C-arm Tracking and Reconstruction Without an External Tracker*

Ameet Jain and Gabor Fichtinger

Department of Computer Science, Johns Hopkins University

Abstract. For quantitative C-arm fluoroscopy, we have developed a unified mathematical framework to tackle the issues of intra-operative calibration, pose estimation, correspondence and reconstruction, without the use of optical/electromagnetic trackers or precision-made fiducial fixtures. Our method uses randomly distributed unknown points in the imaging volume, either naturally present or induced by randomly sticking beads or other simple markers in the image pace. After these points are segmented, a high dimensional non-linear optimization computes all unknown parameters for calibration, C-arm pose, correspondence and reconstruction. Preliminary phantom experiments indicate an average C-arm tracking accuracy of 0.9^o and a 3D reconstruction error of 0.8 mm, with an 8° region of convergence for both the AP and lateral axes. The method appears to be sufficiently accurate for many clinical applications, and appealing since it works without any external instrumentation and does not interfere with the workspace.

1 Introduction

C-arm fluoroscopy is ubiquitous in general surgery, interventional radiology, and brachytherapy, due to its real-time nature, versatility, and low cost. At the same time, quantitative fluoroscopy has not found a *large scale clinical acceptance*, because of inherent technical difficulties involving intra-operative calibration of model parameters, pose tracking, and target matching/reconstruction. While these aspects have been studied extensively, a clinically extant solution appears to be lacking. Advanced commercial and academic systems employ resident calibration structures [1,2,3] and optical/electromagnetic trackers or calibrated radiographic fiducials [4,5,6] to obtain the C-arm pose. The resulting equipage tends to be prohibitively expensive and complex that often interferes with the subject, image space, and clinical work-volume. While some procedures may be more tolerant to these shortcomings, despite pressing clinical needs, quantitative fluoroscopy is completely missing from brachytherapy, which is the motivating application of our project.

It can be observed that point correspondence across images (without any knowledge about their 3D locations) is a very strong constraint for pose estimation, also referred to as *bundle adjustment* in computer vision [7]. In fact,

* This work has been supported by DoD PC050170 and NIH 1R43CA099374-01.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 494–502, 2006.

six *known correspondences* across two X-ray images are sufficient to constrain the relative C-arm pose. If eight or more correspondences are available, well known linear methods exploiting the *fundamental matrix* can be applied [7], while with five correspondences, a maximum of ten degenerate solutions are possible. Thus, in general, five point correspondences across three X-ray images can recover the C-arm locations. Theoretically four correspondences have been shown to be sufficient to recover the poses in general, barring a zero measure set of the configurations that can lead to multiple solutions [8]. Though, these configurations are known to lie on certain special cubic curves (incld. cases when three points project collinearly on any X-ray image), an intuitive understanding of all these point constellations is not yet known. Moreover, it should be noted that certain special point constellations can always be created, such that they can never be resolved uniquely using any fixed number of images.

In many applications, radioactive seeds, screw/needle ends, implanted surgical markers, special anatomy points etc. are naturally present in the images. By enforcing the "consistency" of these feature points across the images, one can potentially solve for all unknown parameters of calibration, pose recovery, matching, and reconstruction in one fell swoop, in one massive high-dimensional non-linear optimization loop. In applications that do not have an adequate information in their images, one can place a few additional sticky beads or wire markers *randomly* on any temporarily stationary part (for example on the patient skin or under the operating table), and then apply the same framework. Thus the subject of this paper is a unified mathematical framework to solve the problems of intra-operative calibration, pose tracking, and target matching, and reconstruction without any sort of pre-fabricated external fiducials or tracking instrumentation. In applications where there is no need for real-time tracking of mobile surgical instruments, for example in prostate brachytherapy, the complete elimination of intra-operative tracking and calibration entourage promises to lead to a wider clinical acceptance of quantitative 3D C-arm fluoroscopy.

2 Methods and Materials

The three integral components of this problem, in decreasing order of complexity are: (1) point correspondences; (2) C-arm pose; and (3) C-arm geometry calibration. We assume that the points have been segmented from the X-ray images. Though generic and extendible to any number of images, we currently develop the framework for exactly three images. The reason for not using two images is that they have reconstruction singularities, solved by introducing a third image, which in turn makes the problem NP-Hard (*i.e.* no algorithm can even verify the optimality of a given solution in polynomial time). The theoretical complexity of four or more images is similar to that of three images. Thus we propose a detailed solution for three images, which is easily extendible to multiple images.

2.1 Mathematical Framework

Let N be the number of points chosen arbitrarily from the clinical work volume and let N_m be the number of points detected in images I_m with pose $[R_m, T_m]$ and projection model M_m. We do not assume that the 3D points are distinctly visible in all the images, but are allowed to be hidden under one another. Though this makes the correspondence problem significantly harder, it is a more realistic representation of the clinical setting. Let s_{lm} be the position of lth point in m^{th} image. When three images are used, the problem can be formulated as a large optimization problem.

$$\underset{\mathcal{R},\mathcal{T},\mathcal{M},f}{\arg\min} \quad \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{k=1}^{N_3} \mathcal{C}_{ijk} f_{ijk}$$

where

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_{ijk} \geq 1 \ \forall\ k\ ; \quad \sum_{j=1}^{N_2}\sum_{k=1}^{N_3} f_{ijk} \geq 1 \ \forall\ i; \tag{1}$$

$$\sum_{k=1}^{N_3}\sum_{i=1}^{N_1} f_{ijk} \geq 1 \ \forall\ j \quad \textit{and} \quad f_{ijk} \in \{0,1\}$$

$\mathcal{C}_{ijk}$ is the the cost (described later) of matching point s_{i1} to points s_{j2} and s_{k3}. Note that it varies with any variation in $\mathcal{R}, \mathcal{T}, \mathcal{M}$. f_{ijk} is a discrete variable taking a value 1/0, and deciding the correctness of the match $< i, j, k >$. The inequalities force every segmented point to be chosen atleast once. Thus, f represents any feasible global match (and vice versa), with the cost of that correspondence given by $\sum\sum\sum \mathcal{C}_{ijk} f_{ijk}$. The problem hence is to compute $\mathcal{R}, \mathcal{T}, \mathcal{M}, f$ that minimize the total cost. It should be noted that since the images represent a real situation, this optimization has a solution with a near-zero cost. The only case in which a unique solution might not exist is when the information is not sufficient, *i.e.* when the number of beads are less than 7 or when they lie in a degenerate configuration.

Complexity: This is a non-linear optimization in $N^3 + 20$ variables with $3N^2$ constraints. The pose and model parameter optimization is in a continuous 20 dimensional space (2×6 for each pose, 3×3 for each model, one less for scale), while that for the correspondences is in a discrete combinatorial space. Note that we assume pixel sizes to be constant and known. Even if the pose & model parameters are known, it can be shown that the combinatorial optimization reduces to the minimum-weight tri-partite matching problem, known to be NP-Hard. This severely ill-conditions the problem, necessitating methods to constrain the problem adequately. It should be noted that though the global optima for two image matching can be proved to have only a cubic complexity, in many cases it suffers from singularities (Figure 2 (a)), forcing the use of a third image and hence an exponential complexity.

3D Reconstruction: Though crucial, 3D reconstruction is not explicitly incorporated into the framework. By optimizing for the cost, we also indirectly

compute the best reconstruction. Thus when the true pose parameters have been computed, the correct 3D reconstruction will come out as a byproduct. In the meanwhile, we shall only concern ourselves with the explicitly incorporated pose, correspondence and calibration parameters.

C-arm Imaging Model: C-arm imaging is typically approximated as a 5-parameter projection camera [1,2,3,4] to be calibrated intra-operatively for *each* individual image. Recently, however, it has been shown [9] that C-arm calibration might not always be necessary (Figure 1), implying that any reasonable calibration can be *assumed* for each image without actually calibrating at any time. Note that we assume the imager pixel sizes to be fixed and known.

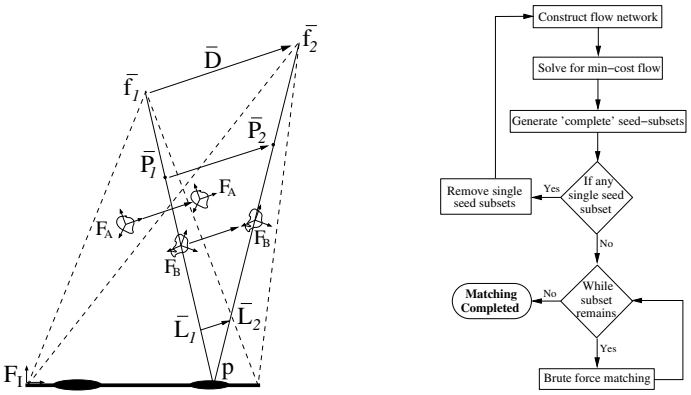

Fig. 1. Mis-calibration shifts all reconstructed objects, but keeps the relative pose nearly constant (left). The flowchart for the correspondence algorithm (right).

The central intuition is that *while an incorrect calibration gives erroneous estimates for the absolute transformations, nevertheless it still provides acceptable relative estimates.* Experimental results corroborate the theoretically derived bounds in that mis-calibration by as much as 50 mm still allows for tracking with an accuracy of 0.5 *mm* in translation and 0.5^o in rotation, and such mis-calibration does not impose any additional error on the reconstruction of small objects [9]. Thus to condition the optimization in Equation (1) better, it it advisable not to solve for the imaging parameters using sparse data from the image, but just use nominal values that may be known from a pre-calibration or the manual/header. An alternate perspective is to notice that since reconstruction errors change only negligibly with calibration errors, any attempt to calibrate using a sparse point set (instead of a very accurate calibration fixture) will ill-condition the problem by allowing for a whole space of feasible solutions. Thus it is wiser to fix the value at a choice that is practically close to reality. Later, if needed, the calibration can be further refined after the optimization converges.

Though we do not explicitly address the issue of distortion correction, advancements in intensifier tubes allow for lesser distortion and more recently the

advent of flat panel detectors obviates this step altogether. Furthermore, many application like brachytherapy, with their limited C-arm workspace, allow for a constant pre-operative distortion correction.

Correspondence: Assuming known pose parameters, we converted the point correspondence problem to a weighted tri-partite matching problem in Equation (1), an NP-Hard combinatorial optimization problem. An attractive approximate solution using a network-flow-based combinatorial optimization has recently been extended to efficiently deal with "hidden seeds", (i.e. points that overlap in some images) in practically $O(N^3)$ times [10].

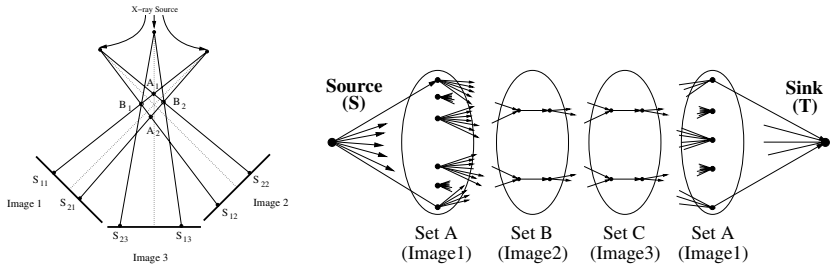

Fig. 2. A third image is needed to resolve two-image singularities (left). The flow network formulation used to solve the correspondence problem.

Sets A, B, C, and D, represent an image each. Links with a cost connect every feasible match between any two images. A flow of value N originates at the source S and ends at sink T. The problem reduces to computing a min-cost flow, easily computed using the cycle cancelling algorithm (pushes negative cycles until there are none left). A flowchart is illustrated in Figure 1. Since the problem is NP-hard the network cannot completely constrain the same point in both set A and D. Nevertheless, it works well, producing near perfect matchings.

Cost Metric: In general, any cost-metric that directly measures the deviation from the observation should perform well. The metric should incorporate all the available information, making the global minimum sharper and the algorithm robust. One good choice for a metric is *projection error* (PE). For any given set of poses and correspondence, the intersection of the three lines that join each projection to its respective x-ray source can be computed using a closed form solution that minimizes the L_2 norm of the error. PE can be computed by projecting this 3D point in each image and then measuring the distance between the projected location and the observed location of the point.

2.2 Optimization Strategy

Due to the convolution of both continuous and discrete parameters, the optimization in Equation (1) becomes ill-conditioned. Incorrect pose estimates will

invoke incorrect correspondences, especially for *dense* point clouds. However, it should be observed that *incorrect pose estimates and the subsequent correspondences invoked by it, are typically inconsistent.* This is because, while it is true that any given pose estimates invoke a correspondence, it is also true that any given correspondence also invokes a unique pose where Equation (1) will be minimized for that correspondence. Thus the desired minima will be such that the current pose invokes a correspondence, and the correspondence in turn will also invoke the same pose (with near zero cost). This order of stability we believe, will exist only at the true unique global minima. We propose an iterative strategy that exploits this observation, in spirit similar to a coordinate descent method.

Another observation to make is that given any generic estimate of the pose, the correspondence is usually completely incorrect. Nevertheless, some other pose in the vicinity can usually establish atleast a few correct correspondences ($\sim 10\%$). This new pose estimate will *behave* like a local minima. We say behave because, for any fixed pose estimate, computing f such that Equation (1) is minimized is not a polynomial time computation ($N!^2$). Thus only a working algorithm can be practically available. Even though our correspondence algorithm has been experimentally shown to be over 98% accurate near the correct pose, its assumptions start breaking down at incorrect estimates. Nevertheless, if we can estimate these few correctly matched points and block the rest, we can quickly converge to the correct answer. In the absence of any additional information, a working strategy is that if a flow in the network originates at a vertex i in set A and also ends up at vertex i in set D, then this flow is *self-consistent.* We choose a subset of self-consistent points as matched points (typically the ones closer to the average PE). These few points can now be easily used to update the pose estimates, which in turn could provide a improvement in the correspondences at a later stage. Thus the algorithm iteratively establishes the best possible correspondences (keeping the pose relatively constant) and then uses a *self-consistent* subset of points to refine the pose. As the iteration progresses, it can stop only at a self-consistent parameter choice where the pose and correspondence perfectly complement each other.

For very high density clouds, the optimization might require non-practical number of iterations to converge. Two main methods to constraint such cases are: (1) establishing a good initial estimate using prior knowledge about the surgical protocol and workspace constraints. (2) in the absence of good initial estimates, a couple of known correspondences can prove to be sufficient. These might be naturally available or artificially induced. In any case, this might become a necessary step since projective geometry can recover the 3D reconstruction only up to an arbitrary scale. To recover the scale, information external to the image is required (ex. length of an inserted screw), allowing for a few known correspondences.

3 Phantom Experiments and Results

A radiographic fiducial was used to track the C-arm (0.56 mm translation; 0.33^o rotation accuracy), and was accurately attached to a point cloud phantom as

shown in Figure 3. The cloud phantom comprises of multiple slabs, thus capable of multiple *random* point configurations. 100 points with 1.56 *points/cc* were used. X-ray images within a 20° cone around the AP-axis were *randomly* taken using an *Philips Integris V3000* fluoroscope and dewarped. Thus both the seed locations and X-ray pose were not biased/optimized in any way, closely representing an uncontrolled surgical scenario. Each image was hand segmented to establish the true segmentation and correspondence. The true C-arm pose and reconstruction was compared to that computed from the algorithm.

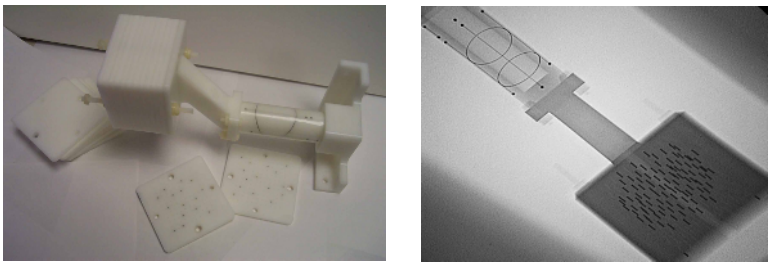

Fig. 3. An image of the point phantom attached to the fiducial (left). The phantom can replicate many point configurations. A typical X-ray image of the combination (right).

We divided the experiments into three separate cloud sizes: sparse having an average 3D point separation $\geq 25\ mm$, medium having $\sim 15\ mm$, while a high density one being $\leq 10\ mm$. These represent different types of surgical scenarios, ranging from orthopedic to brachytherapy. We generated random clouds using $10-20, 20-40$ & $40-100$ points. Region of convergence (ROC), accuracy of C-arm tracking and reconstruction error (RE) are the three metrics used to evaluate performance. Since the scale is not directly recoverable, only the rotation errors are used to study errors in the pose. To study RE, the scale is established using two points from the fiducial.

Figure 4 plots the performance of the algorithm. Each data point is averaged using 10 random runs of the initial estimate and the point cloud. When no prior correspondences are available, the algorithm could have some difficulty in converging reliably. Pose recovery accuracies vary with the number of available points, the average being about 0.9^o, while RE remains fairly stable at $0.8\ mm$. The ROC for sparse and medium sized clouds is $8-10^o$ (individually along both AP and lateral axes), while it is about 6° for dense implants.

Runtime: The algorithm was implemented in Matlab 7 on a Windows PC (3.2 GHz P4, 1GB RAM). The algorithm would typically converge in anywhere between 3-7 total iterations, taking 2-7 minutes, depending on the point cloud density and initial estimate. However, it should be noted that it spends about 50-70% of the time for file I/O (a Matlab constraint). Thus a C/C++ implementation is expected to run in 30s.

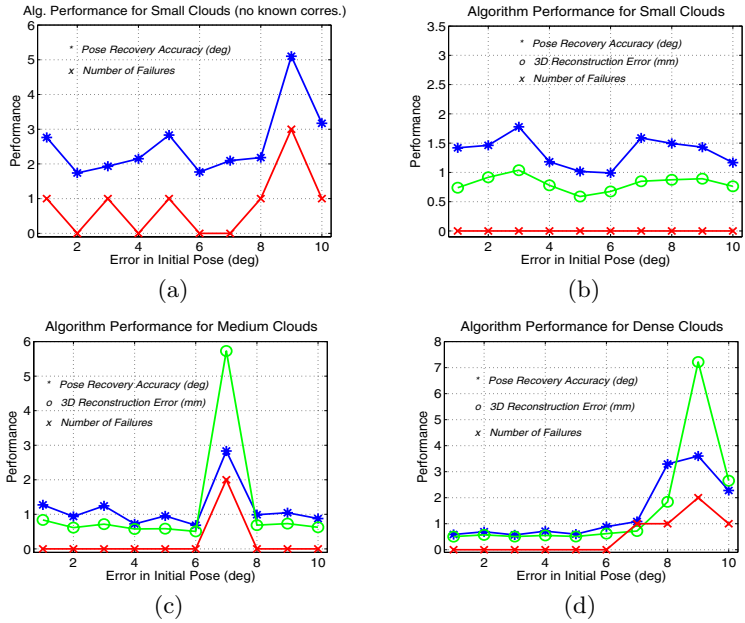

Fig. 4. The performance of the algorithm as a function of initial estimate. (a) is with no known correspondences, while (b)-(d) are with a few known ones.

4 Conclusion

A unified framework for point correspondence, C-arm tracking and reconstruction has been proposed and experimentally validated on phantoms. The experiments indicate an accuracy of 0.9^o for tracking, 0.8 mm for 3D reconstruction and a convergence region of 8° (each) in both the AP and lateral axis of rotation. The framework does not need external fiducials for C-arm pose estimation and is capable of using information naturally present in the X-ray images of a family of clinical applications, such as prostate brachytherapy. In applications where this information is not present, or a greater accuracy is desired, the framework easily extends by randomly attaching beads around the patient. Our technique does not compromise on the available clinical work volume. The framework can also accommodate any available prior information on projection angles or correspondences to constrain the optimization better, and thereby to achieve a higher accuracy.

The main concern for the clinical use of methods relying heavily on high dimensional optimization is that of providing uniformity and reliability in performance. We have conducted our validation on randomly selected views and number/distribution of the points, indicating the robustness of the algorithm to these issues. Nevertheless, this is only a first step and further work to achieve better uniformity in the results is desirable. Though the alternative of well designed

calibration fixtures and image acquisition procedures are also available, they become cumbersome in many procedures. Further development of the current approach or even an amalgamation of the two approaches, could significantly improve the current clinical viability of intra-operative quantitative fluoroscopy. The development of a clinical prostate brachytherapy system to further validate our approach is currently underway. Note that, even though the driving application was prostate brachytherapy, the method also has potential in many synergistic applications in orthopedics and angiography.

References

1. Navab, N., et al: Dynamic geometrical calibration for 3-d cerebral angiography. In: SPIE Medical Imaging. (1996) 361 – 70
2. Tang, T.: Calibration and point-based registration of fluoroscope images. Master's thesis, Queen's University (1999)
3. Livyatan, H., Yaniv, Z., Joskowicz, L.: Robust automatic c-arm calibration for fluoroscopy-based navigation: A practical approach. In: MICCAI. (2002) 60–68
4. Hofstetter, R., Slomczykowski, M., Sati, M., Nolte, L.: Fluoroscopy as an imaging means for computer-assisted surgical navigation. CAS **4(2)** (1999) 65–76
5. OEC 9800 FluoroTrak^{TM}: (GE Healthcare, Waukesha, WI)
6. Jain, A., Mustufa, T., Zhou, Y., Burdette, E.C., Chirikjian, G., Fichtinger, G.: A robust fluoroscope tracking (ftrac) fiducial. Med Phys **32** (2005) 3185–98
7. Ma, Y., et al: An Invitation to 3-D Vision. Springer (2000)
8. Holt, R and et al: Uniqueness of solutions to three perspective views of four points. IEEE Trans. on PAMI **17** (1995) 303–307
9. Jain, A., Kyon, R., Zhou, Y., Fichtinger, G.: C-arm calibration - is it really necessary? In: MICCAI. (2005) LNCS 3749, 639–646
10. Kon, R., Jain, A., Fichtinger, G.: Hidden seed reconstruction from c-arm images in brachytherapy. In: IEEE ISBI. (2006) 526–29