Simultaneous Two-View Calibration of Biplane Fluoroscopic Imaging Systems

Michael H. Rosenthal, Marc Pollefeys, Elizabeth Bullitt MD, and Henry Fuchs

1 Department of Computer Science
2 Department of Neurosurgery
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599 USA

Abstract. We present a novel method for calibrating biplane fluoroscopy systems using sequences of paired views. This method uses standard techniques to estimate the system calibration on a frame-by-frame basis, then uses a global parameter optimization (bundle adjustment) to improve the quality and consistency of the solution. The method is validated using images of a new calibration phantom. The system is calibrated using a random subset of the available frames, and the resulting calibration is applied to a separate set of evaluation frames to reconstruct the visible phantom features into three dimensions. The mean 3D geometric error after reconstruction was found to be 0.78mm for the local, single-frame approach and 0.75mm for the global approach.

1 Background

Calibration is the process through which the parameters describing a system become known. In the case of fluoroscopy or machine vision, a successful calibration process yields an accurate representation of the imaging system’s geometry. This representation allows one to track, reconstruct, and otherwise interpret the three-dimensional world from a set of two-dimensional projective views.

The body of scientific literature related to the calibration of fluoroscopic systems can be broadly divided into two sections: works in the realm of machine vision, most commonly published in journals such as IEEE-PAMI or the International Journal of Computer Vision; and works published specifically on medical fluoroscopy, which most often appear in journals such as Medical Physics. Perhaps because they tend to be published in journals with different audiences, the works in these two sections are not always well connected and have developed in somewhat different manners. The machine vision literature has included extensive developments on quantitative methods for single-view calibration[1, 2], epipolar geometry[3], and higher-order camera configurations[4, 5]. The fluoroscopy literature has had some developments in geometric calibration methods[6, 7], but has had much more effort dedicated to correction of the large distortion errors associated with fluoroscopic imaging[8].
Fluoroscopy and optical imaging have several important differences that must be considered when bridging knowledge between the systems. Most objects are partially transparent when viewed with x-ray fluoroscopy, which means that correspondence ambiguities between views can not be solved using the simple occlusion cues that generally exist in optical imaging. In addition, fluoroscopy systems generally have narrow fields-of-view ($6-12^\circ$ for Siemens Neurostar) relative to most optical imaging systems, which often have fields-of-view of $25^\circ$ or more. The narrow field-of-view makes estimation of focal length, optical center, and depth substantially more difficult by limiting the perspective effect throughout the image.

Much of the research in fluoroscopy has been conducted with single camera systems because biplane systems are both uncommon and costly. Unlike the parallel stereo setups that are common in machine vision, biplane fluoroscopy tends to involve two cameras with orthogonal views that intersect in the volume of interest. This means that calibration objects must be unambiguous and provide sufficient coverage of two simultaneous orthogonal views in order to be useful for full stereo calibration with a single pair of frames. Such a setup does, however, offer the possibility of improved determination of focal length and translation depth in each view due to the orthogonal view provided by the second camera. This paper presents a novel approach to exploiting the features of an orthogonal biplane system to yield improved calibration and 3D reconstruction results.

Calibration for machine vision extends at least as far back as Sutherland’s work in the 1960s [9]. Since that time, several prominent works have shaped the science of calibration. The first paper proposing a direct linear estimate of the pose between two cameras occurred in 1981 [10]. Beginning in 1986, Roger Tsai and colleagues published a series of works on methods for explicit calibration of intrinsic and extrinsic parameters using known calibration objects [1, 2]. Faugeras and colleagues published extensively on calibration, including linear methods for direct calculation of the projection matrix [3, 11].

The fluoroscopy literature shares some similarities with that of machine vision. Starting in 1989, Fencil and Metz published several papers on the calibration of biplane fluoroscopy systems using correspondences of unknown geometry [6, 7]. This paralleled the development of self-calibration techniques in the machine vision literature. Much of the remaining fluoroscopy literature explores the significant distortion problems encountered in fluoroscopic imaging and is beyond the scope of this work.

2 Methods

2.1 Simultaneous Two-View, Multi-Frame Calibration

When calibrating a camera based upon a single frame, errors in target point localization or the calibration phantom can yield inaccurate estimates of the parameters for that frame. As the view varies across frames, these errors can yield variations in the calibration results that appear as jumps between frames -
a highly undesirable feature of any calibration method. In the interest of finding an optimum calibration across such a series of frames, a nonlinear optimization scheme was developed that simultaneously optimizes the global camera calibration parameters and the local frame transformations.

First, we assume that the true intrinsic parameters for both cameras and the true intercamera extrinsics are constant across the series of frames. We then define $K_A$ and $K_B$ as the 3x3 intrinsic projection matrices for the A and B (anteroposterior and lateral) cameras, respectively. Each of these upper triangular matrices has five degrees of freedom (four if skew is assumed to be zero, as in this paper) and is defined up to a scale factor. We also define $R_{A \rightarrow B}$ and $t_{A \rightarrow B}$ as the intercamera extrinsic rotation matrix and translation vector, respectively (a 3D Euclidean transform). These describe the transformation $X_B = R_{A \rightarrow B}X_A + t_{A \rightarrow B}$ and add another six degrees of freedom to the system.

If the calibration phantom and/or cameras are moved during the sequence, then the extrinsic transformation mapping the world or calibration object coordinate system to that of camera A will vary by frame. We define $R_{W \rightarrow A}$ and $t_{W \rightarrow A}$ as the rotation matrices and translation vectors comprising the Euclidean transformation from World to camera A coordinates for each frame $i$. The per-frame extrinsic estimates add $6 \times N$ parameters to the system, where $N$ is the number of pairs of frames in the sequence. The transformations are composed in the same manner as for the intercamera transformation above.

Uncertainty in the calibration model points can be accommodated by adding global parameters describing the true location of each model point. In the case of the custom calibration phantom described above, the location of each spherical target along a cylindrical channel is uncertain. A set of one-dimensional coefficients $k_j \in \mathbb{R}$ defining the $j^{th}$ bearing’s displacement along the drill axis vector $d_j$ was added to the model to compensate for these uncertainties.

The least-squares function to be optimized, $F(x)$, is described below. This cost function has three components: $Err(x)$, the sum of the geometric error distances between each image point and the projection of the corresponding model point; $Var(x)$, the weighted sum of the variances of the parameters that are being globalized; and $Constr(x)$, a cost element intended to restrict the correction of the model points within the physically feasible range of $\pm 0.5$mm. In these equations, $m$ represents the number of frame pairs, $n$ represents the number of model points, $P_{A/B}^i$ represents the projection matrix for camera A or B, $X_{mod}^j$ represents the three-dimensional coordinates of the $j^{th}$ model point, and $x_{A/B}^j$ represents the 2D image coordinate of the $j^{th}$ on camera A or B. In addition, $\Gamma$ represents the set of parameters to be globalized, as described above, $p_l$ represents the $l^{th}$ such parameter in the $i^{th}$ frame, and $\mu_l$ is the mean of the $l^{th}$ parameters across all frames.

$$F(x) = Err(x) + Var(x) + Constr(x)$$

$$Err(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( P_{i}^{A} X_{mod}^{j} - x_{A}^{j} \right)^{2} + \left( P_{i}^{B} X_{mod}^{j} - x_{B}^{j} \right)^{2}$$
\[ Var(x) = \sum_{t \in T} w_t^2 \sum_{i=1}^{m} \frac{(p_{li} - \mu_l)^2}{m - 1} \]  
\[ Constr(x) = \sum_{j=1}^{n} (2k_j)^{20} \]  

The variance weighting factor \( w_l \) was chosen as the inverse of the initial variance of the corresponding parameter in order to equalize the optimization ‘pressure’ across parameters with differing magnitudes. In addition, this weighting factor was multiplied by a scale factor increasing from \( 2^0 \) to \( 2^{10} \) over 11 sequential iterations of the optimizer; this increasing weight on interframe variance is intended to allow a gradual convergence from local parameterization to global parameterization.

The global optimization routine uses a region trust method (Matlab’s \texttt{lsqnonlin} function for large scale problems) with a preconditioned conjugate gradient (PCG) step. The PCG step essentially rescales the elements of the function’s Jacobian matrix so that it is well-conditioned. This is critical to achieving convergence when the partial derivatives of some parameters are significantly different than those of other parameters.

2.2 Experimental Design

Image Capture A custom imaging phantom containing radio-opaque metal bearings in two orthogonal planes was used to provide radiographically distinct reference targets for calibration [12]. Several series of biplane views of the calibration phantom were captured in a single session, yielding a complete data set comprising 512 pairs of images. In order to vary the calibration images over the session, the phantom was translated and rotated during and between these imaging series. The biplane fluoroscopes remained stationary throughout the session. The images were captured from an 884x884x12-bit grayscale analog video signal and converted into 884x884x8-bit grayscale images using Foresight Imaging I-50 video capture cards.

Image Processing and Calibration For each image in the data set, all of the fully visible calibration targets were automatically extracted by thresholding the image and identifying connected regions with appropriate size and shape characteristics. The location of each target was calculated as the binary centroid of its elliptical region.

The intrinsic and extrinsic parameters for each frame pair were estimated separately using the normalized direct linear transformation method and jointly refined via nonlinear optimization of the in-plane geometric distance error. Both methods are described well in [13] and are based upon work by Sutherland, Hartley, and others. This implementation was validated using synthetic data and was found to produce parameters correct to a factor of approximately 1e-10. A subset of twenty-one pairs of frames were selected and their calibrated
Data Analysis The sequence of available frames was randomly sampled into separate sets of calibration and evaluation frames. Each of thirteen calibration sets included five frames and was used to perform local and global calibration. The resulting calibration parameters were used to reconstruct the visible features in the evaluation frame set into three dimensions (twenty-six frames). A separate bundle adjustment process was used to calculate the corrected calibration target world coordinates using only the evaluation frame set. These corrected model coordinates were used as ground truth in evaluating the reconstructions for both the local and global methods. The means and frequencies of 3D geometric errors between the corrected model coordinates and reconstructed 3D coordinates are reported below.

3 Results

3.1 Stability of Parameter Estimates

Figure 1 shows the variation of the extrinsic parameters of the imaging system over a typical frame series. The world to camera extrinsics are expected to vary

Fig. 1. Variability of local extrinsic camera parameters across frames. Parameters were calculated for each frame and view using the methods described in the Methods section. The large jumps in the angle measures on the AP view chart are due to transitions from $+\pi$ to $-\pi$, which represents a minor variation in real terms.
Variability of local intrinsic camera parameters across frames. Parameters were calculated for each frame and view using the approach described in the Methods section.

because they are measured relative to the calibration phantom’s internal coordinate frame and the target was moved throughout the imaging series. The intercamera transformation should, however, remain constant across the frames, so the changes in camera extrinsics should be matched (with appropriate coordinate changes) between the views. As seen in Figure 1, this was not always the case. The global optimization forces all frames to a common intercamera calibration.

Figure 2 shows the variation of the intrinsic (projection) parameters over the same frame series. The intrinsic parameters are expected to be stable over the image series because the cameras were not altered or zoomed during the image acquisition. The per-frame calibration approach again demonstrated substantial variability in the intrinsic parameters of both cameras; global optimization eliminates this variability.

3.2 Quality of Reconstruction

The results of the reconstruction tests are reported in Figure 3 and Table 1. $\mu_{3D}$ and $\sigma_{3D}$ are the mean and standard deviation of the three-dimensional geometric error between ideal and reconstructed points. $\mu_{ray}$ and $\sigma_{ray}$ are the mean and standard deviation of the minimum distance between the back-projected reconstruction rays. The $\tilde{v}$ forms show each of the corresponding items broken into their three-dimensional component means; this allows comparison of errors.
between the three spatial axes. The mean three-dimensional distance between a 3D point reconstructed from two views and its corresponding model point was found to be 0.78mm for the per-pair calibration and 0.75mm for the globally optimized calibration. The global method showed somewhat lower error variance and a reduction in the number of large reconstruction errors, with the fraction of errors over 1.5mm being 6.1% for the local method and 4.1% for the global method. The distance between the backprojected reconstruction rays was found to be slightly larger for the per-pair calibration (1.17mm ± 3.94) than for the global calibration (1.07mm ± 3.91mm). Figure 3 shows the distributions of geometric errors for each method.

![Local Calibration Method vs. Global Calibration Method](image.png)

**Fig. 3.** Distribution of three-dimensional geometric errors between ideal and reconstructed world points. These errors were calculated by calculating calibration parameters using one set of frames and using those results to reconstruct the phantom features in a second set of frames.

4 Conclusions

The global calibration method was found to produce slightly lower mean geometric errors and a lower fraction of large geometric errors as compared to local, per-frame calibration (0.75mm vs. 0.78mm). This improvement is not nearly as significant as expected, so additional exploration of the residual errors is warranted. Possible sources of residual error include distortion in the fluoroscopic im-
Table 1. Three-Dimensional Reconstruction Errors

<table>
<thead>
<tr>
<th>Method</th>
<th>$\mu_{3D}$</th>
<th>$\sigma_{3D}$</th>
<th>$\bar{\mu}_{3D}$</th>
<th>$\bar{\sigma}_{3D}$</th>
<th>$\bar{\mu}_{ray}$</th>
<th>$\bar{\sigma}_{ray}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>0.78mm</td>
<td>0.48mm</td>
<td>0.00</td>
<td>0.57</td>
<td>0.06</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>1.17mm</td>
<td>0.17</td>
<td>0.59</td>
<td>1.16</td>
<td>3.94mm</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.41</td>
<td>0.07</td>
<td>0.41</td>
<td>0.07</td>
<td>0.27</td>
</tr>
<tr>
<td>Global</td>
<td>0.75mm</td>
<td>0.39mm</td>
<td>0.00</td>
<td>0.52</td>
<td>0.06</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>1.07mm</td>
<td>1.06</td>
<td>0.56</td>
<td>1.06</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.34</td>
<td>0.06</td>
<td>0.34</td>
<td>0.06</td>
<td>0.24</td>
</tr>
</tbody>
</table>

ages, errors in target feature localization, and target manufacturing errors. The incorporation of distortion correction into the global bundle adjustment method is presently underway; preliminary results indicate a substantial improvement in reconstruction accuracy. Such improvements in reconstruction accuracy will support the demanding needs of intraoperative fluoroscopic guidance tasks.

5 Acknowledgements

The authors would like to acknowledge Mark Van Horn, Ph.D, Matthew Mauro, M.D., and Susan Weeks, M.D. for their support of and contributions to this research. This work was funded in part by NIH-NCI grant 1-R01-HL69808-01.

References