

MULTI-VIEW STEREO RECONSTRUCTION OF TOTAL KNEE REPLACEMENT FROM X-RAYS

Kush R. Varshney^{a,b}, Nikos Paragios^b, Alain Kulski^c, Remy Raymond^d, Phillipe Hernigou^c, and Alain Rahmouni^d

^a Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA

^b Laboratoire de Mathématiques Appliquées aux Systèmes, École Centrale Paris, Châtenay-Malabry, France

^c Service de Radiologie et d'Imagerie Médicale, ^d Service d'Orthopédie, Hôpital Henri Mondor, Créteil, France

ABSTRACT

In total knee replacement surgery, also known as total knee arthroplasty, prosthetics are implanted in the knee joint as treatment for progressive diseases such as arthritis or trauma. In this paper, we aim to recover the 3-D shape of bones and prosthetic devices in patients who have undergone total knee replacement. Such an objective is addressed using a collection of 2-D X-rays acquired from multiple viewpoints. Then, both bone and prosthetic shape recovery are formulated as a multi-view reconstruction process through the minimization of an objective function. The objective function is designed to recover 3-D surfaces, which once projected to 2-D images, minimize edge-based and region-based criteria. The calculus of variations and gradient descent are used to obtain a surface with minimum cost through a 3-D level set implementation. Promising results are given for real X-ray data.

Index Terms— orthopedics, X-ray imaging, multi-view stereo reconstruction, 3-D level set methods

1. INTRODUCTION

Total knee replacement, a surgical procedure also known as total knee arthroplasty, places prosthetics in the knee joint as treatment for osteoarthritis most commonly, but also as treatment for other progressive diseases or for trauma to the joint. A tibial plate with stem and a curved femoral component make up the prosthetic. Post-operatively, it is of clinical interest to observe the joint in a non-invasive manner, both statically and dynamically.

To obtain images of the human skeletal structure, a fast, low-dose, relatively inexpensive, and effective technique is X-ray radiography (plain film and fluoroscopic). Statically, a collection of X-ray images from different viewpoints around the knee may be acquired using a rotating sensor. Dynamically, a movie sequence of X-rays from one viewpoint may be acquired while the patient flexes the joint. Prosthetics used today are able to reproduce a large part of the very complicated kinematics of the knee, but not fully. By tracking the three-dimensional (3-D) movement of the bones and prosthetics dynamically, it is possible to understand the frictions and forces at work in the post-operative joint.

Such 3-D tracking with the available X-ray data may be decomposed into two phases: first, determining a 3-D shape model of the bones and prosthetics using the collection of two-dimensional (2-D) X-ray images taken from different viewpoints, and second, performing rigid body 3-D/2-D registration and tracking using the 3-D model from the first phase and the single viewpoint movie sequence [1]. The focus of this paper is on the first phase of the problem, approached as stereoscopic reconstruction from many views using a surface evolution technique implemented with 3-D level sets.

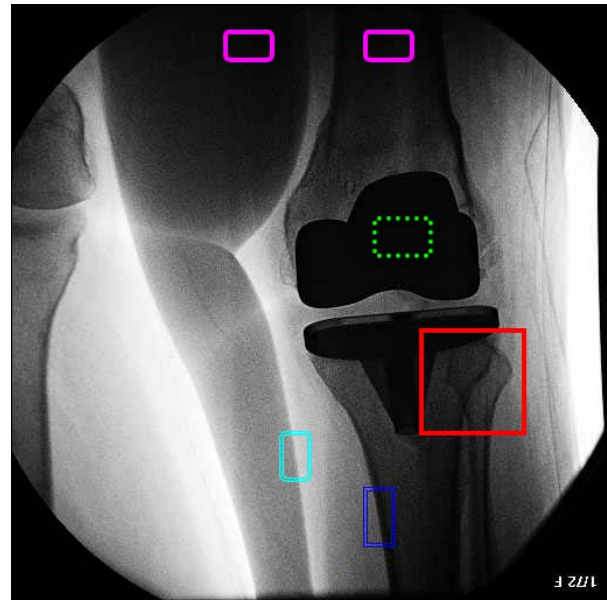


Fig. 1. X-ray image of a knee joint that has a prosthetic. In the red rectangle (solid line), one bone behind another does not result in occlusion. In the green rectangle (dotted line), the prosthetic has no texture whatsoever. In the blue rectangle (double line), a strong edge delimits the bone, but in the cyan rounded rectangle (double line), there is another strong edge in the background. The two magenta rounded rectangles (solid line), one inside the object of interest and the other outside have almost identical intensities.

3-D surface reconstruction of bones using 3-D X-ray computed tomography images is well-studied, but the problem using 2-D X-ray images from different viewpoints is not so well-studied. The only methods of which we are aware involve the morphing and deforming of a fairly accurate prior shape model [2].

The general problem of 3-D shape estimation from a set of 2-D images, however, has received attention in the computer vision community. A large body of literature exists focused on the problem with exactly two images taken from very similar viewpoints (like the human vision problem). The problem we encounter, with a large number of images from disparate viewpoints is the *multi-view stereo reconstruction* problem. Multi-view stereo reconstruction has traditionally been posed in the context of optical imaging for objects such as statuettes in computer vision [3]. Consequently, assump-



Fig. 2. X-ray images from rotating sensor of knee with prosthetic.

tions relating to radiance, occlusion, etc. have been built in that do not necessarily apply when dealing with X-ray imaging.

Let us look at a typical X-ray image, shown in Fig. 1. First of all, getting oriented, the two dark pieces are the femoral and tibial components of the prosthetic. The other leg can be seen on the periphery. Many characteristics are apparent that differentiate X-ray imaging from optical imaging (upon which multi-view stereo reconstruction techniques are based).

Focusing on the red rectangle (solid line), the tibia would occlude the fibula in an optical image, but the X-ray modality shows both bones with pixel intensities darkened in the overlap region. The effect is neither transparency nor shadowing, but can be interpreted similarly. Another related feature seen in the image is that the boundaries of the bones are darker than the centers. This does not change image to image, so a point on the surface of a bone may appear dark in one image and light in an image from a different viewpoint.

Multi-view stereo techniques based on local correspondence rely on the assumption that a point on the surface appears the same in all of the images in which it is visible and that radiance is locally computable. The idea is to find the shape such that the local correspondence error is minimized [4, 5]. Due to the phenomena exhibited by the X-ray modality, local correspondence is not a reliable feature for 3-D reconstruction of bone.

Now, looking at the prosthetic components and specifically the region outlined in green (dotted line), it can be seen that there is no texture at all; consequently, local correspondence is ill-posed for reconstructing the shape of prosthetics. The prosthetics are homogeneous in image intensity, but the bones are not. The background is full of clutter — in some places, for example within the two magenta rounded rectangles (solid line), the image intensities inside and outside the bone are very similar.

Variational methods with a region-based flavor have been applied to multi-view stereo reconstruction as an alternative to local correspondence [6]. In this formulation, a key ingredient is that the background is treated like ‘blue sky,’ i.e. it is homogeneous, behind everything, far away, and different from the object. Also, the surface of the object is assumed to be homogenous in some respect. Unlike local correspondence, there is no comparison of points between pairs of images for consistency, but instead the comparison is integrated over regions. Inhomogeneity in the bones, but more significantly, the background clutter and lack of occlusion are barriers to the direct application of such techniques to X-ray imaging.

Another method for multi-view stereo, given the name ‘shape from silhouette,’ is well matched to the prosthetics and can be applied readily. In this technique, the silhouette of the object is projected back in 3-D for each image as a cylinder with the silhouette

as its cross-section; the intersection of all of the cylinders — the visual hull — is then the estimated 3-D shape [7]. The estimate of the shape is conservative because it is the largest shape that meets the constraints set by the images. It takes no prior information, for example regarding smoothness or geometry, into account.

Edges, such as seen inside the blue rectangle (double line), stand out in X-ray images. Edges have not been exploited enough in previous work on multi-view reconstruction. Unfortunately, the edges of bones and prosthetics are not the only edges in the image; the edge inside the cyan rounded rectangle (double line) is caused by the boundary of the overlapping region of soft tissue from the two legs. Nevertheless, the use of edge features is an avenue that we pursue for this challenging application, combined with some region-based features and an approach inspired by shape from silhouette.

2. MULTI-VIEW STEREO FROM X-RAYS

We now present an approach for 3-D shape reconstruction motivated specifically by the appearance of bones and prosthetics in X-rays. Our variational method extends the 2-D geodesic active regions (GAR) functional to three dimensions by projecting cylinders in a manner similar to the visual hull calculation. Optimization is by surface evolution implemented using 3-D level sets.

2.1. Notation

A set of 2-D images $\mathcal{I} = \{I_1, I_2, \dots, I_N\}$ with corresponding 2-D domains $\Omega_1, \Omega_2, \dots, \Omega_N$ is given in the problem as input. Each Ω_i has local image coordinates (u_i, v_i) . The goal is to determine the 3-D solid enclosed by a surface S that is depicted in the images of \mathcal{I} . The surface S is in \mathbb{R}^3 with global Cartesian coordinates (x, y, z) , or alternatively global cylindrical coordinates (r, θ, z) .

The relationship between the global coordinates (x, y, z) and the local image coordinates (u_i, v_i) is assumed known or known approximately, i.e. the views or cameras are calibrated. These relationships are given by projections $\pi_i : \mathbb{R}^3 \rightarrow \Omega_i$. The mappings are not invertible in general because many different points in \mathbb{R}^3 project onto the same point in Ω_i . The coordinate axis in 3-D perpendicular to the plane Ω_i is taken to be w_i . Within a 2-D plane Ω_i , a curve C_i is parameterized by a variable $s_i \in [0, 1]$ and has a line element ds_i . If the context requires only one 2-D plane, then the subscript i may be dropped.

We have already seen one data image in Fig. 1; two more images are shown in Fig. 2. The images, acquired using a rotating sensor, have viewpoints that are in a ring outside the leg. Each Ω_i plane is parallel to the global z -axis and forms an angle θ_i to the x - z plane, which is known approximately. Parallel projection is a valid assumption in X-ray images; hence, we take π_i to be such that $u_i = x \cos \theta_i + y \sin \theta_i$ and $v_i = z$.

2.2. Geodesic Active Regions

We first describe segmentation into regions R , the inside, and R^c , the outside, using GAR in 2-D before taking the leap to 3-D. The GAR functional is the convex combination of two terms, the geodesic active contours (GAC) functional and a region-based functional [8]:

$$E_{\text{GAR}}(C) = \alpha E_{\text{GAC}}(C) + (1 - \alpha) E_R(C). \quad (1)$$

The GAC functional has minima where the curve C falls along strong edges in an image I [9]:

$$E_{\text{GAC}}(C) = \oint_C g(C(s)) ds, \quad g(I) = \frac{1}{1 + |\nabla I|^p}, \quad (2)$$

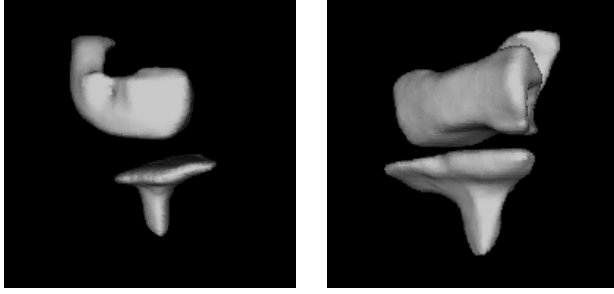


Fig. 3. Two views of the reconstructed 3-D shape of the prosthetic (femoral component and tibial plate with stem).

with $p \in [1, 2]$. To prevent local minima and fractally solutions, oftentimes a curve length penalty is also included in the GAC functional:

$$E_{\text{GAC}}(C) = \oint_C g(C(s)) ds + c \oint_C ds. \quad (3)$$

The region-based portion assumes some prior knowledge regarding the image intensities of R and R^c . It is assumed that pixel values are independent given the region label and have probability distribution function $p_R(I(u, v))$ or $p_{R^c}(I(u, v))$. The functional is essentially a log likelihood ratio:

$$E_R(C) = - \iint_R \log(p_R(I(u, v))) du dv - \iint_{R^c} \log(p_{R^c}(I(u, v))) du dv.$$

Starting from an initial curve, a curve evolution approach is taken to flow towards a minimum of the functional. Using a level-set implementation with the signed distance function $\varphi(u, v; t)$, the level-set update equation is $\varphi_t = F|\nabla\varphi|$, with:

$$F(u, v) = \alpha \left(\kappa(g(I(u, v)) + c) - \left\langle \nabla g(I(u, v)), \frac{\nabla\varphi(u, v)}{|\nabla\varphi(u, v)|} \right\rangle \right) + (1 - \alpha) (-\log(p_R(I(u, v))) + \log(p_{R^c}(I(u, v)))) ,$$

where κ is the curvature of C , and $\langle \cdot, \cdot \rangle$ indicates the inner product.

2.3. Multi-View GAR Surface Evolution

Now we come back to the problem at hand in 3-D. The idea is to have one surface in 3-D and evolve that surface based on information provided by all of the images. All points on a line in 3-D map to a point (u_i, v_i) in Ω_i . Our approach is to apply the GAR flow at (u_i, v_i) to all points (x, y, z) that project to it. The overall force applied to a point (x, y, z) is the superposition of forces from all N images.

We construct a multi-view geodesic active regions functional with a GAC term and a region-based term. The GAC portion of the functional is:

$$E_{\text{GAC-MV}}(S) = \sum_{i=1}^N \oint_{C_i} g(C_i(s_i)) ds_i + c \oint_{C_i} ds_i. \quad (4)$$

Making the assumption that the shape S has a pixel intensity distribution p_R when seen in an image, and the background has a distribution p_{R^c} , the region-based portion is:

$$E_{R\text{-MV}}(S) = - \sum_{i=1}^N \iint_{\pi_i(S)} \log(p_R(I_i(u_i, v_i))) du_i dv_i - \sum_{i=1}^N \iint_{\Omega_i - \pi_i(S)} \log(p_{R^c}(I_i(u_i, v_i))) du_i dv_i. \quad (5)$$

Using a 3-D signed distance function $\varphi(x, y, z; t)$ the 3-D level set update equation takes the form $\varphi_t = F|\nabla\varphi|$, with:

$$F(x, y, z) = \alpha \sum_{i=1}^N \kappa(g(I_i(u_i, v_i)) + c) - \left\langle \nabla_i g(I_i(u_i, v_i)), \frac{\nabla_i \varphi(x, y, z)}{|\nabla_i \varphi(x, y, z)|} \right\rangle + (1 - \alpha) \sum_{i=1}^N -\log(p_R(I_i(u_i, v_i))) + \log(p_{R^c}(I_i(u_i, v_i))). \quad (6)$$

By ∇_i , we mean a gradient with respect to the u_i and v_i axes of Ω_i . The force has effect only in the direction normal to S in 3-D. Explicitly, $\frac{\partial \varphi(x, y, z)}{\partial u_i} = \frac{\partial \varphi(x, y, z)}{\partial x} \cos \theta_i + \frac{\partial \varphi(x, y, z)}{\partial y} \sin \theta_i$, and $\frac{\partial \varphi(x, y, z)}{\partial v_i} = \frac{\partial \varphi(x, y, z)}{\partial z}$.

The surface evolution couples information provided by each image in \mathcal{I} . Individual 2-D functionals for each image in \mathcal{I} extend back along the w_i axis in a cylinder of influence in 3-D. It is in this way that the approach relates to visual hull calculation. Occlusions are not modeled for the reasons discussed in Sec. 1.

2.4. Prosthetic Reconstruction

We have described our surface evolution procedure for multi-view stereo reconstruction from X-rays above. We now apply it to the problem of recovering the 3-D shape of the femoral component and tibial plate of the prosthetic, treating soft tissue, bone, and air as background. Our data is $N = 16$ X-ray images with an approximate spacing of 0.038 radians in θ ; the images are as in Fig. 1 and Fig. 2.

We do not require any special initialization — we use a centered cube that projects to cover about half of an image as an initial surface. The probabilities p_R and p_{R^c} must also be specified. Although not completely accurate, we take the pixels to be i.i.d. Gaussian with prespecified mean μ_R for the prosthetics and i.i.d. Gaussian with prespecified mean μ_{R^c} for the background.

Two views of the prosthetic from the surface estimate using our procedure are shown in Fig. 3. The figure shows a triangulated surface obtained from the implicit level set representation by the marching cubes method. In the final 3-D reconstruction, the two pieces of the prosthetic have been recovered; they are correct visually. By lifting the geodesic active regions functional to the multi-view 3-D context, the homogeneous prosthetics are reconstructed successfully.

2.5. Bone Reconstruction

The pixel intensity in X-ray images of bones is not homogeneous throughout the bone, but follows a predictable pattern. The boundary is dark and the shading gets lighter as the distance away from the boundary increases. In other words, pixel intensity values generally increase as a function of signed distance.

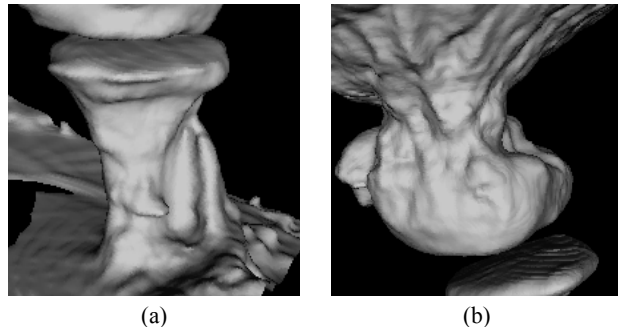


Fig. 4. Reconstructed shape of the (a) tibia and fibula, and (b) femur.

When reconstructing the 3-D shape of bones, we set p_R to account for this phenomenon. We once again take p_R to be Gaussian and independent among different pixels, but not identically distributed. We put in a spatially varying mean $\mu_R(\varphi_i)$ that is a function of signed distance.

The signed distance function $\varphi(x, y, z)$ of the surface evolution is in 3-D, but we need distances in the Ω_i domains. Thus, the 3-D signed distance is projected down to the plane Ω_i as $\varphi_i(x, y, z)$ for each i . Specifically:

$$\varphi_i(x, y, z) = \varphi(x, y, z) \sin \left(\frac{\frac{\partial \varphi(x, y, z)}{\partial x} \cos \theta_i + \frac{\partial \varphi(x, y, z)}{\partial y} \sin \theta_i}{|\nabla \varphi(x, y, z)|} \right).$$

In this work, we use a simple predetermined functional form for $\mu_R(\varphi_i)$ that approximates the pixel intensity values in the images.

Treating the bone and prosthetic as the object of interest and everything else as background, we obtain 3-D shapes such as those in Fig. 4. The image on the left shows the tibia and fibula, whereas the image on the right shows the femur. The circular fields of view in the images leave an artifact of a thin layer along the boundary as part of the surface. The tibia and fibula are not completely separated because they cannot be distinguished in X-rays from certain angles. Obtaining the shape of the bones is more difficult than obtaining the shape of the prosthetic; the results show promise however.

3. CONCLUSION

In this paper, we have looked at a challenging problem in computer vision with an important clinical application. Reconstructing the 3-D shape of metal prosthetics, but even more so of bones from multiple X-ray images is not straightforward due to the fact that radiance is not locally computable, the prosthetics exhibit no texture, the bones are not homogeneous, and the background is cluttered. We have proposed an approach using active contours, 3-D surfaces in particular, that uses edge and region information in the 2-D data images to direct surface evolution.

The procedure can be considered similar to constructing the visual hull from silhouettes, but has a coupling effect among images, has regularization built in, and offers an opportunity to incorporate prior information in both 2-D and 3-D. To account for the lack of homogeneity in the bones, a distribution of image intensity as a function of distance from the surface has been incorporated.

The results shown in Fig. 3 and Fig. 4 indicate that the approach is promising. The method may be enhanced by including separate

partitions for three classes: prosthetic, bone, and background, via multi-phase segmentation, i.e. segmentation with more than two categories. It may also be enhanced by including shape priors and by improving the pixel intensity models p_R and p_{R^c} via learning from data [10] or using a generative model based on X-ray absorption. Instead of using the 2-D GAR functional as a foundation for the 3-D problem, an approach that jointly optimizes pixel intensity could also be used [11].

As mentioned in Sec. 1, the work in this paper may be viewed as a first stage of a larger system that tracks the dynamics of the knee joint in 3-D, allowing inference of the velocities, accelerations, and other mechanical properties affecting it. 3-D/2-D tracking is a difficult problem which will be complicated by the same peculiarities of X-ray imaging discussed, as well as others such as motion blur.

X-ray imaging is an exciting, clinically important, and challenging domain that brings forth issues for computer vision methods not encountered with other types of imaging. We have started to attack the issues here, and hope to continue doing so in future work.

4. REFERENCES

- [1] Á. Czopf, C. Brack, M. Roth, and A. Schweikard, “3D-2D registration of curved objects,” *Periodica Polytechnica Electrical Engineering*, vol. 43, no. 1, pp. 19–41, 1999.
- [2] G. Zheng and L.-P. Nolte, “Surface reconstruction of bone from X-ray images and point distribution model incorporating a novel method for 2D-3D correspondence,” in *Proc. IEEE CVPR*, vol. 2, New York, June 2006, pp. 2237–2244.
- [3] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *Proc. IEEE CVPR*, vol. 1, New York, June 2006, pp. 519–526.
- [4] K. N. Kutulakos and S. M. Seitz, “A theory of shape by space carving,” *Int. J. Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.
- [5] O. Faugeras, J. Gomes, and R. Keriven, “Computational stereo: A variational method,” in *Geometric Level Set Methods in Imaging, Vision, and Graphics*, S. Osher and N. Paragios, Eds. New York: Springer-Verlag, 2003, pp. 343–360.
- [6] A. J. Yezzi, Jr. and S. Soatto, “Stereoscopic segmentation,” *Int. J. Computer Vision*, vol. 53, no. 1, pp. 31–43, 2003.
- [7] A. Laurentini, “The visual hull concept for silhouette-based image understanding,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 2, pp. 150–162, Feb. 1994.
- [8] N. Paragios and R. Deriche, “Geodesic active regions: A new framework to deal with frame partition problems in computer vision,” *J. Visual Communication and Image Representation*, vol. 13, pp. 249–268, 2002.
- [9] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” *Int. J. Computer Vision*, vol. 22, pp. 61–79, 1997.
- [10] M. E. Leventon, O. Faugeras, W. E. L. Grimson, and W. M. Wells, III, “Level set based segmentation with intensity and curvature priors,” in *Proc. Workshop Mathematical Methods in Biomedical Image Analysis*, Hilton Head, South Carolina, June 2000.
- [11] C. V. Alvino and A. J. Yezzi, Jr., “Tomographic reconstruction of piecewise smooth images,” in *Proc. IEEE CVPR*, vol. 1, Washington, DC, June-July 2004, pp. 576–581.