Bone tracking from X-Ray sequences

Laura Sevilla Lara

May 15, 2009

1 Introduction

1.1 Definition of the problem

Bone tracking from X-Ray sequences is important for biological and evolutionary applications like motion analysis of species, and also in human health, for example in surgical guidance, study of kinematics, or post-operative implant assessment. Since X-Ray data is 2-dimensional, it does not provide enough information about the 3D position of the bones. C-Arm technology (Fig. 1) reduces this problem by capturing two simultaneous views of the bone. Using this stereo vision and a volumetric model of the bone, typically a CT scan, the search is performed by comparison of a synthetically computed view (Digitally Reconstructed Radiography or DRR) and the X-Ray image. The objective is to solve for the transformation of the bone that produced each X-Ray image. The high density of the CT scan data makes DRR synthesis a computationally expensive task. Also, high frame rate in the sequences is desired for tracking accuracy, but it results in noisy and low contrast X-Ray images [1], which makes the registration difficult. Furthermore, comparing multi-modal data like in the case of X-Ray and CT is challenging because they don't follow linear transformations [2]. In this project we explore three techniques, that address these issues.

- We define a function that compares the DRRs and the X-Rays. For a given set of values of the rotation and translation parameters, this function outputs a similarity measurement between the DRRs and the X-Rays. The objective is to minimize this function and retrieve the rotation and translation parameters of the position of the bone. Unlike

current techniques, we explicitly compute the derivative of this function with respect to the rotation and translation parameters. This avoids the numerical computation of the gradient, speeding up the process.

- To decrease the size of the data and prevent multi-modal data comparison, we use only the surface of the bone and compare the edges of the synthetic view and the reference X-Ray image.

- To incorporate knowledge about the intensity in the search, we take a segmentation approach by minimizing the entropy of the background.

Each of this approaches has strengths and weaknesses. In future work we will consider the principled combination of these methods.

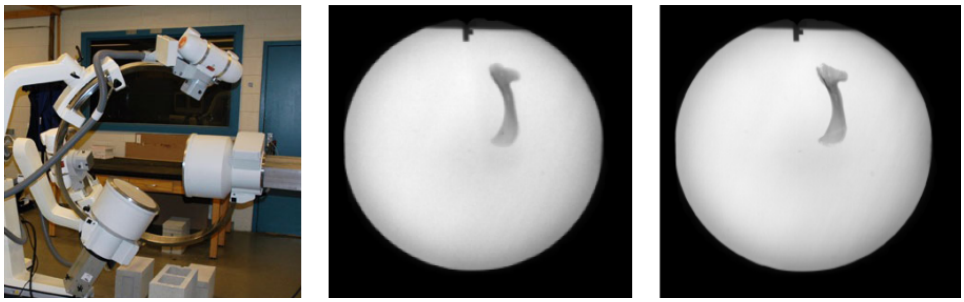

Figure 1: C-Arm Technology provides stereoview (left). X-Ray image (center). DRR (right).

1.2 Previous work

X-Ray registration methods can be classified into two categories: geometry-based and intensity-based. *Geometry-based* methods solve for the transformation that converts a set of points or lines in 3D space into their correspondence in 2D space. These points can be either detected features or implanted fiducial. One major advantage of this approach is that it reduces the size of the data, thus decreasing the computational cost. However, the quality of the registration relies on the assumption of finding natural landmarks in the case of feature-detection and artificial markers have the disadvantage of being an invasive method that requires surgery. *Intensity-based* methods match the two data sets by minimizing a distance function based on the intensity of the

pixels or voxels. Typically, this is done by synthesizing the 2D projection of the volume into what is called a Digitally Reconstructed Radiograph (DRR), simulating the physical process of a X-Ray projector. The DRRs are then compared to the real X-Ray sequences to minimize their difference. This turns into an optimization problem, where the goal is to minimize a distance function with respect to the 6 DOF in the bone position. Since multi-modal data is difficult to compare, some previous techniques have used non-linear distance metrics, like mutual information [3, 4]. Normalized cross correlation (NCC) is the traditional distance metric [5, 6], but it is combined with other characteristics like VLNC (Variance of Local Normalized Correlation), in [6] and edge detection in [5]. Other similarity metrics are pattern intensity [7]. The biggest bottleneck in these DRR based approaches is the computation time. There have been efforts to deal with this obstacle at different levels. In [8], they focus in the graphics problem. In [6] they sacrifices accuracy, computing only parts of a DRR that are considered of interest. In [5] they use parallelization, limiting the application of the system to cluster availability. In this project we seek a solution that can run in a desktop machine.

1.3 Data set

The data used in our experiments was collected by the Department of Evolutionary Biology at Brown University, in collaboration with the Rhode Island Hospital. We have used two alligator bones, a coracoid and a humerus. The 3D data consists of a CT scan of the two bones, represented using a 3-dimensional matrix of high level of detail (0.1925 mm in two of the dimensions and 0.625 mm in the third). Using the Marching Cubes algorithm for isosurface extraction, we also obtained a mesh of the bone surface. The 2D data consist of pairs of X-Ray sequences, of high spatial resolution (1024x1024 pixels). The planes of the views form approximately a 90 degree angle as shown in Fig. 1. These sequences have been corrected for distortion. For the purpose of study, these bones had been inserted metal markers. The markers have been manually identified to provide ground truth for the experiments.

2 Methods

2.1 DRR generation

Computation time still remains a challenge in the DRR approach. A 512x512 image takes 40-50 seconds to converge using 13 parallelized machines, using the state of the art technique [5]. Our objective is to speed up the optimization process by providing a explicit derivative of the function to minimize. This is difficult because the traditional approach to generate DRRs uses ray tracing over the voxels of the volume. This means that the function of the intensity of a pixel with respect to the movement of the object has discontinuities in the borders of the voxels, and therefore cannot be differentiated. We generalize ray tracing to a continuous function. A 3D Gaussian is placed around the center of each bone voxel. The intensity at a given pixel of the DRR is the integral of the distance between the ray and every point in the volume according to the Gaussian function, as shown in Fig 2. The objective is to minimize the difference between this DRR and the X-Ray using NCC. We use correlation instead of NCC to simplify the derivative. In our experience with this data set this simplification did not affect the results. We optimize f_1 (1) using an optimization technique based on [9, 10]. Each iteration involves the approximate solution of a large linear system using the method of preconditioned conjugate gradients.

$$f_1(\theta, t) = \sum_{(x,y)} \frac{S_I(\theta, t) * R_I}{\|S_I(\theta, t)\| \|R_I\|} \tag{1}$$

Where θ are the three parameters that describe the rotation, t are the three parameters that describe the translation, $\|S_I\|$ is the norm of the image S_I as a vector, $*$ is the matrix multiplication component by component, R_I is the reference X-Ray image, and $S_I(\theta, t)$ is the synthesized DRR image computed as follows. The intensity of pixel (x, y) in the synthetic DRR is:

$$S_{I_{(x,y)}} = \sum_{p \in V} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2\sigma^2} \frac{\|r_{(x,y)} \times p\|}{\|r_{(x,y)}\|}} \tag{2}$$

Here $p \in V$ are all the points in the volume and $r_{(x,y)}$ is the ray that passes through the center of the camera and pixel (x, y). The term $\|r_{(x,y)} \times p(\theta, r)\| / \|r_{(x,y)}\|$ represents the distance from the ray to point p. Finally, $p(\theta, r)$ is point p of

the volume, after the transformation of θ and t has been applied

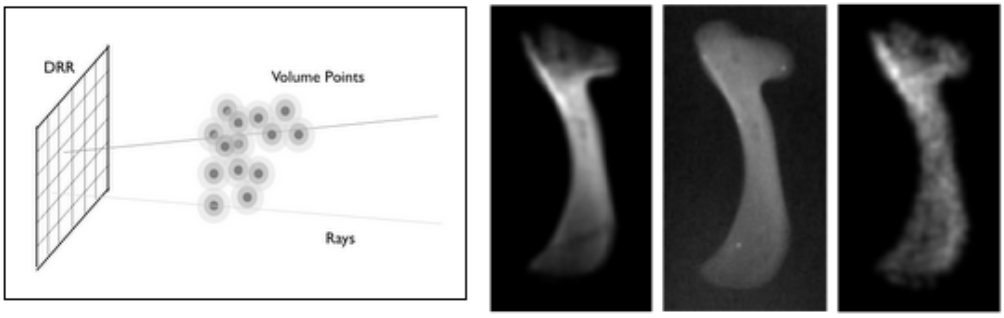

Figure 2: Continuous Ray Tracer Model (left). Full DRR, created using all points in the volume (center-left), Reference Image (center-right) and DRR created with subsampled volume, approximately 1/50 number of points of the full volume (right)

We compute the first and second derivatives of f_1 with respect to the transformation parameters.The function $S_I(\theta, t)$ is written as S_I for clarity.

$$\frac{\partial f_1}{\partial t_1} = \frac{\sum_{(x,y)} R_I S_I' \|S_I\| - \sum_{(x,y)} R_I S_I \|S_I\|'}{S_I^2 \|RI\|} \tag{3}$$

Where S_I' is the partial derivative of the synthetic image with respect to each of the 6 parameters of the transformation and $\|S_I\|'$ is the partial derivative of the norm of the synthetic image, also with respect to each of the 6 parameters. To evaluate the accuracy and speed of the proposed method we shifted the bone in each dimension of the translation and searched for the local minima. We have compared the time for the optimization to converge using finite differences as in [12] to the time using explicit derivatives, and it is improved by a factor of 6. The accuracy in both cases is 0.14 mm, starting from 2.5 mm away. In this project we only implemented translation transformation. To improve performance, the volume was downsampled using a Poisson disk distribution. In practice this did not influence accuracy. Since the results are promising in speed up and accuracy, implementing also rotation is included in our future work.

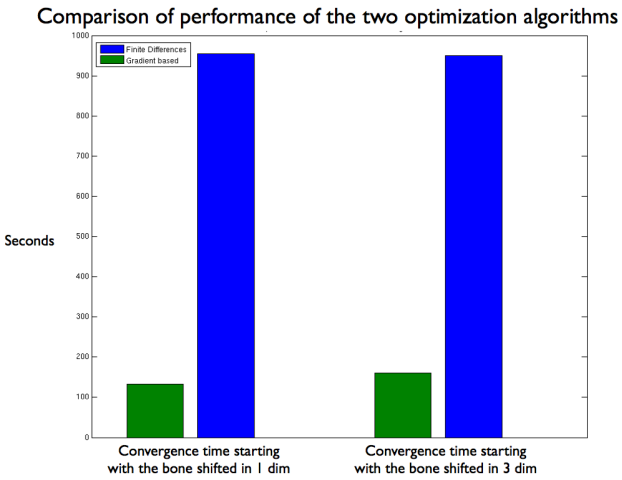

Figure 3: Comparison between finite differences and explicit derivative technique

2.2 Silhouette matching

Operating only with the silhouette provides a major reduction in the size of the data, thus speeding up the computation. We want to minimize the function of the distance between the model's projected edges and the reference image edges. This is computed by adding the distance of every pixel in one edge to the closest point in the other edge. This is done both ways to better constrain the function. The edges in the reference image are detected using Canny algorithm [11].

The new energy function to optimize is f_2:

$$f_2(\theta,t) = \sum_{(x,y)} \frac{Edge(Project(Mesh,\theta,t)) * Chf(Canny(R_I))}{|Edge(Project(Mesh,\theta,t))|} + \frac{Canny(R_I) * Chf(Project(Mesh,\theta,t))}{|Canny(R_I)|}$$

Where M_P is the projected mesh, Chf is the Chamfer distance of an edge, $Edge(I)$ returns the edge of an image I and $*$ is matrix multiplication component by component.

This function f_2 is optimized using the Downhill Simplex algorithm. This technique produces an average accuracy of 3.37 mm with a maximum value

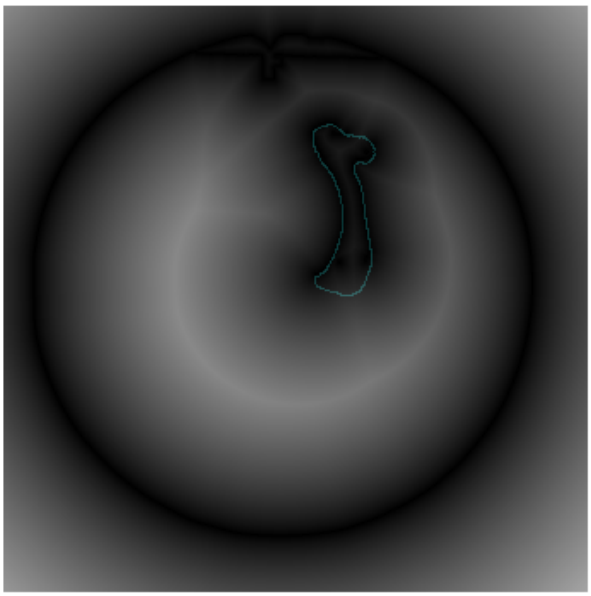

Figure 4: Chamfer distance computed over the Reference Image Edges. Synthetic edge overlaid in blue. Notice, at the bottom of the bone that X-Ray edges can be incomplete

of 5.18 mm.

2.3 Background subtraction

Using the edge eliminates the computational load of dealing with 3D data. The main disadvantage is that the intensity of the X-Ray is not being used explicitly. Using this added information can improve the accuracy and robustness of the results. In this approach we treat the pose estimation problem as a background subtraction case. The goal is to minimize the error in background subtraction. We make the assumption that a well segmented image has low entropy in the background and foreground distributions, while maximizes the distance between them. As shown in figure 5, the mesh silhouette is overlaid on top of the reference image. The pixels inside the silhouette of a well segmented bone should have lower entropy than the pixels of a silhou-

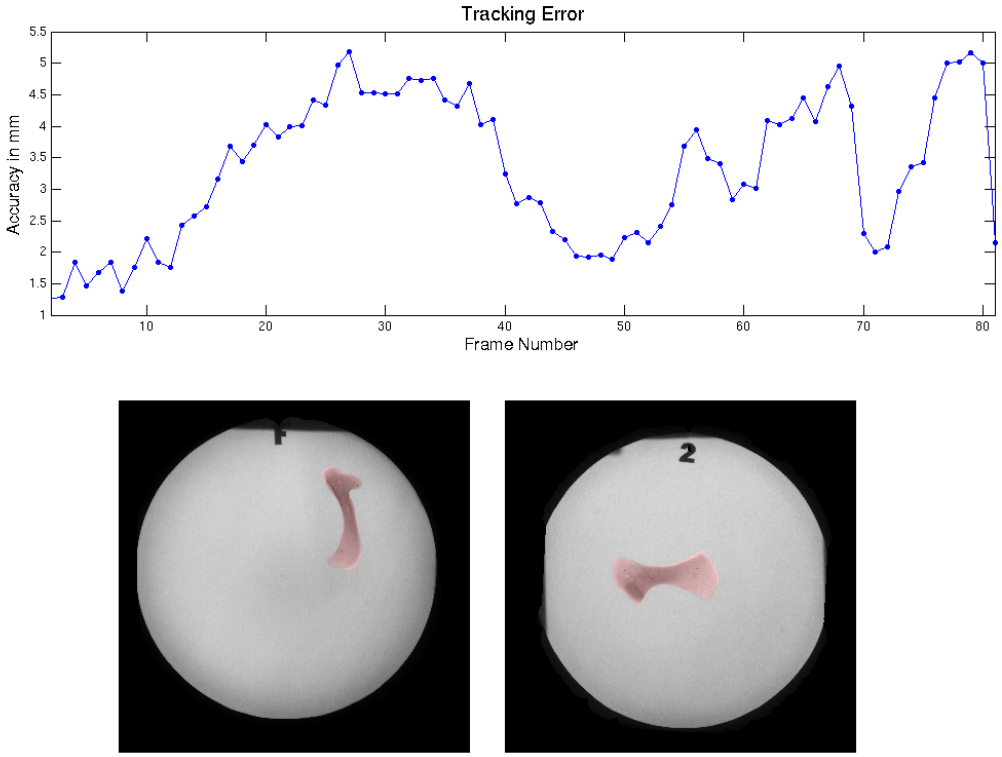

Figure 5: Tracking results using Silhouette Matching. Tracking error (top). Worst registration frames in both cameras (bottom)

ette that covers part of the background. Similarly, the distribution of pixels outside the silhouette will have a lower entropy of none of the darker bone pixels are included.

$$f_3(\theta, t) = H_{Bg}(\theta, t) + \lambda_1 H_{Fg}(\theta, t) + \lambda_2 KL(\theta, t) \tag{4}$$

Where H is the Shannon entropy, and KL is the Kullback-Leibler divergence. In our experience for this particular data set, the background entropy is the strongest term. Since intensities in the background are more uniform than in the bone, the entropy in the foreground increases quickly with small misplacements of the silhouette.

Tracking results with this technique produce a maximum error of 12 mm and and average of 6 mm, as shown in Figure 7.

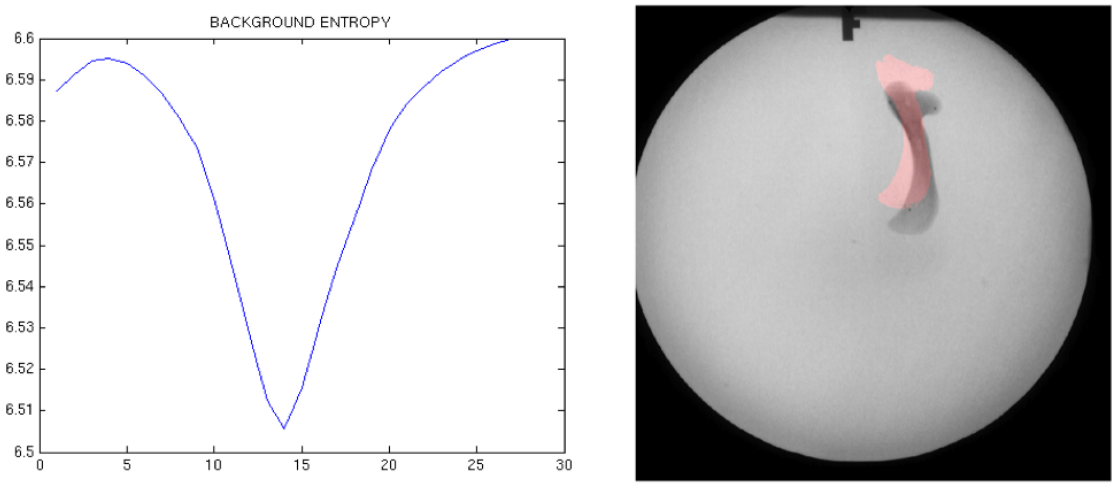

Figure 6: Example of mesh silhouette overlaid on top of the reference image (right). All pixels colored in red are considered foreground, and the rest are considered background. This instance would output high entropy in the foreground and background distributions, and thus high energy in the function to minimize. Entropy with respect to translation (left). Evaluation of frame 14 using the pose of the bone in frames 1-30. The entropy reaches the lowest value in frame 14, where the position is right.

3 Conclusion and future work

We present three different techniques to address the main problems in bone tracking from bi-planar X-Ray: speed and accuracy. Silhouette matching and background subtraction provide a fast way to find a position close to optimal, while the DRR approach finds a more accurate position starting from a close point. These seem to be complementary features, and we will consider combining them into a two step algorithm, that identifies a close to optimal position quickly and then refines it in a slower but more accurate phase.

4 References

[1] Byoung-moon You, Pepe Siy, William Anderst, and Scott Tashman. In Vivo Measurement of 3-D Skeletal Kinematics from Sequences of Biplane

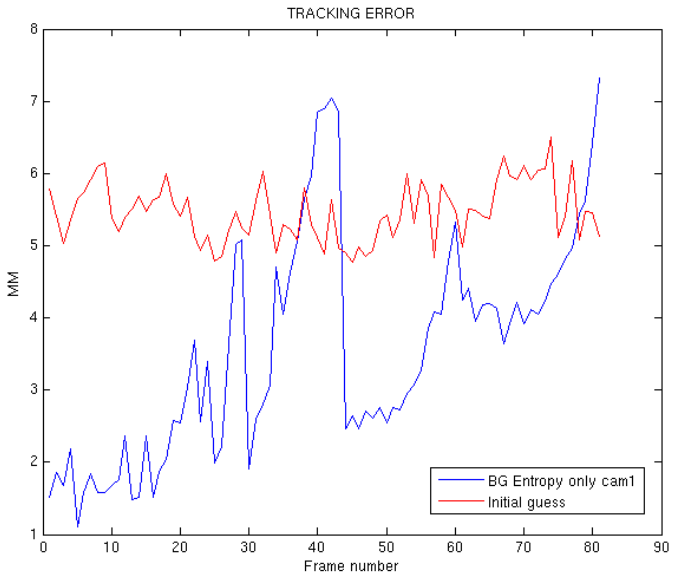

Figure 7: Tracking error across the sequence using background entropy minimization

Radiographs: Application to Knee Kinematics IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 20, NO. 6, JUNE 2001
[2] Alexis Roche, Gregoire Malandain, Nicholas Ayache. Unifying Maximum Likelihood Approaches in Medical Image Registration. INRIA, EPIDAURE Project, 2004
[3] Aouadi, Souha and Sarry, Laurent. Accurate and precise 2D-3D registration based on X-ray intensity. Computer Vision and Image Understanding 110 (2008) 134151
[4] L. Zollei, E. Grimson, A. Norbash, W. Wells. 2D-3D Rigid Registration of X-Ray Fluoroscopy and CT Images Using Mutual Information and Sparsely Sampled Histogram Estimators. IEEE Computer Society Conference on Computer Vision and Image Understanding. 2001
[5] Michael J. Bey, Roger Zauel, Stephanie K. Brock, and Scott Tashman. Validation of a New Model-Based Tracking Technique for Measuring Three-Dimensional, In Vivo Glenohumeral Joint Kinematics. J. Biomech. Eng. – August 2006 – Volume 128, Issue 4, 604
[6] D. Knaan and L. Joskowicz. Effective Intensity-Based 2D/3D Rigid Reg-

istration between Fluoroscopic X-Ray and CT. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003
[7] Weese J; Penney G P; Desmedt P; Buzug T M; Hill D L; Hawkes D J. Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery. IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society 1997;1(4):284-93.
[8] David LaRose; John Bayouth; Takeo Kanade. Transgraph: interactive intensity-based 2D/3D registration of x-ray and CT data. Medical Imaging 2000: Image Processing, Kenneth M. Hanson, Editors, pp.385-396
[9] Coleman, T.F. and Y. Li, "An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds," SIAM Journal on Optimization, Vol. 6, pp. 418-445, 1996.
[10] Coleman, T.F. and Y. Li, "On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds," Mathematical Programming, Vol. 67, Number 2, pp. 189-224, 1994.
[11] Canny, John. A Computational Approach to Edge Detection. IEEE TRANS. PATTERN ANAL. MACH. INTELLIG. Vol. PAMI-8, no. 6, pp. 679-698. 1986 [12] Michael J. Bey, Stephanie K. Klinea, Roger Zauela, Terrence R. Locka and Patricia A. Kolowicha. Measuring dynamic in-vivo glenohumeral joint kinematics: Technique and preliminary results. Journal of Biomechanics Volume 41, Issue 3, 2008, Pages 711-714