical site is highly non-rigid and additionally the remote centre of motion (RCM) constraint on the surgical camera can prevent the collection of sufficiently varied viewpoints to provide accurate calibration. Another alternative solution is to attach a sensor at the end-effector and measure robot motion and sensor motion to identify hand-eye and robot-world transformations [25] and probabilistically solve an asynchronous problem between robotic joint information and sensor data. The approach does not require a correspondence between these two sets data and yields better results than when the exact correspondence is used, however, results have only been shown on simulation data and the method has not been tested with a real robotic system.

For applications in RMIS, alternative calibration objects for hand-eye calibration are the instruments used by the surgeon to manipulate the anatomy. For example if the geometry of the instrument is known, it is possible to use it as a constraint in general calibration of the imaging sensor even if the geometry is simple such as in needle procedures [26], [27]. In RMIS, the instruments and camera are forced to move asynchronously which avoids intractable mutual motion. The physical geometry of the instruments is known in 3D from manufacturing and computer aided design (CAD) models

are available. The laparoscopic camera can be controlled through the robotic system, which decreases the sensitivity of the calibration procedure to noise [8], [9], [28]. This means that numerous methods for instrument tracking in 3D from the laparoscopic video can be used as an initial platform for automated hand-eye calibration. Typically model fitting approaches making use of color features [29], [30], [31], edge features [32], local gradient descriptors [33] and a combination of shape-based features [30], [34] can be used to align a 3D model of the instrument with laparoscopic images.

In this paper, we introduce a new approach to estimate the hand-eye transformation for RMIS systems which does not require an additional calibration target. The method is an important step towards providing CAI capabilities to RMIS with automatic methods for updating calibration [5], [35]. We estimate the surgical instrument’s pose relative to the laparoscopic camera using prior methods for 3D instrument tracking [30] and we develop a novel algorithm to estimate the hand-eye transformation from the 3D pose data. We demonstrate, through extensive experiments on both synthetic and real data, that hand-eye calibration using CAD models of surgical tools achieves higher accuracy in rotation than using a planar object due to range of motion allowed by the instrument. We additionally compare our results with existing state-of-the-art hand-eye calibration methods [12], [15], [36] demonstrating the potential benefits of using instruments as calibration targets.

II. METHODS

A. Hand-eye Calibration

Hand-eye calibration involves determining the unknown fixed transform $\mathbf{X} = {}^{\text{robot}}\mathbf{T}_{\text{cam}}$ between a robot arm coordinate frame $\mathbf{F}_{\text{robot}}$ and a camera coordinate frame \mathbf{F}_{cam}

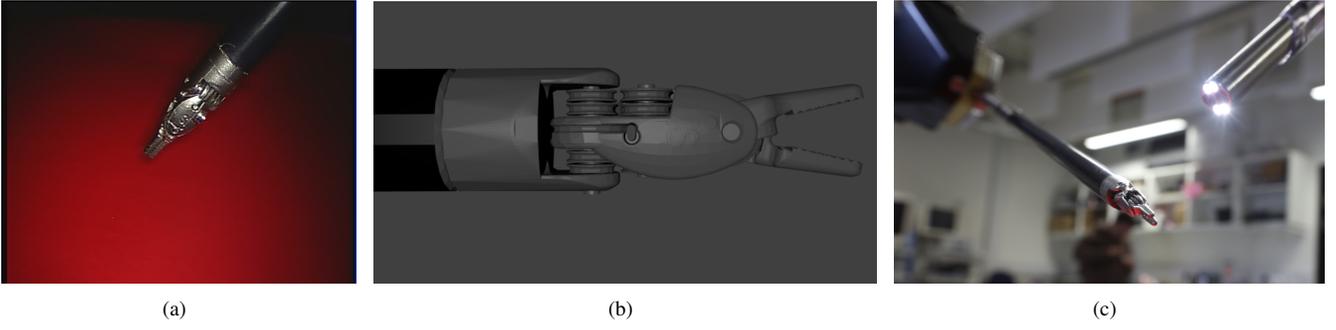


Fig. 3. (a) Shows the CAD model used in the instrument tracking algorithm. (b) Shows an image of the Large Needle Driver (LND) instrument captured by the camera. (c) Shows an image of the instrument moving within the field of view of the camera while it is being tracked.

statically attached to the arm (see Figure 1). The problem is formulated as a linear system of equations [14] as:

$$\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{B} \quad (1)$$

where \mathbf{A} is the relative motion between $\mathbf{F}_{\text{robot}}$ and a fixed reference frame at the base of the robot \mathbf{F}_{base} and is determined using forward kinematics [38], \mathbf{B} represents the relative motion between \mathbf{F}_{cam} and an observed reference frame, normally a chessboard calibration grid [8], \mathbf{F}_{grid} . These transforms can be written as the product of two rigid transformations:

$$\begin{aligned} \mathbf{A} &= {}^{\text{robot}}\mathbf{T}_{\text{base}}(\tau)({}^{\text{robot}}\mathbf{T}_{\text{base}}(\tau'))^{-1} \\ \mathbf{B} &= {}^{\text{cam}}\mathbf{T}_{\text{grid}}(\tau)({}^{\text{cam}}\mathbf{T}_{\text{grid}}(\tau'))^{-1} \end{aligned} \quad (2)$$

where τ and τ' are discrete time values indicating that the two transforms are captured at different time instances, as shown in Figure 2(a), where a solution exists when the two transforms have non-parallel rotation axes [8].

For the hand-eye problem, obtaining a wide range of camera to calibration target transforms increases the accuracy of the calibration and the error in estimating the rotation part of the transform is inversely proportional to the sine value of relative rotation axes [15]. In surgical robotic systems, which are constrained by an RCM, it is challenging to obtain the sufficient range of motion with the camera, however surgical instruments can in principle be positioned to acquire the necessary range. To enable hand-eye calibration using an instrument located at \mathbf{F}_{arm} , which is kinematically linked to \mathbf{F}_{base} , we reformulate Eq. 1 to write \mathbf{A} and \mathbf{B} as

$$\begin{aligned} \mathbf{A} &= {}^{\text{arm}}\mathbf{T}_{\text{base}}(\tau)({}^{\text{arm}}\mathbf{T}_{\text{base}}(\tau'))^{-1} \\ \mathbf{B} &= {}^{\text{tool}}\mathbf{T}_{\text{cam}}(\tau)({}^{\text{tool}}\mathbf{T}_{\text{cam}}(\tau'))^{-1} \end{aligned} \quad (3)$$

where the set-up is shown in Figure 2(b). We estimate ${}^{\text{arm}}\mathbf{T}_{\text{base}}$ and ${}^{\text{base}}\mathbf{T}_{\text{robot}}$ using forward kinematics provided by the DH parameters [38] of the robot arm and the joint encoder data. This leaves 2 unknowns in our reformulation, the hand-eye transformation \mathbf{X} and the transform between the camera coordinate system and the instrument coordinate system ${}^{\text{arm}}\mathbf{T}_{\text{tool}}$.

B. 3D Instrument Tracking

Normally, in standard camera calibration, estimating the camera to grid transform is achieved by solving a PnP

problem whereby the known grid coordinates are matched to easily detect image features. Estimating the transform ${}^{\text{tool}}\mathbf{T}_{\text{cam}}$ between the instrument and the camera from Eq. 3 where the target is a complex model such as a surgical instrument (see Figure 3(a)) is highly challenging. We use a recently proposed instrument tracking method [30] which estimates the 3D pose of robotic surgical instruments by solving a joint cost of aligning a 3D model of the instrument (see Figure 3(b)) with a color-based segmentation and a local optical flow point tracking. It minimizes this cost with gradient descent simultaneously across both eyes of the stereo camera effectively creating stereo constraints and uses a linear Kalman filter for temporal consistency in frame-to-frame tracking.

C. Hand-eye Calibration with Adjoint Transformation

To solve our reformulation of Eq. 1, we have adopted the hand-eye solution in [12] as a starting point and refer the reader to the original paper for a detailed overview of the method. Let $\mathbf{a} + \mathbf{a}'\epsilon$, $\mathbf{q} + \mathbf{q}'\epsilon$ and $\mathbf{b} + \mathbf{b}'\epsilon$ be dual quaternion representing homogeneous transformations \mathbf{A} , ${}^{\text{arm}}\mathbf{T}_{\text{tool}}$ and \mathbf{B} , and $\epsilon^2 = 0$. Thus Eq. 1 can be written as

$$(\mathbf{a} + \mathbf{a}'\epsilon)(\mathbf{q} + \mathbf{q}'\epsilon) = (\mathbf{q} + \mathbf{q}'\epsilon)(\mathbf{b} + \mathbf{b}'\epsilon) \quad (4)$$

The dual quaternion can be separated into a rotation and translation component.

$$\mathbf{a}.\mathbf{q} = \mathbf{q}.\mathbf{b} \quad (5)$$

$$\mathbf{a}'.\mathbf{q} + \mathbf{a}.\mathbf{q}' = \mathbf{q}.\mathbf{b}' + \mathbf{q}'.\mathbf{b} \quad (6)$$

Eq. 5 can be rearranged into the matrix equation,

$$\begin{bmatrix} a_0 - b_0 & -(\vec{a} - \vec{b})^T \\ \vec{a} - \vec{b} & [\vec{a} + \vec{b}]_{\times} + (a_0 - b_0)\mathbf{I}_3 \end{bmatrix} \mathbf{q} = \mathbf{0} \quad (7)$$

$$\mathbf{K}(\mathbf{a}, \mathbf{b})\mathbf{q} = \mathbf{0}$$

where \mathbf{I}_3 is an identity matrix of size 3×3 . The N \mathbf{K} matrices can be stacked into a $4N \times 4$ matrix which allows the rotation quaternion \mathbf{q} to be determined using SVD. [19] proposed an alternative method which through the modification of \mathbf{K} allows the solver to estimate rotation and translation simultaneously. However, this method is more sensitive to noise and the use of a quadratic equation has a possibility of generating a solution that contains imaginary values, which

complicates its use in our proposed application.

Using the initial solution from Eq. 7, we propose a method using an adjoint transformation relationship to enhance the accuracy of hand-eye transformation estimation. Using Eq. 1, we post multiply by ${}^{\text{arm}}\mathbf{T}_{\text{tool}}^{-1}$ and retrieve the equation in the form of a similarity transformation $\mathbf{A} = {}^{\text{arm}}\mathbf{T}_{\text{tool}} \mathbf{B} {}^{\text{arm}}\mathbf{T}_{\text{tool}}^{-1}$, which means that \mathbf{A} and \mathbf{B} are the same transformation applied in different frames. This relationship can be written as

$$\begin{bmatrix} \vec{\omega}_a \\ \vec{v}_a \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{0}_{3 \times 3} \\ [\vec{t}]_{\times} & \mathbf{R} \end{bmatrix} \begin{bmatrix} \vec{\omega}_b \\ \vec{v}_b \end{bmatrix} \quad (8)$$

where \mathbf{R} and \vec{t} are the rotation and translation components of ${}^{\text{arm}}\mathbf{T}_{\text{tool}}$, respectively. \vec{v}_a, ω_a and \vec{v}_b, ω_b are the members of Lie group for a rigid transformation \mathbf{A} and \mathbf{B} respectively, calculated from the matrix logarithm formula explained in [41], [42].

$$\begin{bmatrix} [\vec{\omega}_a]_{\times} & \vec{v}_a \\ \vec{0}^T & 0 \end{bmatrix} = \log \left(\begin{bmatrix} \mathbf{R}_A & \vec{t}_a \\ \vec{0}^T & 1 \end{bmatrix} \right) \quad (9)$$

When considering all the collected poses \mathbf{A} and \mathbf{B} , the first row of Eq. 8 will represent the orthogonal Procrustes Problem [43] which has been satisfied by the dual quaternion approach. Referring to the second row, as the translation and rotation part are already estimated with the dual quaternion, we can deduce that the solution satisfies

$$\vec{v}_a = [\vec{t}]_{\times} \mathbf{R} \vec{\omega}_b + \mathbf{R} \vec{v}_b \quad (10)$$

As the instrument tracking algorithm is less accurate than a grid-based pose estimation, some measurements will have inconsistently high error and to enable the algorithm to estimate a good solution for the hand-eye calibration, we substitute a rotation component of \mathbf{B} with high error with the rotation component of \mathbf{A} . According to the orthogonal Procrustes Problem, the norm of ω_a and ω_b are equal in an ideal case, where the norm of these vectors represent angle of rotation. To eliminate this outlier, the substitution $\omega_a = \mathbf{R} \omega_b$ is applied when there is more than 5 degree difference between the two norms. After the substitution, we will have:

$$\vec{v}_a = [\vec{t}]_{\times} \vec{\omega}_a + \mathbf{R} \vec{v}_b \quad (11)$$

Therefore, to estimate the rotation refinement, we need to compute a real quaternion \mathbf{q} that optimises Eq. 7 and Eq. 11. Additionally, to estimate the translation, we again rearrange Eq. 11 such that.

$$[\vec{\omega}_a] \vec{t} = \mathbf{R} \vec{v}_b - \vec{v}_a \quad (12)$$

Therefore, the solution satisfies both Eq. 11 and Eq. 12. The algorithm then goes back to refine the translation component and this process is repeated until the solution converges. In this paper, the convergence is defined when the change of rotation vectors and translation vectors are both less than 0.001 radians and 1 mm, respectively for 20 iterations. The rotation vector is defined as a principle axis of rotation with norm equals to degree of rotation.

The final stage in estimating the hand-eye transform involves calculating a per-frame hand-eye transformation \mathbf{X}_i

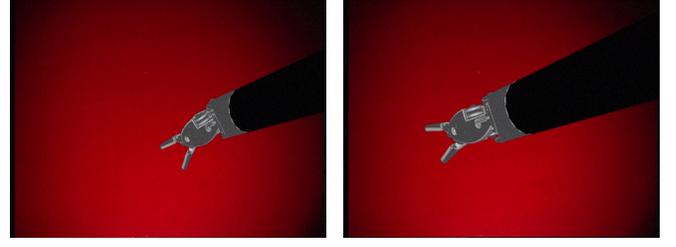


Fig. 4. Synthetic data generated from projecting an instrument CAD tool at poses generated from kinematic data and an artificial hand-eye transform. The images are corrupted with Gaussian noise but the kinematic data is noiseless.

Algorithm 1 Hand-Eye Calibration

```

1: procedure HANDEYE
2:    ${}^{\text{base}}\mathbf{T}_{\text{arm}} \leftarrow$  forward kinematics
3:    $\mathbf{A} \leftarrow$  construct tool's relative motions
4:    ${}^{\text{tool}}\mathbf{T}_{\text{cam}} \leftarrow$  tool tracking
5:    $\mathbf{B} \leftarrow$  construct tracking target's relative motions
6:    $\mathbf{R}_0 \leftarrow$  solve Eq. 7
7:   repeat
8:      $\vec{t}_{\text{refine}} \leftarrow$  solve Eq. 12 using  $R_0$ 
9:      $\vec{t}_0 \leftarrow \vec{t}_{\text{refine}}$ 
10:     $R_{\text{refine}} \leftarrow$  solving Eq. 7, 11 using  $\vec{t}_0$ 
11:     $\mathbf{R}_0 \leftarrow \mathbf{R}_{\text{refine}}$ 
12:   until the solution converges
13:    ${}^{\text{arm}}\mathbf{T}_{\text{tool}} \leftarrow [\mathbf{R}_0, \vec{t}_0; 0 \ 0 \ 0 \ 1]$ 
14:    $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N \leftarrow$  solve eq. 13 for  $N$  frames.
15:    $\mathbf{X} \leftarrow$  average transformations  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ 
16:   return  $\mathbf{X}$  ▷ Hand to eye transformation
17: end procedure

```

before average them across each frame using the algorithm described in [44].

$$\mathbf{X}_i = ({}^{\text{base}}\mathbf{T}_{\text{robot}})^{-1} ({}^{\text{arm}}\mathbf{T}_{\text{base},i})^{-1} ({}^{\text{arm}}\mathbf{T}_{\text{tool}}) ({}^{\text{tool}}\mathbf{T}_{\text{cam},i}) \quad (13)$$

The algorithm to achieve this is summarised in Algorithm 1.

III. EXPERIMENTS AND RESULTS

A. Experiments with synthetic data

Synthetic data is generated by rendering a CAD model of a LND instrument (see Figure 3(b)) onto a virtual camera using OpenGL. To maintain realism in the type of motion that we would expect to see when using a surgical robotic system, we collect joint encoder data using the da Vinci Research Kit (DVRK) [45]. We use this to compute model-view transforms using forward kinematics for a virtual da Vinci classic robot. We generate a ground truth hand-eye transformation which is combined with the computed transforms. For each new set of joint values, we compute a new set of transforms and render the instrument at the appropriate pose. Examples images from this dataset are shown in Figure 4(a) and 4(b).

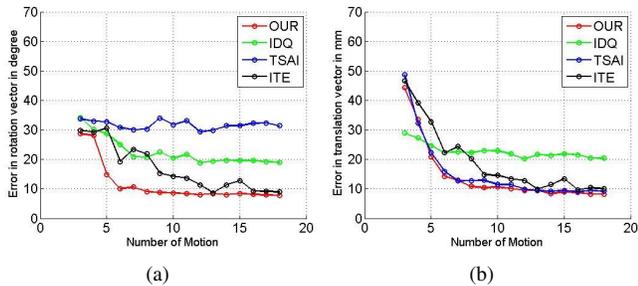


Fig. 5. Performance of each algorithm when it is tested with synthetic data with increasing number of poses used in the calibration. (a) Error in rotation estimation. (b) Error in translation estimation.

Our hand-eye calibration algorithm is compared with algorithms proposed in recent state-of-the-art calibration methods [12], [15], [36]. We use synthetic data to understand how increasing the number of captured poses affects the estimated transformation and how the algorithm behaves when the image and kinematic data are corrupted with increasing Gaussian noise. Despite the fact that Gaussian noise is not an ideal noise model for the data simulation, it is added to this synthetic image to simulate image distortion. We add noise to the kinematic data because commercial surgical robotic systems, such as da Vinci, use tendon actuation for the joints which introduces absolute posing errors in estimating the transforms using forward kinematics [37]. This results in a discrepancy between an observed pose in the camera coordinates and an instrument pose estimated through forward kinematics, as the error introduces a small displacement between them.

The hand-eye result computed using each method is compared with the ground truth. The error in the rotation is calculated with Rodrigues’ formula to determine the relative degree of rotation between the two rotation matrices, while for the translation we use root mean square (RMS) error. The nomenclatures used in this section and the next are “OUR” for our algorithm, “IDQ” for [12], “TSAI” for [15] and “ITE” for [36].

1) *Increasing number of poses*: The results in Figure 5(a) and 5(b) show the performance of each algorithm when the number of poses is increased, which has been found to reduce the geometric error in hand-eye calibration [14]. In this experiment, no noise is added to the joint data, however we add constant Gaussian noise to the image data. Therefore, as the kinematic joint data is perfect, the error in the hand-eye calibration is only caused by the error in estimating the 3D instrument pose using the tool tracking algorithm. Compared with the literature state-of-the-art methods, our algorithm shows superiority with quicker convergence rate and lower error in both rotation and translation part.

2) *Behaviour with increasing Gaussian noise*: Most existing research [13], [19], [23], [46] regarding hand-eye calibration observes a linear relationship between noise and error in both the rotation and translation estimation. However, as illustrated in Figure 6(a) and 6(b), our experiments show that the linear behaviour is less evident when using

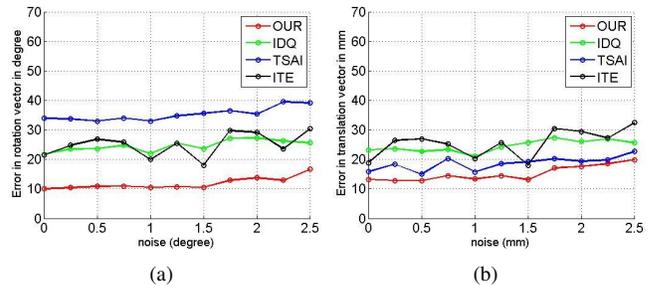


Fig. 6. Performance of each algorithm when it is tested with synthetic data with increasing intensity of Gaussian noise in the kinematics, this replicates the real noise that is seen when estimating poses from tendon-driven robots such as da Vinci. (a) Error in rotation estimation. (b) Error in translation estimation.

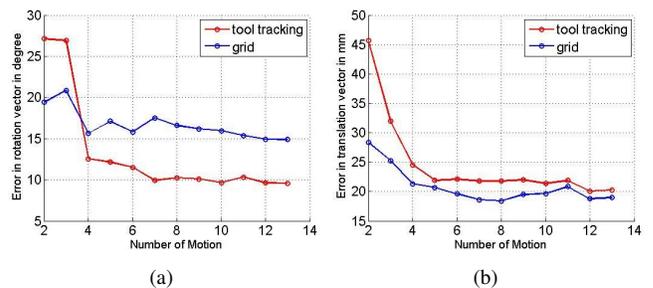


Fig. 7. Comparison between calibration performance with a standard checkerboard and with a surgical instrument. The error in the rotation is significantly improved when using the instrument and translation error is comparable between both calibration targets.

our instrument tracking approach compared with using a calibration grid. This is caused by higher errors in estimating ${}^{\text{tool}}\mathbf{T}_{\text{cam}}$ using tool tracking methods compared with estimating ${}^{\text{grid}}\mathbf{T}_{\text{cam}}$ using convention methods. Additionally, the performance of the tool tracking algorithm is sensitive to drift, resulting in increasingly noisy estimates of the transform, which in turn increases the noise in the relative transformation \mathbf{B} in Eq. 3.

B. Comparison between instrument and grid hand-eye calibration

We compare the accuracy of using a surgical instrument as a hand-eye calibration target with a conventional checkerboard using real image data to demonstrate the advantage of using an instrument as a calibration target. As we do not have a ground truth hand-eye calibration for the real data, we have to use the camera pose prediction as our validation method. The concept is to find the error between the predicted camera pose when using our estimated hand-eye transform and the extrinsic pose obtained from camera calibration. Referring to Eq. 2, we can predict camera pose as:

$${}^{\text{arm}}\mathbf{T}_{\text{base}}(\tau_{\text{predicted}}) = \mathbf{X}\mathbf{B}\mathbf{X}^{-1} {}^{\text{arm}}\mathbf{T}_{\text{base}}(\tau') \quad (14)$$

Figure 7(a) shows the improvement in rotation estimation when using a surgical tool as a calibration target. However, Figure 7(b) shows that a conventional grid-based calibration still has some advantages over our proposed method. Despite the higher accuracy in the rotation estimation that can also

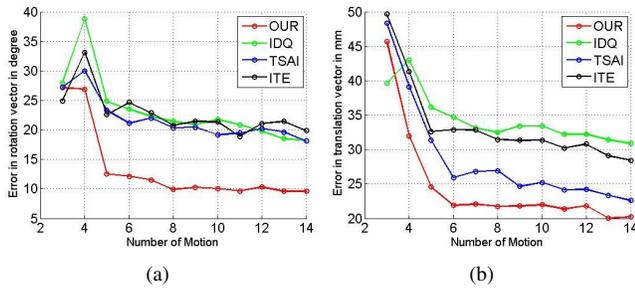


Fig. 8. (a) Shows error in rotation component when the algorithms are tested with real data against increasing the number of poses used in the calibration. (b) Shows error in translation component.

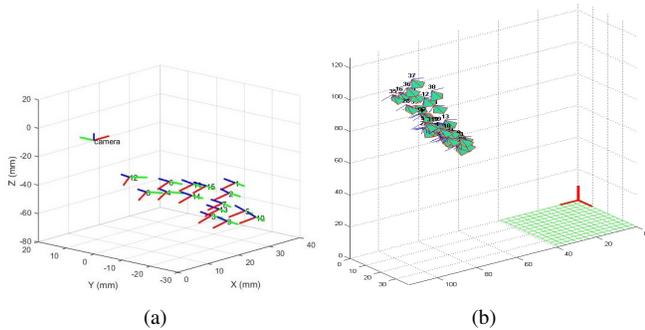


Fig. 9. (a) Shows examples of the surgical instrument pose with respect to camera. The surgical tool has a much wider motion range than the camera and can therefore be used to create much more varied poses for the calibration. (b) Shows examples of the camera pose with respect to a calibration grid frame [17]. Due to the RCM constraint of surgical robotic systems, the range of poses is confined within a very restricted space.

suppress error in translation, the translational error when calibrating with a grid is lower than when using an instrument due to inaccuracies in correctly estimating the instrument pose, particularly when using few images. However, after the addition of several images, the error in estimating the translation is comparable between the two methods. Note that we only select a few frames out of whole data to avoid confusion in the figures. Figure 9(b) illustrates that the RCM constrained camera does not cover a sufficiently varied range of poses, only moving with a small translation along the insertion axis and a small rotation about the axis, while a motion of surgical tool (shown in Figure 9(a)) spreads widely across the frame.

C. Comparison between hand-eye algorithms

We also compare our novel hand-eye algorithm with other existing hand-eye solutions using the same real data. Figure 8(a) and 8(b) shows the same trend as the synthetic result, except that the errors in the experiment with real data are higher than with synthetic which is expected. For a small number of poses, every algorithm is still corrupted by noise from tracking error. When it tends to convergence, our algorithm achieves about 20 mm error in translation and 10 degree error in rotation which is the lowest error.

IV. CONCLUSION

In this paper, we proposed a hand-eye calibration method using a robotic surgical instrument CAD model as the calibration object. The approach is a practical way of performing hand-eye calibration during robotic assisted surgery without requiring additional calibration objects. Additionally, the method makes it easier to collect the sufficiently wide range of poses necessary for accurate hand-eye calibration because the surgical tool can be moved directly at the surgical site. This results in simpler workflow for any practical applications that require multiple calibrations during a surgical procedure. As well as improving the ergonomics of calibration, our results on simulation data show that the developed algorithm provides higher accuracy than existing methods and this result is further confirmed through our experiments with real data. Our method improves the estimation of the rotation component of the transformation while having comparable errors in estimation of the translation component even when only several calibration frames are used. In future work we aim to enhance the algorithm by better integration with automatic instrument detection and pose estimation and to integrate automatic camera calibration which at present is considered to be known.

ACKNOWLEDGEMENT

We would like to thank Simon Di Maio and Mahdi Azizian at Intuitive Surgical, CA for providing us with CAD models of the instrument and the DVRK community for their support. The work was part funded by the EPSRC (EP/N013220/1, EP/N022750/1, EP/N027078/1, NS/A000027/1), The Wellcome Trust (WT101957, 201080/Z/16/Z) and the EU-Horizon2020 project EndoVESPA (H2020-ICT-2015-688592).

REFERENCES

- [1] L. R. Kavoussi, R. G. Moore, J. B. Adams and A. W. Partin, "Comparison of robotic versus human laparoscopic camera control", in *The Journal of Urology*, 1995 Dec; 158(4): 2134-2136.
- [2] J. Marescaux, M. K. Smith, D. Fölscher, F. Jamali, B. Malassagne and J. Leroy, "Telerobotic laparoscopic cholecystectomy: initial clinical experience with 25 patients", in *Annals of Surgery*, 2001 July; 234(1): 1-7.
- [3] M. Hayashibe, N. Suzuki, M. Hashizume, Y. Kakeji, K. Konishi, S. Suzuki and A. Hattori, "Preoperative planning system for surgical robotics setup with kinematics and haptics", in *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2005 Jan; 1(2): 76-85.
- [4] Intuitive Surgical, Inc. ISI API User Guide, 2014.
- [5] D. Stoyanov, "Surgical vision", in *Annals of Biomedical Engineering*, 2012 Feb; 40(2): 332-34.
- [6] S. A. Bowyer, B. L. Davies and F. R. Baena, "Active constraints/virtual fixtures: a survey", in *IEEE Transactions on Robotics*, 2014 Feb; 30(1): 138-157.
- [7] A. Ruszkowski, C. Schneider, O. Mohareri and S. Salcudean, "Bimanual teleoperation with heart motion compensation on the da vinci@research kit: implementation and preliminary experiments", in *ICRA*, 2016, pp. 4101-4108.
- [8] C. Chen, S. Stitt and Y. F. Zheng, "Robotic eye-in-hand calibration by calibrating optical axis and target pattern", in *Journal of Intelligent and Robotic Systems*, 1995 June; 12(2): 155-173.
- [9] F. Mourgues and É. Coste-Manière, "Flexible calibration of actuated stereoscopic endoscope for overlay in robot assisted surgery", in *MICCAI*, 2002, pp. 25-34.

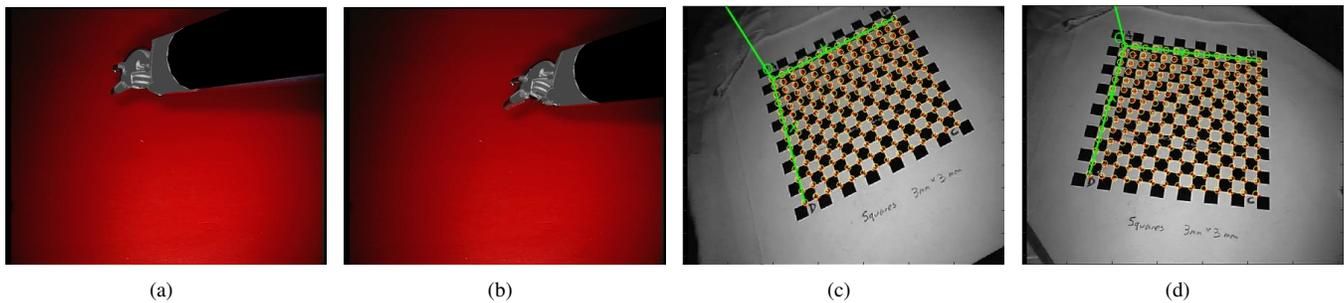


Fig. 10. (a)-(b) shows instrument tracking result from the algorithm of [30] when it is re-projected onto images. We use the estimates from this algorithm to compute the hand-eye transform. (c)-(d) shows example of calibration grid images taken from camera which are used for extrinsic calibration afterwards to estimate the pose of camera with respect to the grid.

- [10] J. Schmidt, F. Vogt and H. Niemann, "Robust hand-eye calibration of an endoscopic surgery robot using dual quaternions", in *DAGM Symposium*, 2003, pp. 548-556.
- [11] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration", in *IROS*, 2006, pp. 4647-4653.
- [12] A. Malti and J. P. Barreto, "Robust hand-eye calibration for computer aided medical endoscopy", in *IEEE ICRA*, 2010, pp. 5543-5549.
- [13] Z. Kukulova, J. Heller and T. Pajdla, "Hand-eye calibration without hand orientation measurement using minimal solution", in *ACCV*, 2012, pp. 576-589.
- [14] R. Y. Tsai and R. K. Lenz, "Real time versatile robotics hand/eye calibration using 3D machine vision", in *IEEE Transactions on Robotics and Automation*, 1988 Apr; 1: 554-561.
- [15] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration", in *IEEE Transactions on Robotics and Automation*, 1989 June; 5(3): 345-358.
- [16] H. Zhuang, "Hand/eye calibration for electronic assembly robots", in *IEEE Transactions on Robotics and Automation*, 1998 Aug; 14(4): 612-616.
- [17] J. Y. Bouguet (2013, December 2). *Camera Calibration Toolbox for Matlab* [Online]. Available: http://http://www.vision.caltech.edu/bouguetj/calib_doc/
- [18] S. J. D. Prince, *Computer Vision Models, Learning and Inference*, 1st ed. UK: Cambridge University Press, 2012.
- [19] K. Daniilidis, "Hand-eye calibration using dual quaternions", in *The International Journal of Robotics Research*, 1999 Mar; 18(3): 286-298.
- [20] N. Andreff and R. Espiau, "Robot Hand-Eye Calibration Using Structure-from-Motion", in *The International Journal of Robotics Research*, 2001 Mar; 20(3): 228-248.
- [21] D. Stoyanov, A. Darzi and G.-Z. Yang, "Laparoscope self-calibration for robotic assisted minimally invasive surgery", in *MICCAI*, 2005, pp. 114-121.
- [22] J. Schmidt, F. Vogt and H. Niemann, "Calibration-free hand-eye calibration: a structure-from-motion approach", in *DAGM Symposium*, 2005, pp. 67-74.
- [23] J. Heller, M. Havlena, A. Sugimoto and T. Pajdla, "Structure-from-motion based hand-eye calibration using L_∞ minimization", in *IEEE CVPR*, 2011, pp. 3497-3503.
- [24] P.-L. Chang, A. Handa, A. J. Davison, D. Stoyanov and P. Edwards, "Robust real-time visual odometry for stereo endoscopy using dense quadrifocal tracking", in *International Conference Information Processing in Computer Assisted Intervention*, 2014.
- [25] H. Li, Q. Ma, T. Wang and G. S. Chirikjian, "Simultaneous hand-eye and robot-world calibration by solving the $AX = YB$ problem without correspondence", in *IEEE Robotics and Automation Letters*, 2016 Jan; 1(1): 145-152.
- [26] F. Vasconcelos, D. Peebles, S. Ourselin and D. Stoyanov, "Spatial calibration of a 2D/3D ultrasound using a tracked needle", in *International Journal of Computer Assisted Radiology and Surgery*, 2016 June; 11(6): 1091-1099.
- [27] F. Vasconcelos, D. Peebles, S. Ourselin and D. Stoyanov, "Similarity registration problems for 2D/3D ultrasound calibration", in *ECCV*, 2016 [Accepted]. Available: <http://arxiv.org/abs/1608.00247>.
- [28] J. Schmidt and H. Niemann, "Data selection for hand-eye calibration: a vector quantization approach", in *The International Journal of Robotics Research*, 2008 Sep; 27(9): 1027-1053.
- [29] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly and D. Stoyanov, "Towards detection and localization of instruments in minimally invasive surgery", in *IEEE Transactions on Biomedical Engineering*, 2013 Apr; 60(4): 1050-1058.
- [30] M. Allan, P.-L. Chang, S. Ourselin, D. J. Hawkes, A. Sridhar, J. Kelly and D. Stoyanov, "Image based surgical instrument pose estimation with multi-class labelling and optical flow", in *MICCAI*, 2015, pp. 331-338.
- [31] Z. Pezzementi, S. Voros and G. D. Hager, "Articulated object tracking by rendering consistent appearance parts", in *IEEE ICRA*, 2009, pp. 3940-3947.
- [32] S. Voros, E. Orvain, P. Cinquin and J.-A. Long, "Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders", in *The International Journal of Robotics Research*, 2007 Nov; 26(11-12): 1173-1190.
- [33] A. Reiter, P. K. Allen and T. Zhao, "Feature classification for tracking articulated surgical tools", in *MICCAI*, 2012, pp. 592-600.
- [34] M. Allan, S. Thompson, M. J. Clarkson, S. Ourselin, D. J. Hawkes, J. Kelly and D. Stoyanov, "2D-3D pose tracking of rigid instruments in minimally invasive surgery", in *IPCAI*, 2014, pp. 1-10.
- [35] J. Lin, N. T. Clancy, D. Stoyanov and D. S. Elson, "Tissue surface reconstruction aided by local normal information using a self-calibrated endoscopic structured light system", in *MICCAI*, 2015, pp. 405-412.
- [36] H. Zhuang and Y. C. Shiu, "A noise-tolerant algorithm for robotic hand-eye calibration with or without sensor orientation measurement", in *IEEE Transactions on Systems, Man, and Cybernetics*, 1993 Jul/Aug; 23(4): 1168-1175.
- [37] J. H. Jang, S. H. Kim and Y. K. Kwak, "Calibration of geometric and non-geometric errors of an industrial robot", in *Robotica*, 2001 May; 19(3): 311-321.
- [38] M. W. Spong, S. Hutchinson and M. Vidyasagar, *Robot Modeling and Control* 1st ed. USA: Wiley, 2006.
- [39] J. C. K. Chou and M. Kamel, "Quaternions approach to solve the kinematic equation of rotation $\mathbf{A}_a \mathbf{A}_x = \mathbf{A}_x \mathbf{A}_b$ of a sensor-mounted robotic manipulator", in *IEEE ICRA*, 1988, pp. 656-662.
- [40] J. C. K. Chou and M. Kamel, "Finding the position and orientation of a sensor on a robot manipulator using quaternions", in *International Journal of Robotics Research*, 1991 June; 10(3): 240-254.
- [41] R. M. Murray, Z. Li and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation* 1st ed. CRC Press Inc., 1994.
- [42] J. E. Blanco, "A tutorial on $SE(3)$ transformation parameterizations and on-manifold optimization", Universidad de Málaga, 2014.
- [43] P. Schönemann, "A generalized solution of the orthogonal procrustes problem", in *Psychometrika*, 1966 Mar; 31(1): 1-10.
- [44] L. Kavan, S. Collins, C. O'Sullivan and J. Zara, "Dual quaternions for rigid transformation blending", Trinity College Dublin, 2006.
- [45] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor and S. P. DiMaio, "An Open-Source Research Kit for the da Vinci Surgical System", in *IEEE ICRA*, 2014, pp. 6434-6439.
- [46] F. Dornaika and R. Horaud, "Simultaneous robot-world and hand-eye calibration", in *IEEE Transactions on Robotics and Automation*, 1998 Aug; 14(4): 617-622.