Cross Modality Registration of Video and Magnetic Tracker Data for 3D Appearance and Structure Modeling

Dusty Sargent[a], Chao-I Chen[b], Yuan-Fang Wang[b]

[a]STI Medical Systems, 733 Bishop Street, Honolulu, HI, USA 96813
[b]Dept. of Computer Science, University of California, Santa Barbara, CA, USA 93106

ABSTRACT

The paper reports a fully-automated, cross-modality sensor data registration scheme between video and magnetic tracker data. This registration scheme is intended for use in computerized imaging systems to model the appearance, structure, and dimension of human anatomy in three dimensions (3D) from endoscopic videos, particularly colonoscopic videos, for cancer research and clinical practices. The proposed cross-modality calibration procedure operates this way: Before a colonoscopic procedure, the surgeon inserts a magnetic tracker into the working channel of the endoscope or otherwise fixes the tracker's position on the scope. The surgeon then maneuvers the scope-tracker assembly to view a checkerboard calibration pattern from a few different viewpoints for a few seconds. The calibration procedure is then completed, and the relative pose (translation and rotation) between the reference frames of the magnetic tracker and the scope is determined. During the colonoscopic procedure, the readings from the magnetic tracker are used to automatically deduce the pose (both position and orientation) of the scope's reference frame over time, without complicated image analysis. Knowing the scope movement over time then allows us to infer the 3D appearance and structure of the organs and tissues in the scene. While there are other well-established mechanisms for inferring the movement of the camera (scope) from images, they are often sensitive to mistakes in image analysis, error accumulation, and structure deformation. The proposed method using a magnetic tracker to establish the camera motion parameters thus provides a robust and efficient alternative for 3D model construction. Furthermore, the calibration procedure does not require special training nor use expensive calibration equipment (except for a camera calibration pattern—a checkerboard pattern—that can be printed on any laser or inkjet printer).

Keywords: Modeling, registration, calibration, magnetic tracker

1. DESCRIPTION OF PURPOSE

Research into computerized modeling of the appearance, structure, and dimension of malignant tissue growth, tumors, and polyps in colonoscopy can be of significant clinical value. We are currently developing such a modeling framework.[1] Sample model building results using images from real colon exams are shown in Fig. 1. Two pairs of images (left column) are used in the analysis to construct two 3D profiles (middle column) that are then merged and smoothed (right column) in Fig. 1(a). Sample novel views rendered based on the constructed computer model are depicted in Fig. 1(b). These images allow the surgeon to inspect the anatomy from many novel viewpoints and also obtain precise dimension measurements of anatomical features of interest. This paper discusses an alternative formulation of a critical step in our model building process that significantly improves the robustness and efficiency of the modeling procedure.

The fidelity of the structural description from the analysis depends on the accuracy of the estimated camera motion parameters. Unfortunately, the most error-prone step in the whole modeling process is in inferring the camera's motion parameters.[2] Video-only inference process often suffers if the image quality is poor, lighting is uneven or weak, or motion blur causes blotting and smear of image appearance. An additional sensing modality, such as a magnetic tracker that reports the six DOFs of the camera movement, can provide corroborating evidence on or perform an independent determination of the camera motion parameters. Such redundancy and cross-validation thus provide a more robust and fail-safe solution. This is the focus of the reported research.

Further author information: (Send correspondence to Dusty Sargent)
Dusty Sargent: E-mail: dsargent@sti-hawaii.com

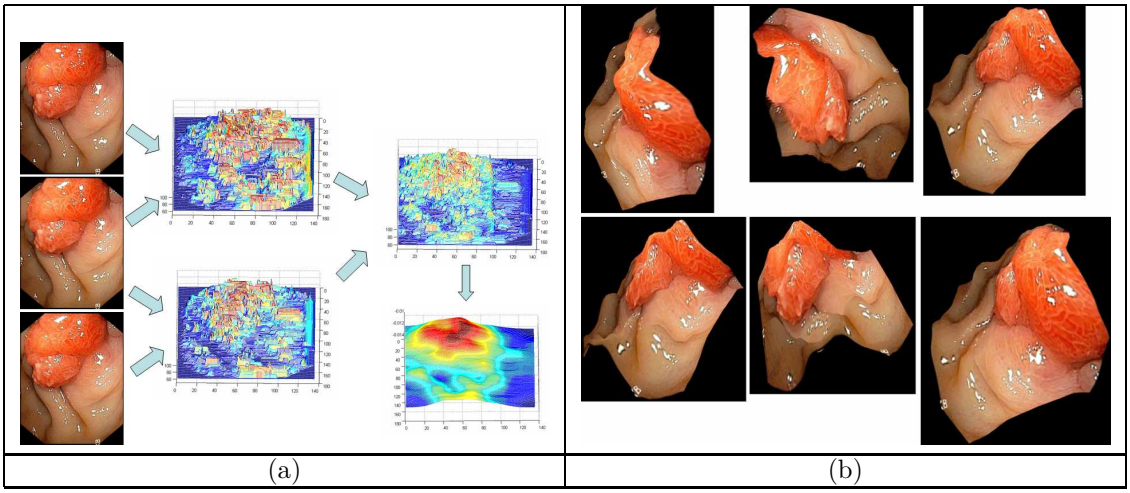


Figure 1. (a) Stereo matching, depth inference, and partial model registration results, and (b) novel view rendering using models constructed in (a).

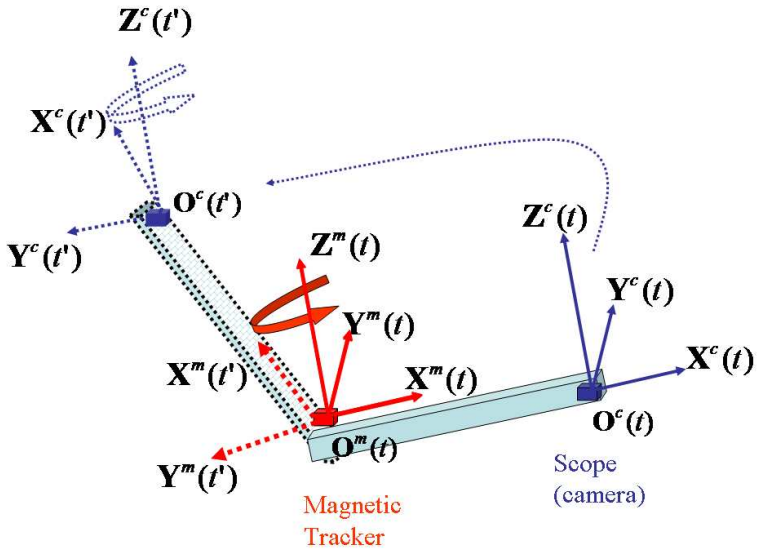


Figure 2. Pose and motion of the tracker are not the same as those of the scope.

2. METHOD

We address a challenging aspect of registering magnetic trackers with video sensors when *the scope and the tracker are in close proximity but are not collocated (e.g., the tracker is inside the working channel of an endoscope), and they move in unison.* The important observation is that the magnetic tracker reports the position and orientation of the tracker's tip, which are *not* the same as those of the scope. Furthermore, because the tracker's and the camera's frames are not collocated, the motion experienced and reported by the magnetic tracker is *not* the same to that experienced by the camera, even when the two move congruently.

This distinction is subtle but important, as it exerts a significant influence on the 3D modeling accuracy. Recall that the tracker is inserted into the scope's working channel without elaborate control and careful calibration, and hence, the relative pose of the two coordinate systems is unknown and not necessarily aligned properly. In Fig. 2, we show a simplified configuration of such a scope-tracker assembly where the coordinate frame of the magnetic tracker (superscript m) and that of the camera (superscript c) are displaced by an (unknown and uncalibrated) distance, but are otherwise aligned. Now if a rotation is executed about the Z axis of the magnetic tracker, the movement experienced by the camera from time t to t' includes the same rotation plus a

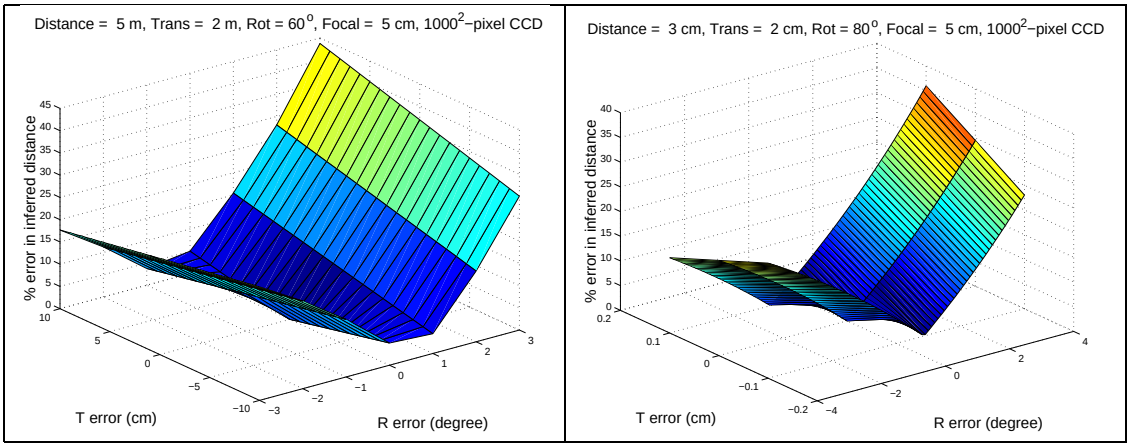


Figure 3. Errors in camera pose on the accuracy of the inferred 3D depth for both far-field, large-motion and near-field, small-motion cases.

non-zero (and unknown) translation. The additional translation experienced by the camera is *not* reported by the magnetic tracker—even when the two are rigidly attached and move in unison. The situation can become significantly more complicated if the assembly undergoes a general 3D motion and if the two coordinate frames are not aligned.

To recover 3D depth, we perform triangulation from the corresponding feature positions in two frames.[2,3] Triangulation requires the knowledge of the relative pose between the two frames, or the movement of the scope. However, the movement experienced by the scope is not the same reported by the tracker. We will then use the wrong scope movement in the analysis if the reported tracker movement is assumed.

One might wonder how much degradation this discrepancy might cause on the modeling accuracy. In Fig. 3, we provide the answer. We assume a simple side-by-side camera movement, using a camera with a focal length of 5cm and a CCD resolution of one-million pixels. We tested both a far-field, large-motion case (a movement of 2m between two shots, and an object about 5m away) and a near-field, small-motion case (a movement of 2cm between two shots, and an object about 3cm away). We added random noise to perturb the translational (up to 5%) and rotational (up to 4^o) components of the camera motion to simulate the error in using the magnetic tracker's pose readings as the camera's pose. We calculate, for each error setting in the camera motion parameters, how much the recovered depth (through triangulation) deviates from the ground truth (assuming no error in the feature position and correspondence). As can be seen the error in the recovered depth can be significant, rising to about 40% for 5% translation error and 4^o rotation error. Hence, it is of paramount importance to establish the correct camera motion parameters. *Feigning the reported tracker pose as the desired camera pose is* **not** *an acceptable solution* (Section 3 validates this claim).

Instead, we have developed a cross-modality registration procedure that is efficient, robust, and accurate. The registration procedure establishes the relative pose between the magnetic tracker and the endoscope in an assembly—*all without tailored hardware, time-consuming procedures, and special training of the surgeon.* Once the relative pose between the scope and the tracker has been calibrated, the movement of the scope can be inferred reliably in real time—simply offsetting the tracker's readings by the calibrated relative pose between the tracker and the scope.

Our calibration routine employs a magnetic tracker and its base station, an endoscope, and a grid or a checkerboard pattern which serves as the "base station" of the scope. Fig. 5 depicts the general configuration of the reference frames of the tracker base (superscript b), the magnetic tracker (superscript m), the camera or the scope (superscript c), and the calibration grid (superscript g). Each reference frame is specified by its origin and the directions of its X, Y, and Z axes. Two of the reference frames (associated with the tracker base and the calibration grid) are static while the other two (associated with the magnetic tracker and the camera) are dynamic, but move in unison.

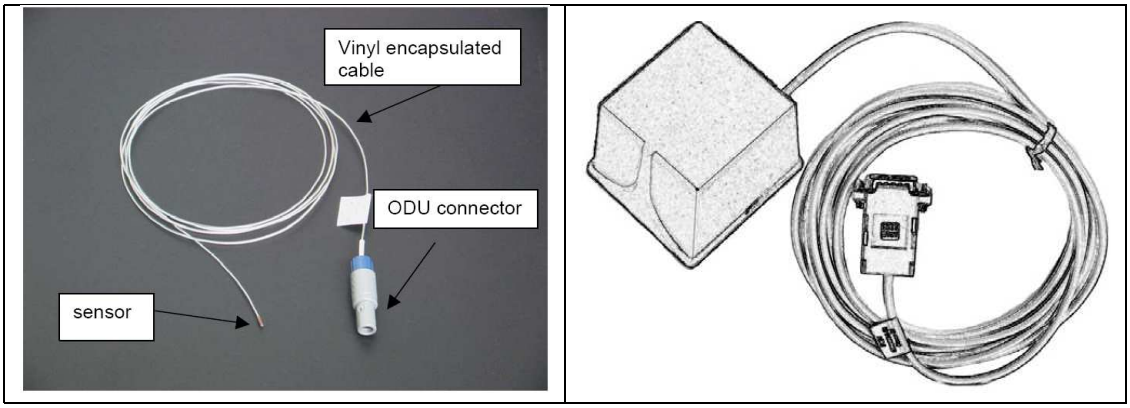


Figure 4. Left: MicroBird tracker from Ascension Corp. Right: tracker base.

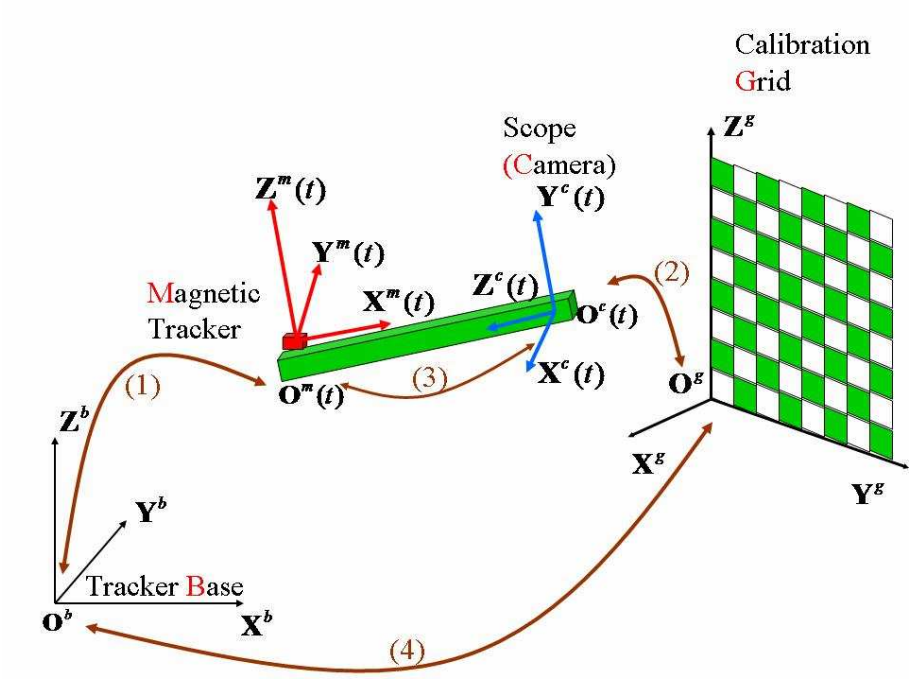


Figure 5. Cross-modality calibration configuration

Four transformations (translation and rotation)—two known and two unknown—between these four coordinate systems are important (Fig. 5): (1) The transformation between the tracker base and the magnetic tracker (known): The six-dimensional relative pose (three translations and three rotations) is reported by electromagnetic sensing. (2) The transformation between the calibration grid and the scope (known): The relative pose between the two is derived using a number of well established camera calibration algorithms.[2–5] (3) The transformation between the magnetic tracker and the scope (unknown): This is the relative pose that the calibration algorithm attempts to recover. Knowing this pose allows us to derive the camera pose from the tracker readings by a proper offset. (4) The transformation between the tracker base and the calibration grid (unknown): We do not require these two systems to be positioned in any specific manner. So the transformation between the two base systems are not assumed known.

The key idea of the calibration procedure is this: We can represent the pose of the scope in its base system (the calibration grid) in two ways: The first way is to compute the pose using standard camera calibration algorithms ((2) above). The second way is to express the pose by a circuitous relation of going from the scope to the tracker ((3) above), then from the tracker to the tracker base ((1) above), and finally from the tracker base to the calibration grid ((4) above). The first way contains no unknown, while the second way contains two unknowns, i.e., the relative pose between the camera and the tracker ((3) above) and that between the tracker base and the calibration grid ((4) above). If we equate the pose of the scope in the calibration grid system derived using the two methods, we will be setting up some constraints on these two unknown relative poses. An overly simplified analogy is that (2) = (3) + (1) + (4).

Certainly, it is not possible to solve for two unknowns with only one constraint. However, one recalls that the relative pose between the calibration grid and the tracker base ((4) above) is constant because both coordinate systems are stationary. While both the camera and the tracker are moving in space, their relative pose ((3) above) is again constant (i.e., the whole scope-tracker assembly moves rigidly). Hence, if we record multiple, synchronized magnetic tracker readings and photos of the calibration grid as the scope-tracker assembly moves in space, we can set up multiple constraints of the type (2) = (3) + (1) + (4) to solve for both (3) and (4). Theoretically, one can prove that five synchronized images and tracker readings are needed to solve for the unknown relative poses. In reality, we will use a lot more than five to ensure robustness and accuracy. Given the video frame rate of 30 frames/second, just a few seconds of videos will provide hundreds of image frames and tracker readings for calibration.

Furthermore, we perform an additional step of nonlinear optimization to improve the accuracy of the scope-tracker transformation. The camera motion between any two frames of video depends on only the tracker's position and orientation at each frame, and the transformation from the tracker to the scope. We also have the pixel coordinates of the grid junctions in each frame to use as correspondences, and the intrinsic and distortion parameters from the camera calibration process. We can therefore obtain the camera motion (rotation and translation) between each pair of frames and use the epipolar constraint to form an objective function to be minimized as follows:

$$e = \sum_{1 \le k,l \le n} \sum_{1 \le i \le m_{kl}} \left([\mathbf{p}_i^{(l)^T}, 1][\mathbf{T}_k^l]_\times \mathbf{R}_k^l \begin{bmatrix} \mathbf{p}_i^{(k)} \\ 1 \end{bmatrix} \right)^2 \quad (1)$$

where n is the number of image frames used, $\mathbf{T}_k^l$ and $\mathbf{R}_k^l$ are the translational and rotational camera motion from frame k to frame l, m_{kl} is the number of corresponding features between frames k and l, and $\mathbf{p}_i^{(k)}$ and $\mathbf{p}_i^{(l)}$ are the positions of the i-th corresponding features in frames k and l, respectively. No matter how many frames are used, there are only 6 variables representing the scope-tracker transformation: 3 for rotation and 3 for translation. While the formulation in Eq. 1 is nonlinear, we can use the output of the first step as an educated initial guess. In particular, we use the trust-region methods, which combine gradient descent and Newton's method,[2,6–10] to guarantee that we find a local minimum in the objective function.

3. EXPERIMENTAL RESULTS

Here we show results of calibration using real video and tracker data. These camera calibration algorithms employ a standard calibration pattern (a checkerboard or a grid pattern in Fig. 6). The camera views the pattern from a number of arbitrary poses (sample images are shown on the left three columns of Fig. 6). Based

on the corresponding grid junction locations in these images (red squares in the left three columns of Fig. 6 which are *automatically* detected and matched given the four color anchors), the calibration algorithm recovers the camera poses relative to the coordinate frame of the calibration grid as shown on the right of Fig. 7. The procedures can handle large fisheye camera distortion (bottom row in Fig. 6 and Fig. 7).

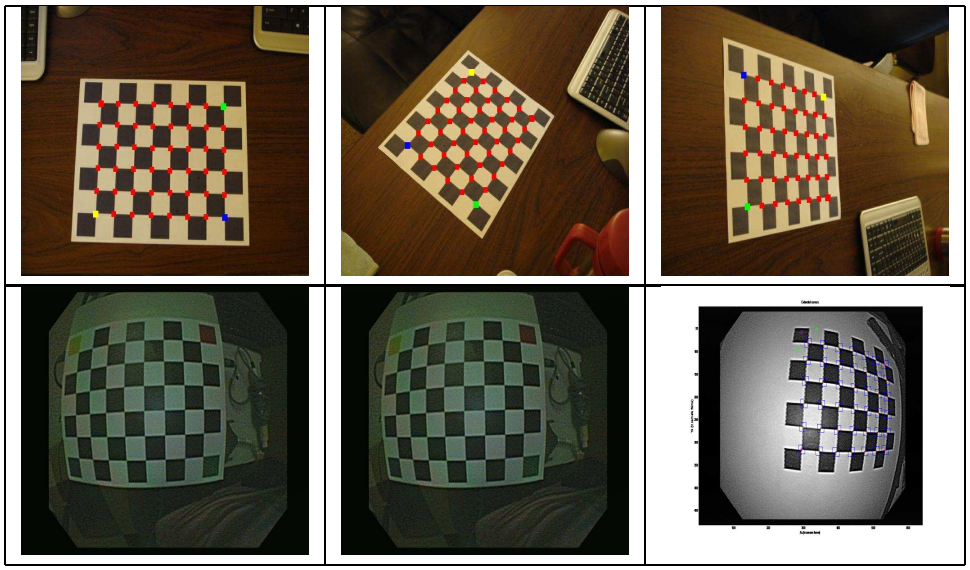


Figure 6. Sample images of the calibration grid from different view points. Top row: an example of small image distortion. Bottom row: an example of large fisheye distortion.

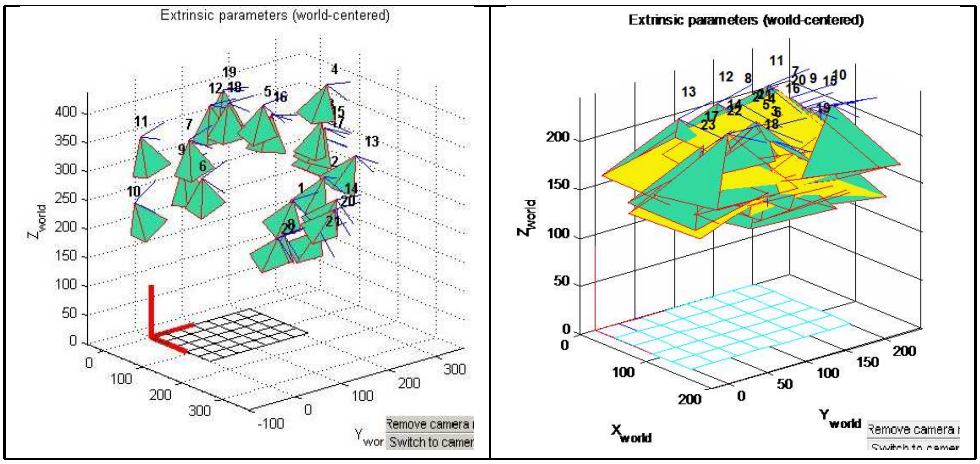


Figure 7. The inferred camera poses relative to the reference frame of the calibration grid in Fig. 6. The green cones represent the camera poses in space where the cone apex is the camera's optical center and the base represents the image plane. The red lines are the X, Y, and Z coordinate axes of the calibration grid. Left: Camera poses of the top row in Fig. 6. Right: Camera poses of the bottom row in Fig. 6.

How do we know if the calibration procedure is accurate? One way is to perform 3D modeling. Recall that to construct a 3D model by triangulating from corresponding image features requires accurate estimate of the camera motion parameters. If the estimation is way off, the 3D shape will become grotesquely distorted. As the camera motion parameters are now inferred from the magnetic tracker readings (which we assume are accurate) and the relative pose between the camera and the tracker (obtained from the calibration procedure), the calibration accuracy directly influences the modeling accuracy. Furthermore, during the calibration process we already gathered many images of the calibration grid pattern with junctions extracted and correspondences

identified in these images. Therefore, We can construct 3D models of the calibration pattern and calculate the reconstruction error (as we know the ground truth—the spacing between junctions and the fact that we have a repetitive planar pattern).

In Fig. 8(a) & (b), we show 3D models of the grid junctions in Fig. 6, rendered from multiple viewpoints in space. As can be seen, we correctly recover the regular, planar shape of the calibration grid. We also compute three quantitative error measures: planarity: the average deviation of a grid junction from the best-fitting plane, linearity: the average deviation of a grid junction from the best-fitting grid line, and orthogonality: the deviation of the angle of intersection of the best-fitting horizontal and vertical grid lines from 90^o. These errors are shown in the first row, with calibrated offset, of Table 1 (percentage error is computed with respect to a grid patten of roughly 24cm by 18cm, i.e., % error = absolute error/ grid line length). The calibration results demonstrate good planarity, linearity, orthogonality.

As mentioned before, feigning the reported tracker pose as the desired camera pose is not acceptable. This point is illustrated in Fig. 8(c) and the second row, without calibrated offset, in Table 1. As can be seen that the grid looks distorted and quantitatively the error measurements are much larger if the pose of the magnetic tracker is masqueraded as the desired camera pose, without the benefit of the proposed calibration algorithm.

4. CONCLUSIONS

We have developed a cross-modality registration procedure that is efficient, robust, and accurate. The registration procedure establishes the relative pose between the magnetic tracker and the endoscope in an assembly. The approach overcomes the difficulty of applying traditional computer-vision techniques for motion estimation in medical images—namely, sensitivity to mistakes in image analysis, error accumulation, and structure deformation. The calibration technique does not require special training of the surgeons, elaborate procedures, or expensive equipment, and hence, should have a low threshold of adoption in real surgery.

REFERENCES

[1] Koppel, D., Chen, C.-I., Wang, Y.-F., Lee, H., Gu, J., Poirson, A., and Wolters, R., "Toward automated model building from video in computer assisted diagnoses in colonoscopy," in [*Proceedings of the SPIE Medical Imaging Conference*], (2007).

[2] Hartley, R. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press, Camridge, MA (2003).

[3] Xu, G. and Zhang, Z., [*Epipolar Geometry in Stereo, Motion and Object Recognition*], Kluwer Academic Publishers, The Netherlands (1996).

[4] Zhang, Z., "A Flexible New Technique for Camera Calibration," *IEEE Trans. Pattern Analy. Machine Intell.* **22**, 1330–4 (2000).

[5] Intel Corp., "http://www.intel.com/technology/computing/opencv."

[6] Demmel, J., [*Applied Numerical Linear Algebra*], SIAM (1997).

[7] Saad, Y., [*Iterative Methods for Linear Systems, 2nd Ed.*], SIAM (2003).

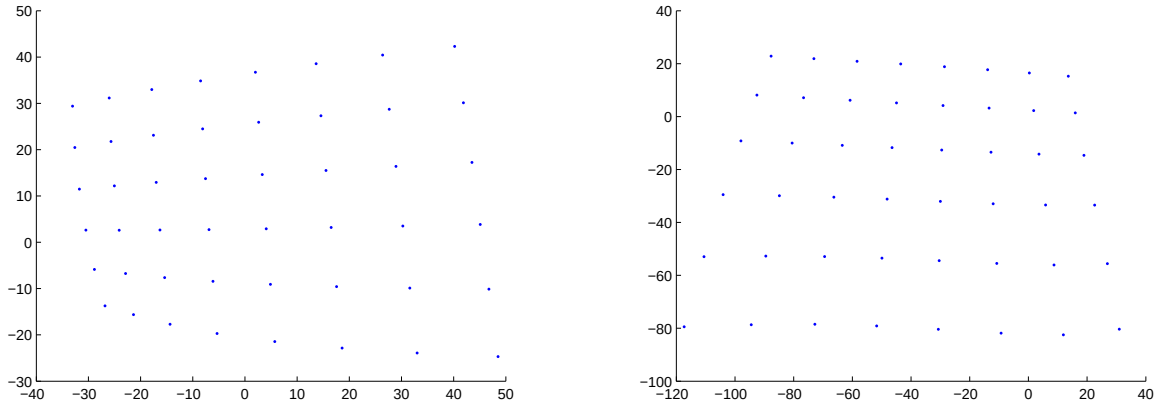
[8] Conn, A. R., Gould, N. I. M., and Toint, P. L., [*Trust-Region Methods (MPS-SIAM Series on Optimization)*], SIAM, Philadelphia, PA (2000).

[9] Dennis, J. E. and Mei, H. H. W., "Two new unconstrained optimization algorithms which use function and gradient values," *Journal of Optimization Theory and Application* **28**, 453–482 (August 1979).

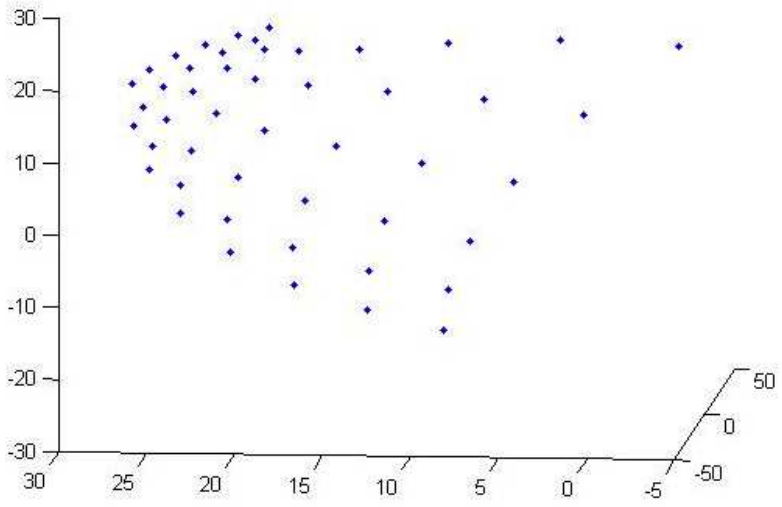
[10] Dennis, J. E. and Schnabel, R. B., [*Numerical Methods for Unconstrained Optimization and Nonlinear Equations*], SIAM, Philadelphia, PA (1996).

Table 1. Error in 3D model using calibrated offset from the proposed algorithm and without.

	Planarity		Linearity		Orthogonality
Tracker data	abs error(mm)	%error	abs error(mm)	%error	angle intersection from 90^o
with calibrated offset	1.0	0.4%	0.6	0.2%	0.4^o
without calibrated offset	3.8	1.6%	2.2	0.9%	32.4^o



(a) (b)



(c)

Figure 8. Reconstructed grid junction points in space. (a) & (b) with proper calibrated offset, and (c) without proper calibrated offset.