



Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2010 April 14; 2010: 1401–1404. doi:10.1109/ISBI.2010.5490260.

C-arm Pose Estimation Using a Set of Coplanar Ellipses in Correspondence

Maria S. Ayad¹, Junghoon Lee², Anton Deguet¹, Everette C. Burdette³, and Jerry L. Prince^{2,1}

¹Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

²Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

³Acoustic MedSystems, Champaign, IL, USA

Abstract

C-arms are increasingly being used to assist in a large number of surgical procedures. Fairly accurate and fast pose estimates are needed for non-encoded c-arms that are commonly available in most operating rooms in order to attain quantitative feedback from the x-ray images. We propose the use of an image-based fiducial composed of a set of coplanar ellipses to track the c-arm. We adopt an existing method for planar homography and propose a variation consisting of three modifications: including a weighting scheme for the linear system used, orthonormalizing the vectors pertaining to the rotation component of the transformation, and fine tuning the estimates using a constrained optimization step. We show that these variations make the approach more robust to noise that typically arises in fluoroscopy imaging and guarantee the orthonormality of the estimated rotation. The performance of the modified algorithm is demonstrated using realistic x-ray simulations. We also run sensitivity analysis for segmentation and calibration errors that are likely to occur in a practical setting. Preliminary results show mean tracking accuracy within 0.5° and 0.9 mm for segmentation error variance up to 2 pixels squared. The algorithm also proves to be robust to calibration errors up to 1 cm.

Index Terms

C-arm; tracking; pose estimation; conics; planar fiducial

1. Introduction

C-arm fluoroscopy has been widely used for qualitative assessment of a variety of computer-assisted surgical procedures [1]. Attaining quantitative feedback from the x-ray images will become more feasible and widely available through developing fast, cheap and reliable techniques for calibration and pose estimation for the c-arm. In this paper we present a clinically friendly solution for pose determination of the commonly-available nonencoded c-arms that exist in most operating rooms (OR)s. There exist two major approaches for pose recovery: reliance on auxiliary devices and employing image-based tracking fiducials. External trackers, such as optical and EM trackers have proven to be both expensive and cumbersome in the OR. Optical trackers require a line of sight, whereas EM trackers are sensitive to the presence of metal objects in the work area. Therefore, fiducial-based pose estimation has gained wider clinical acceptance [2].

For a fiducial to be clinically appealing, it needs to be compact in size, easy to include in the surgical working space and noninterfering with the trajectory of the c-arm motion, which

often requires a wide field of view. This motivated the idea of a flat fiducial that can be cheaply manufactured and easily placed under the patient and tracked without special mounting or fixation issues. If the x-ray absorption that is created by this fiducial is relatively low, then it will be visible to the clinician, but will not interfere with visualization of the anatomy or surgical tools that are within the field of view. We thus needed to choose suitable features that can satisfy a number of requirements. First, they need to be easily embedded in a simple non-invasive flat fiducial; second, they should accurately recover the pose intraoperatively; third, they must neither interfere with the physical constraints of the space nor negatively impact the quality of the images needed for guidance. Our choice was a set of coplanar ellipses as shown in Fig. 1(b). The study of conics for pose estimation and object recognition has been an active research topic for several groups due to a number of reasons. First, conics are more compact than points or segments. Second, segmentation of conics is more immune to noise due to the large number of points on the curves. Third, they are easier to establish correspondence. Fourth, they provide closed-form solutions. Ellipses in particular are especially attractive since a 3D ellipse projects to an ellipse in the image. It has been shown in [3, 4] that c-arm pose estimation for computer-assisted surgery can be achieved by a single ellipse and a point correspondence.

In the current work, we eliminate the need for points and utilize only planar ellipses with known correspondence. With planar targets, the problem of pose determination amounts to solving a planar homography. Solving camera pose and planar homography in general using conics has been previously presented by several authors including De Ma [5], Forsyth et al. [6], Sugimoto [7] and Kannala et al. [8]. While De Ma [5] proposes that two ellipses can uniquely determine the pose, it is not clear how this estimate can deteriorate in the presence of noise similar to our target application. On the other hand, Sugimoto's method [7] utilizes a minimum of seven conics. Kannala et al. however, present an attractive approach that needs three or more corresponding conics for computing planar homographies. We adopt this method and apply three main modifications to be able to solve the problem of pose estimation for c-arms from the noisy fluoroscopic images that are typically taken during surgical procedures. First, we propose a weighting technique that accounts for segmentation errors, being a dominant source of noise. Second, we enforce the constraints needed to ensure that the homography computed can represent a meaningful pose in terms of the rotational component. Finally, a constrained optimization step is added to the algorithmic flow to fine tune all the pose parameters after the third rotation vector is obtained.

2. Materials and Methods

2.1. Perspective projection of a conic

Let $\mathbf{x}_w = [x_w, y_w, z_w]^T$ be a point in the world frame (which is the fiducial frame) and $\mathbf{x} = [x, y, z]^T$ be the same point in the camera frame, as shown in Fig. 1(a). They are related by the equation

$$\mathbf{x} = \mathbf{R} \mathbf{x}_w + \mathbf{t}, \quad (1)$$

where $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$ is the rotation matrix and \mathbf{t} is the translation vector between the two frames. With no loss of generality, consider points in the $x_w y_w$ -plane; for all such points, the world frame coordinates can be given by vectors of the form $\mathbf{x}_w = [x_w, y_w, 0]^T$. In this case, (1) reduces to

$$\mathbf{x}=\mathbf{G} \mathbf{u}_w, \quad (2)$$

where $\mathbf{G} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}]$ and $\mathbf{u}_w = [x_w, y_w, 1]^T$.

C-arm imaging is modeled by the full perspective projection model where the optical center is the origin of the camera frame, the z -axis is the optical axis and the image plane is parallel to the xy -plane at a distance f from the origin. Let $[o_u, o_v]$ be the location of the image center in pixel coordinates and s_u, s_v be the pixel sizes in the u and v directions respectively. Then

the image point $\mathbf{u} = [u, v, 1]^T$ can be given by $u = \frac{f x}{s_u z} + o_u$ and $v = \frac{f y}{s_v z} + o_v$. Therefore

$$\mathbf{x}=z \mathbf{C} \mathbf{u}=\mathbf{G} \mathbf{u}_w, \quad (3)$$

where

$$\mathbf{C}=\begin{bmatrix} s_u/f & 0 & -o_u s_u/f \\ 0 & s_v/f & -o_v s_v/f \\ 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

Let the equation of a conic in the $x_w y_w$ -plane in the world frame be given by

$$\mathbf{u}_w^T \mathbf{Q} \mathbf{u}_w=0, \quad (5)$$

and the equation of the corresponding projected conic in the image plane be given by

$$\mathbf{u}^T \mathbf{A} \mathbf{u}=0. \quad (6)$$

Using (3) in (6), we get

$$\mathbf{u}_w^T \mathbf{G}^T \mathbf{C}^{-T} \mathbf{A} \mathbf{C}^{-1} \mathbf{G} \mathbf{u}_w=0, \quad (7)$$

where \mathbf{C}^{-T} is the transposed inverse of \mathbf{C} . Comparing (5) to (7), we have

$$\mathbf{G}^T \mathbf{C}^{-T} \mathbf{A} \mathbf{C}^{-1} \mathbf{G}=\mathbf{G}^T \tilde{\mathbf{A}} \mathbf{G}=k \mathbf{Q}, \quad (8)$$

where k is a constant and $\tilde{\mathbf{A}} = \mathbf{C}^{-T} \mathbf{A} \mathbf{C}^{-1}$.

2.2. Using one image of multiple conics

An extension of this formulation for estimating the pose using a single image of multiple conics has been adopted from the method proposed by Kannala et al. [8] and is also written

out below. First, consider two coplanar conics \mathbf{Q}_1 and \mathbf{Q}_2 with respective images \mathbf{A}_1 and \mathbf{A}_2 as in (5) and (6). From (8) we have

$$\mathbf{G}^T \tilde{\mathbf{A}}_1 \mathbf{G} = k_1 \mathbf{Q}_1, \quad (9)$$

$$\mathbf{G}^T \tilde{\mathbf{A}}_2 \mathbf{G} = k_2 \mathbf{Q}_2. \quad (10)$$

In order to use the above two equations simultaneously, all the coefficient matrices are normalized to have unit Frobenius norm. Then, the constant k_i is computed by fixing the determinant of the matrix of unknowns to be unity [8]. Thus

$$k_i = \left(\frac{\det(\tilde{\mathbf{A}}_i)}{\det(\mathbf{Q}_i)} \right)^{\frac{1}{3}}, \quad i=1, 2. \quad (11)$$

Each \mathbf{Q}_i is then replaced by $k_i \mathbf{Q}_i$ and we now have

$$\mathbf{M}^T \tilde{\mathbf{A}}_1 \mathbf{M} = \mathbf{Q}_1, \quad (12)$$

$$\mathbf{M}^T \tilde{\mathbf{A}}_2 \mathbf{M} = \mathbf{Q}_2. \quad (13)$$

where \mathbf{M} is a constant multiple of \mathbf{G} and has determinant value equal to one. From (12) and (13) we get that

$$\mathbf{M}^T \tilde{\mathbf{A}}_1 \mathbf{M} \mathbf{Q}_1^{-1} = \mathbf{M}^T \tilde{\mathbf{A}}_2 \mathbf{M} \mathbf{Q}_2^{-1}, \quad (14)$$

which gives

$$\tilde{\mathbf{A}}_2^{-1} \tilde{\mathbf{A}}_1 \mathbf{M} = \mathbf{M} \mathbf{Q}_2^{-1} \mathbf{Q}_1. \quad (15)$$

Let $\mathbf{P}_A = \tilde{\mathbf{A}}_2^{-1} \tilde{\mathbf{A}}_1$ and $\mathbf{P}_Q = \mathbf{Q}_2^{-1} \mathbf{Q}_1$. Thus, we have

$$\mathbf{P}_A \mathbf{M} = \mathbf{M} \mathbf{P}_Q. \quad (16)$$

Equation (16) which results from a pair of conics and their corresponding projections can be rewritten as

$$\mathbf{F}_{12} \mathbf{m} = 0, \quad (17)$$

where \mathbf{m} is a 9×1 vector containing the elements of \mathbf{M} and \mathbf{F}_{12} is a 9×9 matrix obtained from the elements of \mathbf{P}_A and \mathbf{P}_Q .

In the case of more than two conics, \mathbf{F} is formed by stacking matrices \mathbf{F}_{ij} arising from the matrix equation relating conics i and j and their projections. All different ordered pair are considered resulting in an \mathbf{F} matrix of $9n(n-1)$ rows and 9 columns. When \mathbf{F} has a rank equal to 8, \mathbf{m} is directly obtained. It is then rescaled such that elements of the first vector of the rotation matrix form a unit vector, and the sign is determined by ensuring that the object lies between the c-arm source and the image plane. However, in practice, \mathbf{F} is often of full rank due to the errors in the conic coefficients. In this case, in order to avoid a trivial solution for the system (17), we can write \mathbf{F} as $\mathbf{F} = \mathbf{F}_0 + \mathbf{E}$ where \mathbf{F}_0 is the exact rank-deficient matrix that we would get had there been no errors and \mathbf{E} is an error matrix.

Equation (17) now is an application of the total least squares problem where the right hand side is a zero vector. The idea as in [9, 10] is to find a rank-deficient least squares estimate of \mathbf{F}_0 by finding an error matrix \mathbf{E} with minimum Frobenius norm that lowers the rank of \mathbf{F} , i.e.

$$\min \|\mathbf{F} - \mathbf{F}_0\|_F \quad \text{s.t.} \quad \text{rank}(\mathbf{F}_0) = 8. \quad (18)$$

By the Eckart-Young-Mirsky theorem [9,10], we have

$$\mathbf{F}_0 = \sum_{i=1}^8 s_i \mathbf{u}_i \mathbf{v}_i^T, \quad (19)$$

where

$$\mathbf{F} = \mathbf{U} \mathbf{S} \mathbf{V}^T = \sum_{i=1}^9 s_i \mathbf{u}_i \mathbf{v}_i^T \quad (20)$$

is the singular value decomposition (SVD) of \mathbf{F} with singular values $s_1 \geq s_2 \geq \dots \geq s_9$. Now the singular vector corresponding to the least singular value of \mathbf{F} is used to estimate the vector \mathbf{m} that solves (17). Scaling and sign are determined as stated above for the case when \mathbf{F} has rank equal to 8.

2.3. Weighting and Orthonormalization

Although this technique with multiple conic pairs has been shown to outperform the homography estimation results obtained using points only and a single pair of conics, the quality of pose estimates degrades dramatically even with a small amount of segmentation errors (which is inevitable according to the existing ellipse segmentation techniques). In this paper we apply three modifications to the algorithm described in the previous section. With such changes, the performance of this algorithm can be improved in the presence of noise beyond what has been achieved by the original algorithm. First, we incorporate a weighting

scheme that enables us to take the effect of errors in the input data into consideration while solving the homogeneous total least squares (TLS) problem in (17); next, we find a rotation matrix that is a best orthonormal approximation to the one retrieved by the first step, and finally we refine all the pose parameters including the translation vector through a constrained optimization that minimizes the norm of the right hand side of the weighted system of equations and enforces orthonormality conditions on the rotation.

2.3.1. Weighting—With the overdetermined system of equations reached in (17), our simulations showed - in agreement with the literature - that in order to achieve reasonable pose estimates in the presence of realistic amounts of noise, the consequences of the error sources need to be inherently taken into account in the unified framework of homography estimation [10]. This is attained by giving different credence to each of the equations stacked in \mathbf{F} , a process termed equilibration, which relies on the assumption that the error matrix \mathbf{E} is not of i.i.d structure. This process essentially replaces the metric in (18) by

$$\min \|\mathbf{W}_L(\mathbf{F} - \mathbf{F}_0)\mathbf{W}_R\|_F, \quad (21)$$

where \mathbf{W}_L and \mathbf{W}_R are suitably chosen non-singular weight matrices that can be used to handle errors in the rows and columns of \mathbf{F} respectively [11]. For our system of equations, we use a left-hand equilibration approach in which - instead of using \mathbf{F} directly - we pursue finding a rank-deficient approximation of \mathbf{F} after scaling its rows using weights related to the variations that happen in its elements. In this case we replace \mathbf{F} in (17) by $\mathbf{W}_L\mathbf{F}$, where \mathbf{W}_L is a diagonal matrix with i^{th} diagonal entry $1/\sigma_i^2$. To find estimates for each of the σ_i^2 's, one possibility is to average all the variances of the entries of the i^{th} row of \mathbf{E} after simulating the segmentation errors that might have occurred for a given observed image. In practice, one can use the equations of the observed ellipses, together with an assumed error model of the noise that has caused the system to be full rank and simulate such error numerous times, thus generating many such error matrices \mathbf{E} . Applying the SVD to the matrix $\mathbf{W}_L\mathbf{F}$ yields its null vector \mathbf{m} , which is the singular vector corresponding to its least singular value. Again, \mathbf{m} is then rescaled and its sign is determined as denoted above.

2.3.2. Enforcing orthonormality constraints for the rotation—The estimated parameters that we get from the solution to the TLS system is essentially a general planar homography and due to the presence of noise in the observations, orthonormality of the rotation matrix of the pose in our case is not ensured. So, another SVD operation is performed on the first two columns of the rotation and a best approximate is found by multiplying the matrices of the left and right singular vectors by a matrix containing the first two columns of a 3×3 identity matrix. The third column of the rotation is then calculated as the cross product of those two columns.

2.3.3. Final Optimization—For a final tuning of the pose - including also the translational element of the pose - retrieved in the previous steps, we solve a constrained optimization problem that essentially maintains orthonormality gained in the previous step and simultaneously finds an optimal pose for our problem.

$$\begin{aligned} \min \|\mathbf{W}_L\mathbf{F}\mathbf{m}\|^2 \quad \text{s.t.} \quad & \mathbf{m}^T\mathbf{B}_1\mathbf{m}=0, \\ & \mathbf{m}^T\mathbf{B}_2\mathbf{m}=1, \\ & \text{and } \mathbf{m}^T\mathbf{B}_3\mathbf{m}=1. \end{aligned} \quad (22)$$

where \mathbf{B}_1 is the matrix corresponding to the orthogonality condition of the first two columns of the rotation matrix and \mathbf{B}_2 and \mathbf{B}_3 are the matrices corresponding to enforcing each of the first and second columns to have unit norm. Eventually, the final pose estimate (consisting of both rotation and translation) still satisfies a near-zero right hand side for (17).

3. Experiments and Results

3.1. Sensitivity analysis

Simulation studies were conducted to examine the effect of several factors on the accuracy of the estimated pose.

3.1.1. Sensitivity to segmentation errors—Simulated planar sets of 3 ellipses were used (see Fig. 1(b)). In order to simulate ellipse segmentation errors, we uniformly sampled 50 points on each projected ellipse, added Gaussian noise to the sample points, then fitted a new ellipse to the points using the method in [12]. The variance of the Gaussian was increased from 0.5 to 2 pixels squared in increments of 0.5. The focal length was set to 1 m and the pixel size used for our experiments was 0.44 mm for each of the u and v directions. For each error level, 700 experiments with 7 different poses were simulated; the mean and standard deviation of rotation and translation errors in the pose are shown in Fig. 2(a). In order to demonstrate the effect of weighting described in Section (2.3.1), we present results analogous to the ones in Fig. 2(a) that were obtained by running our code on the same datasets before modifying the algorithm to incorporate the weighting step. The results presented in Fig. 2(b) were attained by directly using the unweighted system (17). It is to be noted that in these experiments the rotation orthonormalization and the final optimization step were both performed as before.

3.1.2. Sensitivity to the size of ellipses—In order to test the effect of the size of ellipses used for the fiducial, the exact same simulations were done using a smaller pattern, i.e. with ellipses whose semi-major and semi-minor axes are half of those in the pattern used to generate the results in Fig. 2(a). The relative positioning and orientation were the same. The results for these simulations (based on a total of 2800 experiments) are shown in Fig. 2(c) and again show degradation in accuracy as the level of segmentation error increases. These results show that for a given configuration, the larger we can make the pattern, the better for pose accuracy.

3.1.3. Sensitivity to calibration errors—To assess the effect of miscalibration of the c-arm, we generated a realistic simulation for possible errors in the position of the focal spot. The magnitude of the error was increased from 0 to 15 mm in increments of 1 mm. During simulation, the fact that the uncertainty in the focal length is much larger than that in the image plane was taken into consideration and the direction of the in-plane drift of the principal point was uniformly sampled over all angles. For this experiment, segmentation noise has not been taken into account and therefore the system (17) has a unique solution without need for the weighting matrix \mathbf{W}_L . Experiments were done as previously described using the same set of three ellipses and the results are shown in Fig. 2(d). The algorithm proves to be robust to calibration errors upto 1 cm.

4. Conclusion

We proposed the use of a set of coplanar ellipses for c-arm pose estimation to assist in image-guided procedures. It has been shown in the literature on 3D reconstruction from 2D fluoroscopic images that reasonably accurate pose estimates for the c-arm (within a few degrees for the rotation and several mm for the translation) are sufficient for most clinical

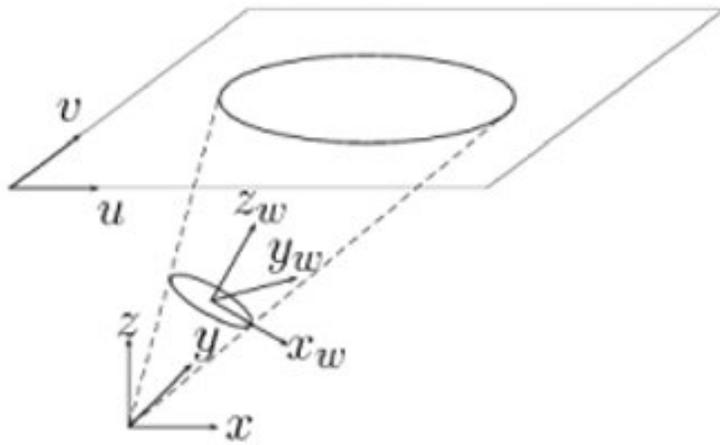
purposes. We showed using preliminary simulation studies that even with only three coplanar ellipses, we can attain a tracking accuracy within 0.5° and 0.9 mm for segmentation error variance up to 2 pixels squared and we are currently pursuing to apply this approach for precisely machined mechanical phantoms. Practical issues related to the specific fabrication parameters are still ongoing research issues. The presence of ellipses in the field of view must be considered and understood relative to the surgical procedure as well. In summary, this is a first step towards a simple clinically amiable tracker that may spare the need for proper positioning, mounting and yet provides a wide field of view for image capture in the OR. It provides proof of concept and encourages further investigation of this approach.

Acknowledgments

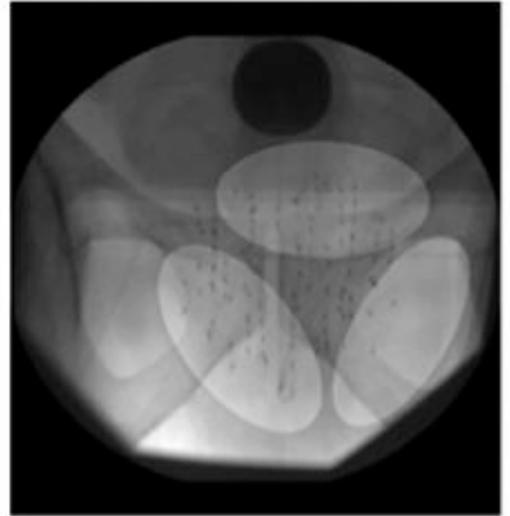
This work has been supported by NIH/NCI 2R44CA099374, and a National Science Foundation Graduate Research Fellowship.

References

1. Hofstetter R, Slomczykowski M, Sati M, Nolte LP. Fluoroscopy as an imaging means for computer-assisted surgical navigation. *Computer Aided Surgery* 1999;4(no. 2)
2. Yaniv Z, Joskowicz L. Precise robot-assisted guide positioning for distal locking of intramedullary nails. *IEEE Transactions on Medical Imaging* 2005;24(no. 5):624–635. [PubMed: 15889550]
3. Burkhardt D, Jain A, Fichtinger G. A cheap and easy method for 3d c-arm reconstruction using elliptic curves in. *Proceedings of SPIE* 2007;6509:65090B.
4. Chintalapani, G.; Jain, AK.; Burkhardt, DH.; Prince, JL.; Fichtinger, G. Ctrec: C-arm tracking and reconstruction using elliptic curves. *Computer Vision and Pattern Recognition Workshops, 2008. CVPR Workshops 2008. IEEE Computer Society Conference on; 2008; p. 1-7.*
5. De Ma S. Conics-based stereo, motion estimation, and pose determination. *International Journal of Computer Vision* 1993;10(no. 1):7–25.
6. Forsyth D, Mundy JL, Zisserman A, Coelho C, Heller A, Rothwell C. Invariant descriptors for 3 d object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1991;13(no. 10):971–991.
7. Sugimoto A. A linear algorithm for computing the homography from conics in correspondence. *Journal of Mathematical Imaging and Vision* 2000;13(no. 2):115–130.
8. Kannala, J.; Salo, M.; Heikkila, J. Algorithms for computing a planar homography from conics in correspondence. *Proc. the 16th British Machine Vision Conference (BMVC; 2006; p. 77-86. Citeseer*
9. Golub GH, Hoffman A, Stewart GW. A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra Applic* 1987;88:317–328.
10. Muhlich M, Mester R. The role of total least squares in motion analysis. *Lecture Notes in Computer Science* 1998;1407:305–321.
11. Muhlich M, Mester R. A considerable improvement in non-iterative homography estimation using tls and equilibration. *Pattern Recognition Letters* 2001;22(no. 11):1181–1189.
12. Fitzgibbon A, Pilu M, Fisher RB. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1999;21(no. 5):476–480.



(a)



(b)

Fig. 1.
 (a) Different coordinate systems involved in the projection of a conic. (b) X-ray image including ellipse projections.

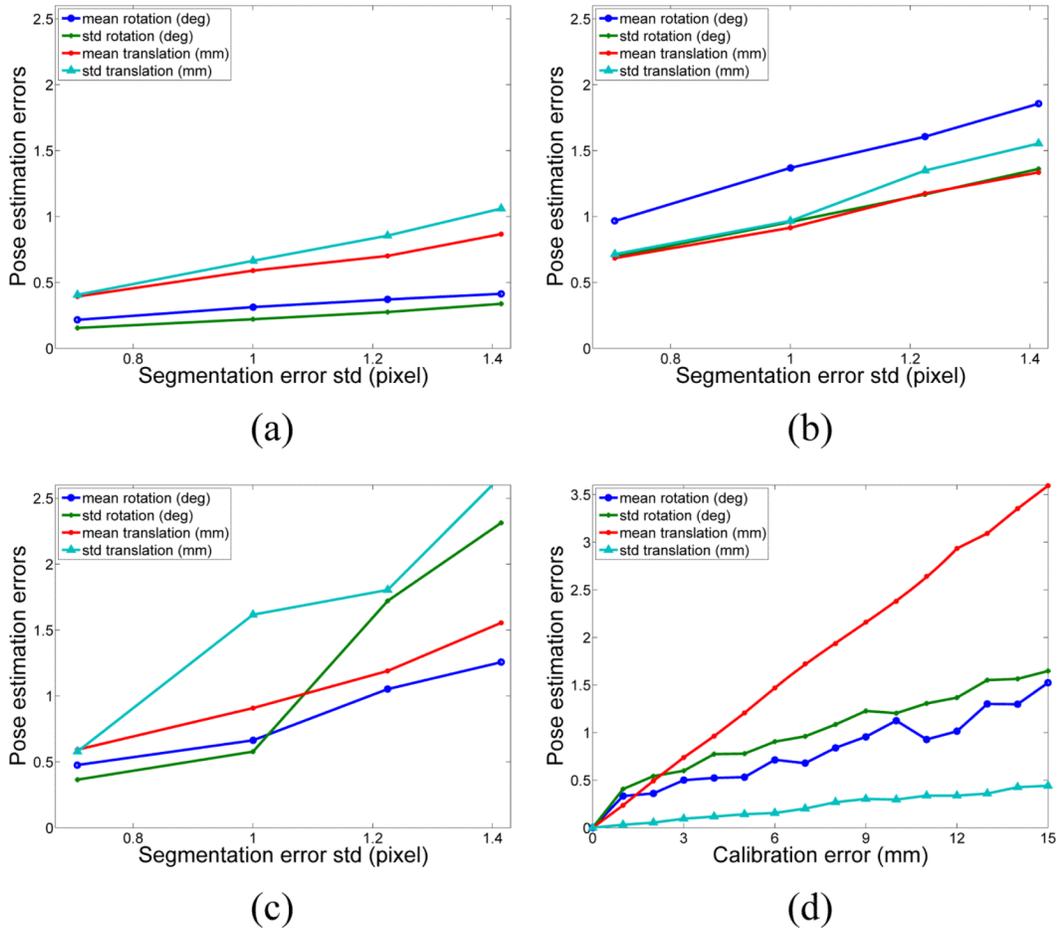


Fig. 2. Pose estimation errors. (a) Sensitivity to segmentation errors (Large Pattern-Weighted results). (b) Sensitivity to segmentation errors (Large Pattern-Unweighted results). (c) Sensitivity to segmentation errors (Small Pattern-Weighted results). (d) Sensitivity to calibration errors (Large Pattern-Weighted results).