

Efficient Stereo Image Geometrical Reconstruction at Arbitrary Camera Settings from a Single Calibration

Songbai Ji^{1,2}, Xiaoyao Fan¹, David W. Roberts^{2,3}, and Keith D. Paulsen^{1,2,3}

¹ Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA

² Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA

³ Dartmouth Hitchcock Medical Center, Lebanon, NH 03756, USA

{songbai.ji, xiaoyao.fan, david.w.roberts,
keith.d.paulsen}@dartmouth.edu

Abstract. Camera calibration is central to obtaining a quantitative image-to-physical-space mapping from stereo images acquired in the operating room (OR). A practical challenge for cameras mounted to the operating microscope is maintenance of image calibration as the surgeon's field-of-view is repeatedly changed (in terms of zoom and focal settings) throughout a procedure. Here, we present an efficient method for sustaining a quantitative image-to-physical space relationship for arbitrary image acquisition settings (S) without the need for camera re-calibration. Essentially, we warp images acquired at S into the equivalent data acquired at a reference setting, S_0 , using deformation fields obtained with optical flow by successively imaging a simple phantom. Closed-form expressions for the distortions were derived from which 3D surface reconstruction was performed based on the single calibration at S_0 . The accuracy of the reconstructed surface was 1.05 mm and 0.59 mm along and perpendicular to the optical axis of the operating microscope on average, respectively, for six phantom image pairs, and was 1.26 mm and 0.71 mm for images acquired with a total of 47 arbitrary settings during three clinical cases. The technique is presented in the context of stereovision; however, it may also be applicable to other types of video image acquisitions (e.g., endoscope) because it does not rely on any *a priori* knowledge about the camera system itself, suggesting the method is likely of considerable significance.

1 Introduction

Camera images provide texture intensity from the surface of objects in the scene, and are an increasingly popular form of data in image-guided procedures such as neurosurgery [1]. Calibration is central to obtaining quantitative geometrical information from the camera system to project 2D image pixels into their 3D coordinates in physical space in the case of stereovision. Techniques for calibrating a camera system at fixed zoom and focal settings are well studied [2]. However, many cameras offer a wide range of zoom factors and focal lengths that can be arbitrarily varied to obtain an optimal view [3]; thus, maintenance of camera calibration becomes a practical challenge. Because these images depend on acquisition settings, recovering the camera calibration parameters efficiently and for an arbitrary setting is

essential for applications like stereovision in the operating room (OR) where the surgeon is repeatedly altering the field-of-view through the operating microscope.

Existing techniques for camera calibration at an arbitrary setting either actively re-calibrate at a given setting on-demand [3] or interpolate camera parameters via bivariate fitting by explicitly modeling each as a polynomial function of zoom and focal length based on data from a dense set of pre-calibrations [4, 5]. Although calibration at a given setting can be fully automated with an on-demand approach [2], repeatedly imaging an instrumented calibration target [3] is inconvenient and cumbersome in the OR. While interpolation of pre-determined camera parameters minimizes disruption of surgical workflow, a dense combination of zoom and focal length settings have to be calibrated (and re-calibrated for quality assurance and/or when camera extrinsic parameters are changed, e.g., from repositioning) which too adds to pre-operative activity and personnel time requirements. Consequently, suggestions of a fixed zoom and focus have been made to ensure optimal accuracy [5], but such restrictions significantly limit the effective OR use of camera systems.

In this study, we present a method to recover geometry from stereo images at arbitrary camera settings using a single calibration at a fixed (reference) zoom and focal setting. The approach is especially appealing for OR applications because it does not disrupt surgical workflow nor does it require tedious calibration at numerous zoom-focus combinations. The performance of the technique is evaluated on a physical phantom and in three clinical cases involving open cranial surgery with a microscope-mounted stereovision system. However, the general strategy appears to be applicable to other types of stereo/video images (e.g., endoscope).

2 Material and Methods

A custom-designed stereovision system consisting of two C-mount cameras (Flea2 model FL2G-50S5C-C, Point Grey Research Inc., Richmond, BC, Canada) was rigidly mounted to a Zeiss surgical microscope (OPMI[®] Pentero[™], Carl Zeiss, Inc., Oberkochen, Germany) through a binocular port [6]. The position and orientation of the microscope was available from a StealthStation[®] navigation system via StealthLink (Medtronic, Inc., Louisville, CO) through a rigidly-attached tracker. In addition, the microscope zoom, m , and focal length, f , are also directly available from StealthLink, which eliminates the need to manually record the acquisition settings.

The technical details of stereovision calibration at a single acquisition setting and subsequent surface reconstruction have been well studied [2]. Both the left (I_L) and right (I_R) camera images depend on image acquisition settings such as m and f . Conceptually, the following functional forms define the images acquired:

$$I_L = G_L(m, f), \text{ and } I_R = G_R(m, f). \quad (1)$$

For notational simplicity, we drop the subscripts throughout the rest of the paper when an image is not specifically associated with either the left (L) or right (R) camera. We also denote the image acquired at a set of reference settings, S_0 , as $I_0 = G(m_0, f_0)$, which represents the lowest magnification (m_0) and shortest focal length (f_0) that the microscope offers. The choice of S_0 is selected for convenience because of the ease with which the microscope can be returned to these settings;

however, a different set of reference settings or multiple settings could be utilized (see Discussion). An image obtained at an arbitrary setting, S , is referred to as a “deformed” image ($I = G(m, f)$). The “deformation field” relating the deformed image to the reference or “undeformed” image is found via optical flow (OF) motion-tracking, which has been well studied [7] and successfully employed in image-guided procedures [1] including stereovision in neurosurgery [6]. Essentially, our technique for stereovision reconstruction at S warps the deformed images into the reference image as if the data were acquired at S_0 using deformation/distortion fields obtained from a series of phantom images. Because stereo images at S_0 have been calibrated, the warped stereo images acquired at S can then be reconstructed with the same single calibration once the warping is complete.

2.1 Image Deformation due to the Change in Acquisition Settings

To determine image deformation due to the change in image acquisition settings, m and f , a phantom was created by printing squares in random positions and intensities on paper. The phantom was first imaged at the reference setting, S_0 , and then a series of images was acquired by successively changing either m or f (while maintaining the partnered parameter at its respective reference value). Image acquisitions at multiple m values at each f settings were unnecessary because the resulting 2D image deformation induced by a change in m was independent of f settings, at least for the Zeiss Pentero surgical microscope based on setting values from StealthLink.

Because the OF algorithm is designed to detect small displacements, deformation fields between images obtained from two adjacent m or f values (instead of relative to the reference values) were computed. The resulting displacement vectors were found to vary radially relative to the focal point along the optical axis (Fig. 1). Thus, a local cylindrical coordinate system was established with its origin at the focal point in order to fit the deformation field as a function of radial distance, r . Because the OF algorithm can produce artifacts especially in image corners with poor lighting conditions, regions near the boundary (<100 pixels) were excluded from the processing. A least-squares linear fitting was found to be sufficient to represent the deformation field (difference between the measured and recovered values was 0.06 pixels on average with a maximum value of 0.2 in the region used for fitting). This procedure yielded an analytical expression for the magnitude of radial displacement

$$F_m^{i \rightarrow (i-1)}(r) = k_m^{i \rightarrow (i-1)} r, \text{ and } F_f^{i \rightarrow (i-1)}(r) = k_f^{i \rightarrow (i-1)} r, \quad (2)$$

where $k_m^{i \rightarrow (i-1)}$ and $k_f^{i \rightarrow (i-1)}$ are linear scaling factors independently determined from fitting the deformation fields obtained from image pairs acquired at the i^{th} and $(i-1)^{\text{st}}$ setting of m and f , respectively (Fig. 1).

The following pseudo-algorithm summarizes the process of generating and fitting image deformation fields resulting from changes in m and f :

1. Set $f = f_0$, successively increase m from m_0 in small steps to acquire images at each setting. Compute deformation field between images obtained from two adjacent m values, and determine the corresponding scaling factor, $k_m^{i \rightarrow (i-1)}$;

2. Set $m = m_0$, successively increase f from f_0 in small steps to acquire images at each setting. Compute deformation field between images obtained from two adjacent f values, and determine the corresponding scaling factor, $k_f^{i \rightarrow (i-1)}$.

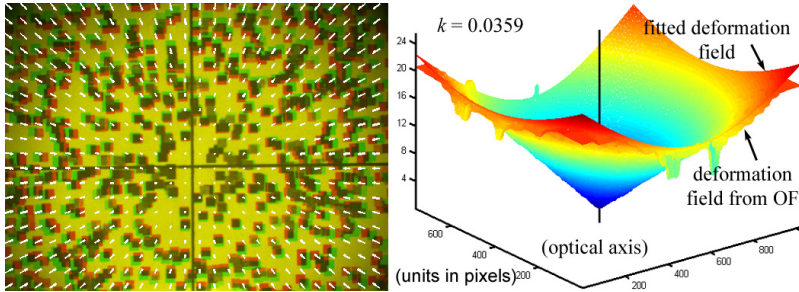


Fig. 1. Typical deformation field for $F_m^{i \rightarrow (i-1)}(r)$ (analogous for $F_f^{i \rightarrow (i-1)}(r)$, not shown). Left: overlays of undeformed (red) and deformed (green) images and the resulting displacement field (reduced lighting condition is evident in the corners). Right: magnitudes of radial displacements (measured and fitted) expressed in a cylindrical coordinate system with the optical axis focal point as the origin (unit in pixels). The corresponding linear scaling factor is also shown.

2.2 Image Warping into Reference Setting

Because the deformation scaling factors were obtained between two adjacent m and f values, tracking the position of each individual pixel in the current image (acquired at m or f ; in the local cylindrical coordinate system) through the chain of image deformations is required to warp it into the equivalent position acquired at m_0 or f_0 . For example, the corresponding position of a specific pixel in the image acquired at setting m^i , can readily be obtained in the image acquired at m^{i-1} as $r \times (1 + k_m^{i \rightarrow (i-1)})$. Following the chain of deformation scaling, the corresponding pixel location at m_0 , and analogously, at f_0 , is obtained with the closed-form equation

$$P_m^{i \rightarrow 0}(r) = r \prod (1 + k_m^{i \rightarrow (i-1)}), \text{ and } P_f^{i \rightarrow 0}(r) = r \prod (1 + k_f^{i \rightarrow (i-1)}). \quad (3)$$

The resulting ratios define the “pixel cumulative radial scaling (PCRS)” in the local cylindrical coordinate system with respect to m and f . Based on the set of m and f values at which phantom images were acquired, the corresponding characteristic curves can be found that define the image deformation behavior of the stereovision system (Fig. 2). These curves were further fit to a polynomial form in order to warp images acquired at arbitrary m or f values (ratios were constrained to 1.0 at the reference settings). A third-order polynomial was sufficient to produce differences with respect to the measured data that were less than 5×10^{-5} for both m and f . The same measurement and data-fitting schemes were applied to both the left and right camera images of the stereovision system. We found that the characteristic curves were virtually identical in the two cases (difference $< 10^{-3}$).

The following pseudo-algorithm summarizes the procedure to warp an image acquired at an arbitrary setting, S , into the reference setting, S_0 :

1. Use the fitted data, $P_m^{i \rightarrow 0}(r)$, to interpolate the PCRS for the given m ;
2. Re-position pixels and interpolate image as if acquired at (m_0, f) ;
3. Use the fitted data, $P_f^{i \rightarrow 0}(r)$, to interpolate the PCRS for the given f ;
4. Re-position pixels and interpolate image as if acquired at (m_0, f_0) .

After both the left and right camera images were warped into S_0 , calibration performed at S_0 is then used to reconstruct the 3D surface.

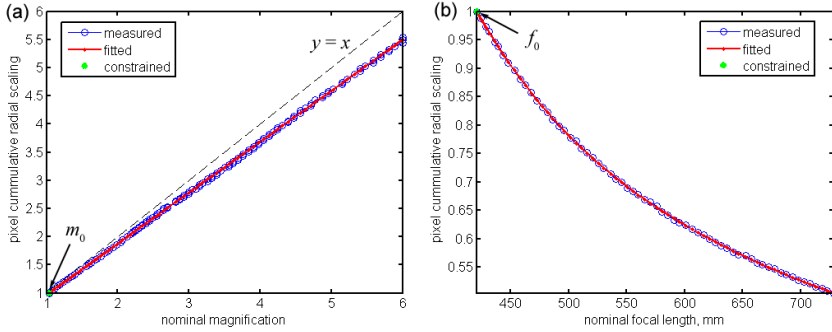


Fig. 2. Characteristic PCRS curves with respect to (a) magnification, m , and (b) microscope focal length, f , for the left camera. The characteristic curves for the right camera (not shown) were virtually identical (difference $< 10^{-3}$). The measured radial scaling ratios were fit to a third-order polynomial with ratios constrained to 1.0 at the reference points, m_0 and f_0 .

2.3 Data Analysis

Six combinations of (m, f) were arbitrarily selected to evaluate the performance of our technique on a phantom skull with hand-drawn feature lines. Images acquired with 47 arbitrary zoom and focal settings, S , during 3 epilepsy surgeries were also considered. For each S , a pair of stereo images of the exposed surface was acquired, warped, and reconstructed into a 3D surface. The “ground-truth” surface was also reconstructed using image pairs acquired at S_0 . For each object (phantom or cortical surface), the overlapping region of all reconstructed surfaces corresponded to the same physical surface, and allowed the reconstruction accuracy at S to be measured as the average surface nodal distances along (d_1) and perpendicular to (d_2) the optical axis of the microscope based on the triangulated surfaces. To compute the lateral distance (i.e., d_2), the two texture-encoded surfaces were projected onto the XY -plane in a local coordinate system (with the optical axis as the Z -axis), and an average displacement magnitude was obtained through OF motion tracking. Finally, the accuracy of the “ground-truth” surface reconstructed at S_0 was evaluated using independently sampled points acquired from a tracked stylus to report an average relative distance between two sets of homologous points (e.g., vessel intersections).

3 Results

Table 1 summarizes d_1 and d_2 for two arbitrary settings, S , from the phantom skull and from each of the 3 patient cases (m and f ranged 1.02–6.05 and 417.9–728.2 mm,

respectively). Based on all settings tested, the average accuracies of d_1 and d_2 were 1.05 ± 0.33 mm (range 0.59–1.4 mm) and 0.59 ± 0.32 mm (range 0.15–0.79 mm) for the phantom, and 1.26 ± 0.31 mm (range 0.61–1.90 mm) and 0.71 ± 0.39 mm (range 0.08–1.56 mm) for the patient cases, respectively. The average accuracy in terms of point-to-point distances between homologous features on the surfaces reconstructed at S_0 using independently sampled probe tip data was 1.57 ± 0.27 mm (based on a total of 16 data points from 3 patient cases). Fig. 3 illustrates the reconstructed surfaces from the phantom skull and cortical surfaces from the 3 patient cases using camera calibration at S_0 and the corresponding surfaces formed from images acquired at a selected arbitrary setting, S , by the warping technique described in Section 2.2. For comparison of reconstruction accuracy, independently sampled probe points from a tracked stylus are also shown together with their homologous features on the cortical surfaces reconstructed at S_0 .

Table 1. Summary of d_1 and d_2 (in mm) between surfaces reconstructed at S_0 and S at representative settings (m, f) for phantom and patient cases, as well as their average values (\bar{d}_1 and \bar{d}_2) from all settings tested (n is the number of settings tested for each case)

Parameter	Phantom	Phantom	Patient 1	Patient 1	Patient 2	Patient 2	Patient 3	Patient 3
M	1.20	4.30	3.02	4.43	2.71	2.16	3.65	6.03
F	417.9	501.7	472.2	545.2	520.9	420.2	521.1	728.2
d_1	0.59	1.40	1.08	1.32	1.65	0.82	1.38	1.54
d_2	0.55	0.35	0.74	0.66	0.10	0.20	0.63	0.84
\bar{d}_1	1.05 ± 0.33 ($n=6$)		1.13 ± 0.24 ($n=24$)		1.34 ± 0.39 ($n=11$)		1.44 ± 0.24 ($n=12$)	
\bar{d}_2	0.59 ± 0.32		0.90 ± 0.33		0.51 ± 0.44		0.48 ± 0.28	

4 Discussion and Conclusion

Efficient image reconstruction at an arbitrary camera setting is critical for effective deployment of intraoperative stereovision especially when the data is presented in the operating room (OR). We have described a simple, yet effective strategy for maintaining quantitative correspondence of camera image data with physical OR coordinates despite intraoperative changes in camera zoom and focal length settings based on a one-time collection of images of a planar phantom from which the deformation field is derived to warp subsequent images acquired at arbitrary settings. The resulting sub-millimeter to millimeter accuracy of the reconstructed surfaces at arbitrary settings relative to “ground-truth” especially in the clinical cases (maximum d_1 and d_2 error of 1.90 mm and 1.56 mm, respectively) was excellent, particularly since the magnitude of cortical surface pulsation, itself, can be up to 1 mm [6]. Since our approach does not require (repeatedly) imaging a calibration target, it is especially appealing for applications in the OR because it does not interrupt surgical workflow. Similarly to fitting individual camera parameters into bivariate functions of m and f [4, 5], image acquisitions of a phantom are still necessary with our approach (in order to derive 2D image deformation fields). However, instead of requiring as many as $p \times q$ individual settings (p and q are the number of discrete m and f settings needed

to calibrate, respectively; [4, 5]), the number of image acquisitions required with our method is limited to $p + q$ because the 2D image deformation induced by a change in m does not depend on f settings for the Zeiss Pentero surgical microscope. If such independence does not hold elsewhere, our method would similarly require $p \times q$ acquisitions. However, our strategy only needs to model 1D radial scaling (or 2D deformation at worst) instead of 11 camera parameters explicitly for every calibration. In addition, accurate interpolation of parameters (Fig. 2) can be obtained with sparse (m, f) pairs because of the smooth deformation field (Fig. 1). In contrast, conventional methods require dense (m, f) pairs because of the jagged/nonlinear parametric surfaces [4]. The simple radial scaling is sufficient to accurately characterize the induced image deformations, as expected from a pinhole camera model. The closed-form expressions for the PCRS curves (Eqn. 3) further simplify our approach because they allow image deformation to be easily interpolated and images to be efficiently warped to the reference setting for accurate 3D reconstruction. When a closed analytical form is otherwise not available (e.g., perhaps for a different camera system), a chain of *implicit* deformation fields can still be used to interpolate and warp the images, in which case the general concept of image warping to a reference setting for reconstruction would be applicable as well.

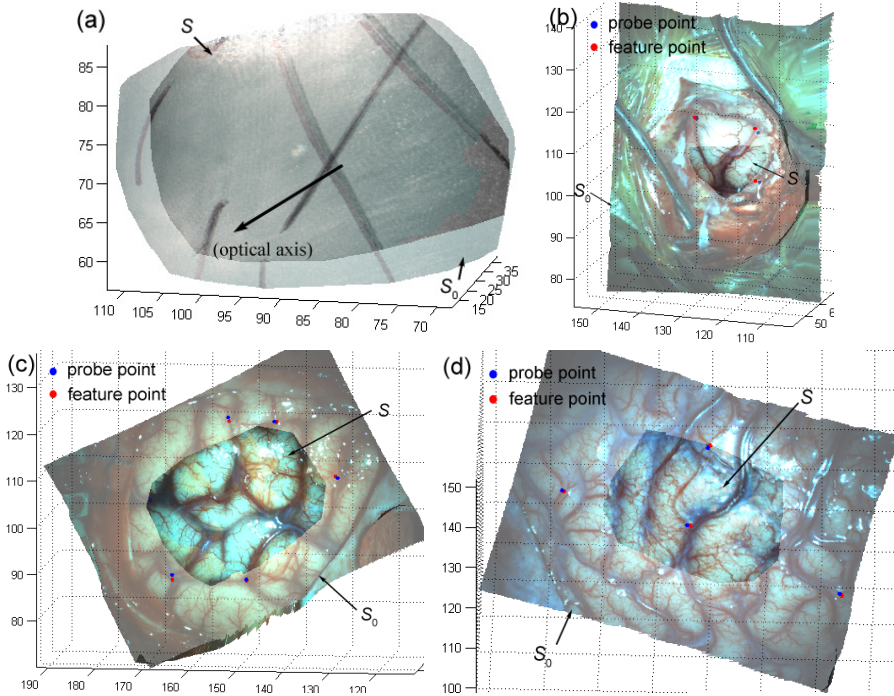


Fig. 3. Overlays of reconstructed surfaces using the reference setting S_0 and an arbitrary setting S for the phantom (a) and three patient cases (c–d) in MR image space (units in mm). Surfaces reconstructed at S (manually masked) through the warping technique virtually coincide with their counterparts reconstructed at S_0 . Independently sampled probe points from a tracked stylus are shown together with their homologous feature locations on the cortical surfaces.

We chose the lowest m and f values as the reference setting in this study because of the ease with which the microscope can be manually returned to this state. Because image reconstruction was performed at a single calibration setting, some loss of resolution or reduction in the field of view occurs when images are acquired at a higher m or f , respectively. However, our technique can be extended by calibrating the camera system at a small number of discrete reference settings, e.g., at higher m and/or f values. For each new reference setting, the PCRS curves can be re-established using the same set of phantom planar images without re-collection.

Essentially, our reconstruction framework treats the calibration process as a transfer function by directly modeling the smooth (and hence, more accurate) relationship between input (changes in m , f) and output (image warping) instead of intermediate parameters that are often nonlinear and co-dependent in parametric space [4]. Therefore, we expect that our general image warping strategy will be advantageous and important for broad deployment of intraoperative stereovision and applicable to other types of stereo/video images as well (e.g., endoscope) because it does not rely on any *a priori* knowledge about the camera system itself other than the induced image deformation remaining constant and being reproducible. In addition, the general strategy does not dictate the specific techniques used for calibration at the reference setting or 3D surface reconstruction [1, 2], suggesting this methodological approach has potential to be of considerable clinical significance.

Acknowledgement. This work was supported in part by National Institutes of Health grant number R01 CA159324-01 and 1R21 NS078607.

References

1. Mirotta, D.J., Ishii, M., Hager, G.D.: Vision-based navigation in image-guided interventions. *Annual Review of Biomedical Engineering* 13, 297–319 (2011)
2. Hemayed, E.E.: A survey of camera self-calibration. In: *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 351–357 (2003)
3. Figl, M., Ede, C., Hummel, J., Wanschitz, F., Ewers, R., Bergmann, H., Birkfellner, W.: A fully automated calibration method for an optical see-through head-mounted operating microscope with variable zoom and focus. *IEEE Trans. Med. Imag.* 24(11), 1492–1499 (2005)
4. Willson, R.: *Modeling and Calibration of Automated Zoom Lenses*, CMU-RI-TR-94-03, Robotics Institute, Carnegie Mellon University (1994)
5. Edwards, P.J., King, A.P., Maurer, C.R., de Cunha, D.A., Hawkes, D.J., Hill, D.L., Gaston, R.P., Fenlon, M.R., Juszczak, A., Strong, A.J., Chandler, C.L., Gleeson, M.J.: Design and evaluation of a system for microscope-assisted guided interventions (MAGI). *IEEE Trans. Med. Imag.* 19(11), 1082–1093 (2000)
6. Ji, S., Fan, X., Roberts, D.W., Paulsen, K.D.: Cortical surface strain estimation using stereovision. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part I. LNCS*, vol. 6891, pp. 412–419. Springer, Heidelberg (2011)
7. Liu, C.: *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. Doctoral Thesis. Massachusetts Institute of Technology (May 2009)