

"In Vivo" Pose Estimation of Artificial Knee Implants Using Computer Vision

Scott A. Walker
Research Assistant
Colorado School of Mines
Rose Biomedical Research
Golden, Colorado 80401

William Hoff, Ph.D.
Assistant Professor
Engineering Division
Colorado School of Mines
Golden, Colorado 80401

Richard Komistek, Ph.D.
Director
Rose Musculoskeletal Laboratory
4545 E. Ninth Avenue
Denver, Colorado 80220

Douglas Dennis, MD
Director Clinical Research
Rose Musculoskeletal Laboratory
4545 E. Ninth Avenue
Denver, Colorado 80220

KEYWORDS

Medical Imaging, Pose Estimation, Implants, Fluoroscopic Analysis.

ABSTRACT

This paper describes an algorithm to estimate the position and orientation (pose) of artificial knee implants from fluoroscopy images using computer vision. The resulting information is used to determine contact position from "in vivo" bone motion in implanted knees. This determination can be used to support the development of improved prosthetic knee implants. Current generation implants have a limited life span due to premature wear of the polyethylene material at the joint surface. To get "in vivo" motion, fluoroscopy videos were taken of implant patients performing deep knee bends. Our algorithm determines the full 6 degree of freedom translation and rotation of knee components. This is necessary for artificial knees which have shown significant rotation out of the sagittal plane, in particular internal/external rotations. By creating a library of images at known orientation and performing a matching technique, the 3-D pose of the femoral and tibial components are determined. By transforming the coordinate systems into one common system contact positions can be determined. The entire process, when used at certain knee angles, will give a representation of the positions in contact during normal knee motion.

INTRODUCTION

More than 100 types of arthritis now afflict millions of Americans, often resulting in progressive joint destruction in which the articular cartilage (joint cushion) is worn away causing friction between the aburnated (uncovered) bones ends. This painful and crippling condition frequently requires total joint replacement using implants with polyethylene inserts.

Although artificial knee joints are expected to last 10 to 15 years, research indicates that most implants last an average of just 5.6 years [1]. With population longevity statistics rising, many patients will require additional surgery to replace dysfunctional prosthetic joints to achieve two decades or more of use. Not only does this result in patient discomfort, it also significantly increases health care expenditures nationwide. Currently, more than 400,000 Americans receive lower extremity implants per year, accounting for over \$1.5 billion annually in health care costs.

After a total knee arthroplasty, TKA, the knee joint undergoes a change in boundary conditions. Therefore, it is important to understand knee kinematics under “in vivo”, weight bearing conditions. Fluoroscopy has been used effectively by researchers to analyze existing TKA’s under these conditions. Fluoroscopy is a relatively safe procedure where X-rays are emitted from a tube, pass through the knee and strike a fluorescent screen where the images are intensified and recorded via video tape [2]. The result is a perspective projections of the knee, recorded as a continuous series of images. Figure 1(a) shows an image from a fluoroscope video sequence.

Direct analysis of fluoroscopy images yields translations and rotations within the plane of the image. To extract the out-of-plane rotations and translation into the plane the image must be compared against another image of known pose.

To create images of known pose, accurate 3-D CAD models of the implant are developed and a library of silhouette images is created at known orientations. The match of the silhouettes of the cad model and the image is determined by a simple a pixel by pixel comparison. The match yields values which can be used to extract the pose of the fluoroscopy image. Figure 1(b) shows an image of the CAD models overlaid on top of the original X-ray fluoroscopy image.

The process has been proven to have an accuracy of 0.7 degree (in all three axes) and 0.25 mm of translational accuracy parallel to the image plane and 0.94 mm accuracy perpendicular to the image plane using synthetic data. No accuracy data on real fluoroscopy images is available at this time. The values using synthetic data give a baseline for all accuracy measurements.

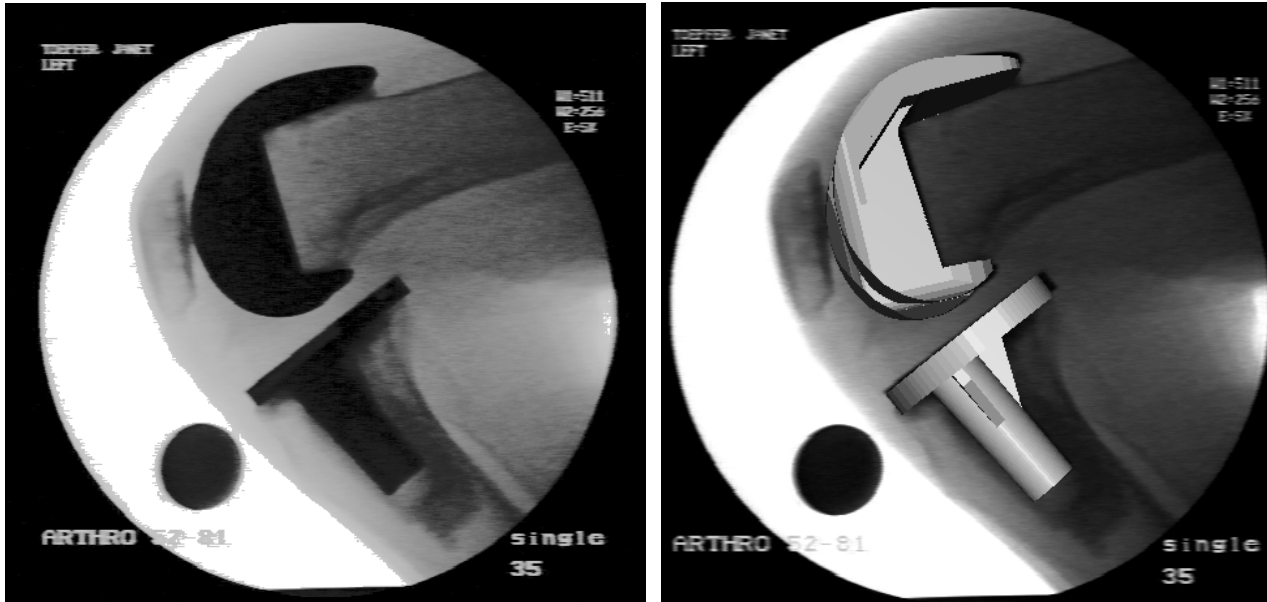


Figure 1. (a) Original X-ray fluoroscopy image of artificial knee. (b) Fluoroscopy image with CAD model overlay.

Previous Work

Work with fluoroscopy images has primarily concentrated on measuring the rotation and translation of components within the plane of the image. Stiehl, Komistek *et. al.* [3] analyzed still photographs of the fluoroscopic video to measure rotations and translations of the knee joint members. Their work was limited to the in-plane rotations and translations of the implant components. In other words, they could measure the x,y translations of the components in the sagittal plane, and the rotation within that plane. However, the actual motion of the components also includes rotations and translations out of the plane of the image.

Our work is designed to provide information on the full 6-degree-of-freedom (DOF) motion of the implant components. The central process in our work is to extract the 3-D position and orientation (pose) of an object (*i.e.*, the implant component) from a single perspective image (*i.e.*, a frame from the X-ray fluoroscopy video). The basic ideas behind this process are well-known and have been used for many years in the photogrammetry community and the computer vision community [4]. The process relies on having an accurate geometrical model of the object to be located, and also knowing the parameters of the imaging system. With this knowledge, the size and shape of the object in the image gives information about its position and orientation relative to the imaging sensor. For example, the apparent size of the object in the image is proportional to its distance to the sensor.

The principles of the photogrammetric techniques are shown in Figure 2. Each point that is imaged under perspective projection contributes two equations: one for its u image coordinate and one for its v image coordinate. There are a total of 6 unknowns for the pose of the object. Thus, a minimum of 3 (non-collinear) points is required to solve for the pose of the object. Additional points may be used for

accuracy, and to eliminate ambiguous solutions [5], in which case an optimization technique is used to find the minimum error solution.

Since knee prosthetic components have a known geometrical shape and the fluoroscope image is a true perspective projection, it is possible in principal to recover all six degrees of freedom (DOF) of the object. Banks and Hodge [6, 7] measured the full six-DOF motion of knee prostheses by matching the projected silhouette contour of the prosthesis against a library of shapes representing the contour of the object over a range of possible orientations. They measured the accuracy of the technique by comparing it to known translations and rotations of prostheses “in vitro”. They report 0.74 degrees of rotational accuracy (in all three axes), 0.2 mm of translational accuracy parallel to the image plane, and 5.0 mm of translational accuracy perpendicular to the image plane. Our process is similar to the Banks and Hodge method. The difference is that Banks and Hodge represent their silhouettes with Fourier descriptors where we work with the actual image. We also utilize larger image libraries to increase the resolution of the angular measurements.

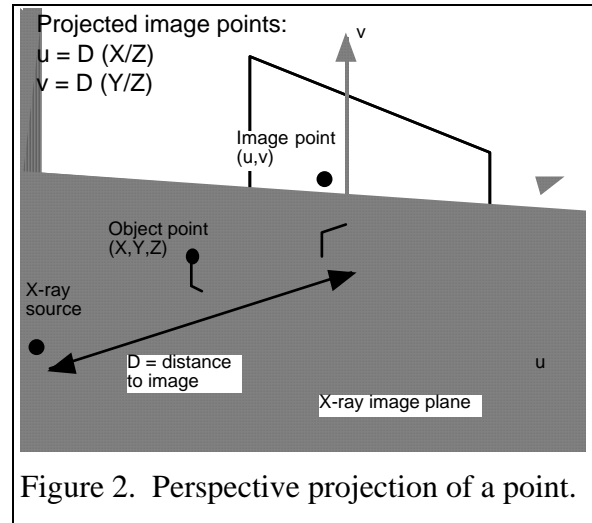


Figure 2. Perspective projection of a point.

Process Description

The central idea in our process is that we can determine the pose of the object (i.e., a knee implant component) from a single perspective image by measuring the size and shape of its projection in the image. One technique for doing this is to match the silhouette of the object against a library of synthetic images of the object, each rendered at a known position and orientation. The image with the best match would directly yield the position and orientation of the object in the input image. However, a library which encompasses all possible rotations and translations would be prohibitively large. As an example, suppose we want to determine an object’s pose within 5 mm and 2 degrees, and the allowable range is 50 mm translation along and 20 degrees rotation about each of the axes. Dividing each range by the resolution results in a 11^6 or 1,771,561 entry library.

To reduce the size of the required library, we use a simplified perspective model, as is done by Banks and Hodge. This model assumes that the shape of an object’s silhouette remains unchanged as the object is moved towards or away from the imaging sensor. This is not strictly true because the fluoroscope is a true perspective projection. However, it is a reasonable approximation if this translational motion is small. In the case of a subject flexing their knee in the sagittal plane, the translation out of the sagittal plane is in fact typically small.

With the simplified perspective assumption, the shape of the object is independent of its distance from the imaging sensor (although its size is dependent on the distance). Therefore, we generate a library of images of the object, all rendered at a constant (nominal) distance from the sensor. When we process an

input image, we correct for any difference in distance by scaling the size of the unknown object silhouette so that its area is equal to the area of the library silhouettes.

The library of images consists of views of the object rendered at different out-of-plane rotations. The object is rotated at one degree increments about the x axis, and at one degree increments about the y axis. The object is always centered in the middle of the image. The object is always rotated within the plane of the image so that its principal axis is aligned with the horizontal (x) axis. Thus, the library is only two dimensional rather than 6 dimensional. The range of rotation is ± 15 degrees about the x axis, and ± 20 degrees about the y axis. Figure 3 shows a portion of the library of images for a femoral component. There are a total of $41 \times 31 = 1271$ images in each library.

Creating the Library

The software modeling program AutoCAD™ was used to create 3-D models of the implant components. The default mode of AutoCAD is an orthographic view. For the library images a perspective view is required. To perform this a dynamic view is set up which imitates the view produced by the fluoroscopy unit. The settings for this viewpoint can be determined by calibration techniques noted in most computer vision texts. With this model a library of images of known position and orientation was created (Figure 4).

We then convert the rendered images into canonical form. Canonical form is achieved by scaling the silhouette to a specified area (15000 pixels) and rotating the object so that its principal axis is aligned with the x axis. The amount of scaling and rotation is recorded, for later use by the pose estimation algorithm. Finally, the library images are converted to binary form (one bit per pixel), and stacked to form a single multi-frame image. The size of a library for a single component is about 3 Mbytes of data.

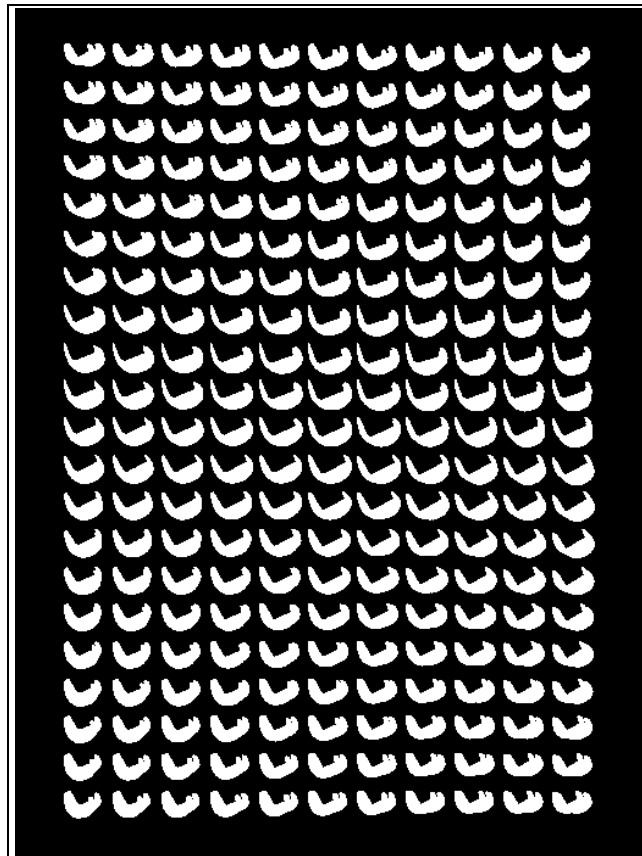


Figure 3. Portion of library for femoral implant component.

The size of a library for a single component is about 3

Pose Estimation

Analysis of a fluoroscopy video begins with digitizing selected images from the sequence, using a Silicon Graphics workstation equipped with a Galileo imaging board. These images are then stretched to equalize the horizontal and vertical scale. The images are then input to the software that extracts the silhouette of the component and estimates its pose.

The image processing algorithms described in this paper were implemented using a public domain image processing software package called Khoros, from Khoral Research Inc. [8]. Khoros is actually a large collection of programs, integrated with a visual programming environment called “cantata.” Khoros is extensible and many new programs were developed and integrated during the course of this work. Figure 5 shows the cantata workspace that performs the silhouette extraction and pose estimation. Each box, or “glyph,” within the workspace performs an operation on an image (or images), and the lines between the glyphs show the transfer of data. We will describe the processing of an image by following the flow of data between the glyphs in this figure.

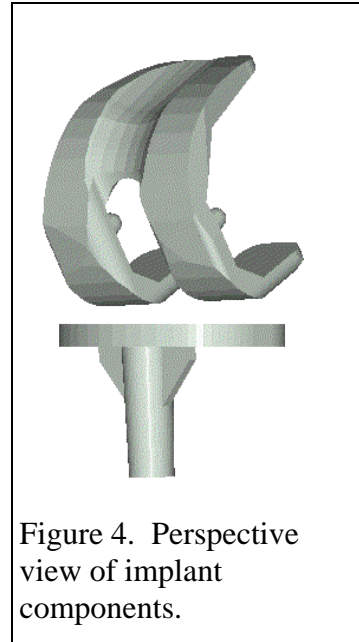


Figure 4. Perspective view of implant components.

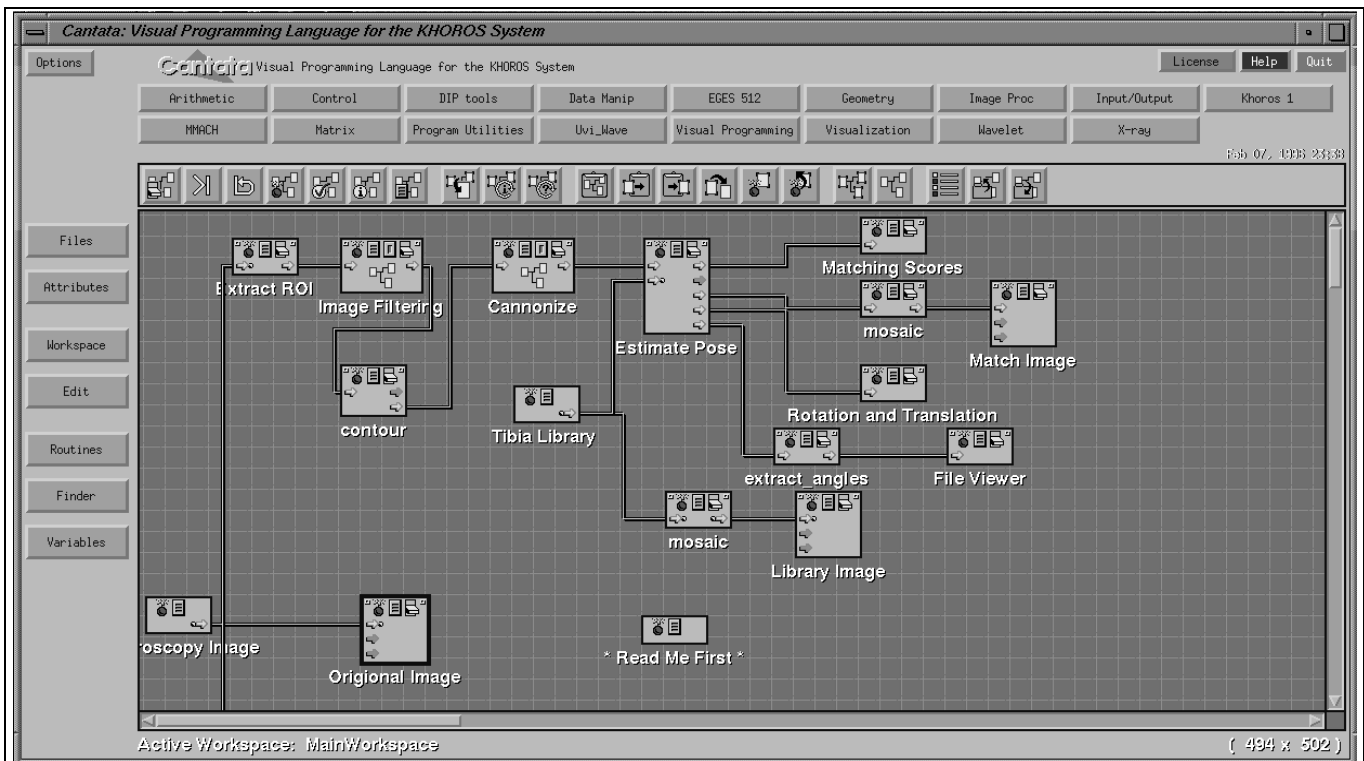


Figure 5. Khoros workspace, showing data flow of the algorithm.

We begin with the glyph in the lower left corner, which represents the input fluoroscopy image. This image is transferred to the glyph titled “Extract ROI”. When executed, this glyph allows the user to manually designate a rough region of interest (a sub-image) surrounding the component of interest. This is done to reduce the size of the image to be processed and speed up the computation time. Next, the reduced region of interest image is passed to the glyph titled “Image Filtering”, which applies a median filter to reduce the effect of noise.

Next is an operation to extract the contour (or silhouette) of the implant component. Currently, this is done manually, by having the user click on points around the boundary of the implant. The resulting binary image is passed to the glyph titled “Canonize”, which converts the silhouette image to a canonical form. As described earlier, the canonization process centers the silhouette, scales it to a constant area, and rotates it so that its principal axis is aligned with the horizontal (x) axis. The resulting canonical image is suitable for direct matching against the library.

The glyph called “Pose estimation” finds the best match of the input canonical image with a library image. This is done by systematically subtracting the input canonical image with each of the library images and generating a “score” for each. The “score” is the number of unmatched pixels. Figure 6 shows the matching results for a particular image. The black areas indicate the unmatched pixels. The library image with the best match determines the two out-of-plane rotation angles of the object, θ_x and θ_y .

We then find the remaining degrees of freedom of the object. The in-plane rotation angle θ_z is determined by taking the difference between the input image’s in-plane rotation angle and the library image’s in-plane rotation angle:

$$\theta_z = \theta_z^{input} - \theta_z^{library}$$

The Z position of the object is determined by dividing scale of the fluoroscopy image by the scale of the library and multiplying that by the initial z distance that the library image was rendered:

$$Z = Z_{library} \cdot (s_{input} / s_{library})$$

To determine the x, y position of the object, we compute the 2D image vector from the projected origin of the object to its image centroid. In the library image, this vector is given by:

$$(c_x^{library}, c_y^{library}) = (r_x^{library} - p_x^{library}, r_y^{library} - p_y^{library})$$

where $(r_x^{library}, r_y^{library})$ is the image location of the object centroid in the library image and $(p_x^{library}, p_y^{library})$ is the

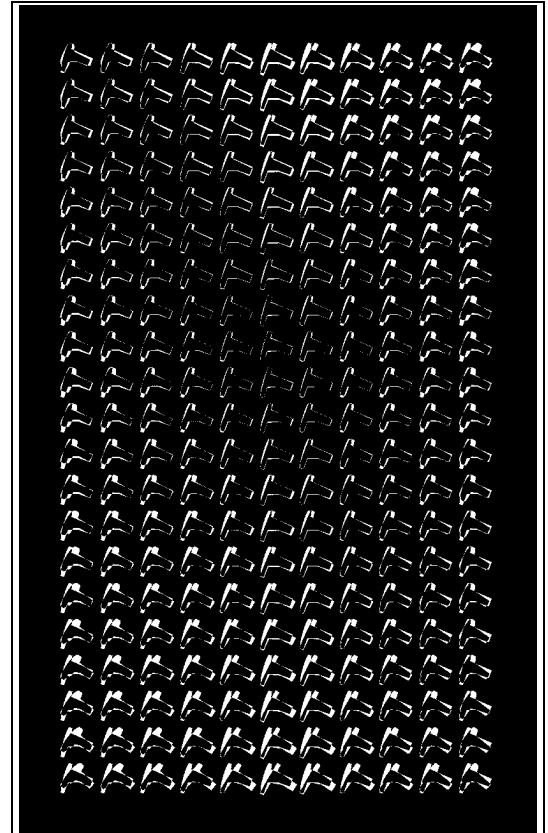


Figure 6. Matching results.

principal point in the library image.

From here the vector must be transformed to what it is in the input image. First we rotate this vector in the plane of this image by the angle θ_Z :

$$\begin{aligned} c_X^{input} &= c_X^{library} \cdot \cos(\theta_Z) - c_Y^{library} \cdot \sin(\theta_Z) \\ c_Y^{input} &= c_X^{library} \cdot \sin(\theta_Z) + c_Y^{library} \cdot \cos(\theta_Z) \end{aligned}$$

Then the vector is scaled by the appropriate amount:

$$\bar{\mathbf{c}}^{input} = \bar{\mathbf{c}}^{input} \cdot (s_{library} / s_{input})$$

The origin of the object in the input image is thus located at:

$$(q_X^{input}, q_Y^{input}) = (r_X^{input} - c_X^{input}, r_Y^{input} - c_Y^{input})$$

where $(r_X^{input}, r_Y^{input})$ = image location of the object centroid in the input image. To calculate the (x,y) location of the object (in meters) relative to the sensor's coordinate frame, we use similar triangles to scale the 2D image vector by the known distance to the object. The (x,y) location of the object's origin is:

$$\begin{aligned} X &= Z \cdot (q_X^{input} - p_X^{input}) / f \\ Y &= Z \cdot (q_Y^{input} - p_Y^{input}) / f \end{aligned}$$

where:

- Z = object position (already calculated)
- $(q_X^{input}, q_Y^{input})$ = object's origin location in pixels in the image
- $(p_X^{input}, p_Y^{input})$ = input image principal point in pixels
- f = focal length in pixels

Finally, we correct the θ_X and θ_Y rotation angles to take into account the effect of x,y translation on the apparent rotation. The result is the full 6 DOF pose of the object (X,Y,Z, θ_X , θ_Y , θ_Z) relative to the sensor frame.

As a check, we can overlay the CAD model of the implant back onto the original X-ray image, using the derived pose data and the known model of the imaging system. The model should fit the actual image silhouette closely. Figure 1(b) shows an example of the CAD models for the femoral and tibial components overlaid on the original image.

RESULTS

The femoral and tibial components have two distinctly different shapes. The femoral component silhouettes are generally rounded, with condylar pegs and condylar "runners" that change shape significantly with x and y rotations. Femoral features are generally large and rounded, and therefore the overall shapes are affected only slightly with added contour noise. The tibial components have more sharper features because of their "T" shaped form with tapered stems. The tibial plateaus are

symmetric with respect to the sagittal (x-y) plane, while the femoral components are usually slightly asymmetric. These differences lead to inconsistencies in the performance of the pose estimation process with respect to different axes.

The accuracy of the process has been determined using synthetic images. The images were renderings of the 3-D CAD models in pre-determined poses. The average accuracy of each component is listed in the table below.

| | Femoral Component | Tibial Component |
|----------------------------------|-------------------|------------------|
| In-Plane Rotational Error | $\pm 0.12^\circ$ | $\pm 0.2^\circ$ |
| Out-of-Plane Rotational Error | $\pm 0.8^\circ$ | $\pm 0.52^\circ$ |
| In-Plane Translational Error | ± 0.24 mm | ± 0.20 mm |
| Out-of-Plane Translational Error | ± 1.24 mm | ± 0.36 mm |

CONCLUSIONS

The pose estimation process is a great aid for determining “in vivo” knee kinematics in implanted knees. When this process is run on successive images at small enough flexion increments the end result is a smooth motion which can be viewed from any angle. The pose estimation process can help to find liftoff, internal rotation, sliding and contact position for any flexion angle.

One of the limitations is that the entire process requires a significant amount of human interaction for development of libraries, and external knowledge of the implants for contour extraction. The library creation could be further automated by implementing AutoLisp routines to rotate and render the library images. The contour extraction is the area where most of the error can be attributed. This is due to human variability in picking the vertices for the contour. An interactive thresholding program for segmentation of x-ray images into binary images, allowing value adjustments to ensure implant components are separated from the background would reduce this error.

The process described in this paper has been used to determine anterior/posterior position along with relative pose at different flexion angles. The data collected using this process has been a large improvement over previous studies using external markers which have been used to estimate “in vivo” motion.

REFERENCES

- [1] M. M. Landy and P. S. Walker, “Wear of ultra-high molecular weight polyethylene components of 90 retrieved knee prostheses,” *Journal of Arthroplasty*, Vol. 3, No. pp. s73-s85, 1988.
- [2] R. L. Perry, “Principles of conventional radiography and fluoroscopy,” *Veterinary Clinics of North America: Small Animal Practice*, Vol. 23, No. 2, pp. 235-252, 1983.

- [3] J. B. Stiehl, R. D. Komistek, D. A. Dennis, W. A. Hoff, R. D. Paxson, and R. W. Soutas-Little, "Kinematic analysis of the knee following posterior cruciate retaining total knee arthroplasty using fluoroscopy," *Journal of Bone and Joint Surgery-British*, Vol. 1996.
- [4] R. Haralick and L. Shapiro, *Computer and robot vision*, Addison-Wesley Inc, 1993.
- [5] W. J. Wolfe, D. Mathis, C. W. Sklair, and M. Magee, "The perspective view of three points," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 1, pp. 66-73, 1991.
- [6] S. A. Banks and W. A. Hodge, "Direct measurement of 3D knee prosthesis kinematics using single plane fluoroscopy," *Proc. of Orthopedic Research Society*, pp. 428, 1993.
- [7] S. A. Banks and W. A. Hodge, "Comparison of dynamic TKR kinematics and tracking patterns measured "in vivo"," *Proc. of Orthopedic Research Society*, pp. 665, 1994.
- [8] J. Rasure and Kubica, "The Khoros application development environment," in *Experimental Environments for Computer Vision and Image Processing*, Vol. H. I. Christensen and J. L. Crowley, Eds., World Scientific, pp. 1994.