LABORATOIRE MATHEMATIQUES APPLIQUEES AUX SYSTEMES

Surface Evolution for 3-D Shape Reconstruction of Knee Joint Prosthetics and Bones

Kush R. Varshney — Nikos Paragios

N° 0703

Février 2007

Projet Orasis

RAPPORT DE RECHERCHE

TECHNICAL REPORT

Surface Evolution for 3-D Shape Reconstruction of Knee Joint Prosthetics and Bones

Kush R. Varshney* †, Nikos Paragios†

Projet Orasis — Imagerie Médicale et Vision par Ordinateur

Rapport de recherche n° 0703 — Février 2007 — 27 pages

Abstract: The problem of multi-view stereo reconstruction is investigated with an application to medical imaging. In particular, the 3-D shape of bones and prosthetic devices in patients having had total knee replacement surgery is to be recovered using a collection of X-rays acquired from multiple viewpoints. After describing existing methods for multi-view stereo, the assumptions of these techniques are discussed in conjunction with the imaging modality and objects being imaged. A brief overview of level set methods follows. A new surface evolution formulation is presented that uses 2-D edge-based and region-based criteria, and an implementation with level sets. Results are given for synthetic data and real X-ray data.

Key-words: level sets, x-ray imagery, multi-view stereo reconstruction, geodesic active regions

The first author is supported by a National Science Foundation (USA) Graduate Research Fellowship.

* Massachusetts Institute of Technology - Laboratory for Information and Decision Systems
† École Centrale Paris - Laboratoire MAS

Laboratoire MAS
Ecole Centrale Paris, Grande Voie des Vignes
92290 CHATENAY-MALABRY Cedex (France)
Téléphone : +33 1 41 13 17 11 — Télécopie : +33 1 41 13 17 35

Évolution de Surface pour 3-D Reconstruction de Forme de la Prosthétique et des Os du Genou Joignez

Résumé : Le problème de la reconstruction multi-vue stéréo est étudié avec une application à la imagerie médicale. En particulier, la forme 3-D des os et des dispositifs prosthétique dans les patients ayant eu la chirurgie totale de remplacement de genou doit être récupérée en utilisant une collection de rayons X acquis des points de vue multiples. Après avoir décrit des méthodes existantes pour le multi-vue stéréo, les acceptations de ces techniques sont discutées en même temps que la modalité de formation image et les objets étant reflètents. Une brève vue d'ensemble des méthodes de courbes de niveau suit. Une nouvelle évolution de surface formulation est présentée que l'informations de frontière et de région d'utilisations. L'exécution emploie de courbes de niveau. Les résultats sont donnés pour les données synthétiques et le vraies données de rayon X.

Mots-clés : courbes de niveau, imagerie rayon X, reconstruction multi-vue stéréo, régions actives géodésiques

Contents

1 Introduction

Aging populations and degenerative disease such as osteoarthritis go hand in hand. The last option in treating osteoarthritis of the knee is total knee replacement, also known as total knee arthroplasty, a surgery in which artificial prosthetics with components made of metal are implanted into the joint. In the drawings of Fig. 1, the left side shows the knee joint before surgery and the right side shows the knee joint after surgery. Fig. 2 contains a schematic diagram of the two pieces of the prosthetic.

The kinematics of the knee are very complex; today's prosthetics can replicate them to a large degree, but not completely. Patients are living longer after knee replacement and have higher expectations for the prosthetics than ever before. For longevity and functional performance, one factor is precise implant positioning [1]. To characterize the kinematics and positioning of prosthetics in the knee, it is of post-operative interest to observe the joint non-invasively, both statically and dynamically.

One of the oldest forms of medical imaging and the original form of radiology, X-ray or Röntgen ray radiography, is a fast, relatively inexpensive, and effective technique for obtaining images of the human skeletal structure. It is often used in the diagnosis of bone fractures, for example. X-ray radiation is passed through the body; some of the radiation reaches a detector on the other side, while the rest is attenuated. Different materials have different attenuation properties. Air and soft tissue of the body are radiolucent, meaning that they allow X-rays to pass through, whereas bone and metal are radiopaque, meaning that they do not allow X-rays through. The image accompanying the original announcement of the discovery of X-rays by Röntgen, shown in Fig. 3, illustrates these radiopacity ideas. The picture is of a hand wearing a ring; the bones and metal ring have an opposite shading to the soft tissue and air.

In our work, we use a collection of X-ray images taken from different viewpoints around the knee joint to determine a three-dimensional (3-D) model of first, the prosthetic components, and afterwards, of the bones as well. This 3-D shape extraction can be considered as a first stage of a larger system that registers [3] and tracks the prosthetics and bones in a dynamic sequence of two-dimensional (2-D) images of the joint. However, the 3-D shape recovery problem is interesting in and of itself.

One may ask why computed tomography (CT) technology is not used for this task. X-ray radiography is less expensive and exposes the patient to less radiation than CT, but more significantly, streaking artifacts arise in the presence of foreign metal objects when CT volumes are reconstructed using conventional techniques [4]. (Work is ongoing in the development of statistically-based advanced image formation techniques that mitigate artifacts [5].) Even if a 3-D volume of voxels were available, the task of extracting the shape of the prosthetic and bones would still remain. In this light, we approach the problem as stereoscopic reconstruction from many views.

This technical report is organized as follows. Sec. 2 discusses a number of approaches to the problem of reconstructing 3-D shape from 2-D images acquired from different viewpoints, paying particular attention to the assumptions required for their effectiveness and which of these assumptions apply to X-ray images. Sec. 3 provides background material on level set

Figure 1: Before (left) and after (right) drawings of total knee replacement surgery. Image from MedilinePlus Medical Encyclopedia (http://www.nlm.nih.gov/medlineplus/ency/imagepages/9494.htm).

Figure 2: Diagram of knee prosthetic device. Image from Joint Replacement Center at Florida Hospital Celebration Health website (http://www.celebrationjointreplacement.com).

Figure 3: The first X-ray image (1895). Image from Glasser [2].

methods. Sec. 4 comes back to the problem at hand: recovering the shape of prosthetics and bones from X-ray images. Taking the peculiarities of X-ray images into account, an approach using active contours is suggested. The section shows results on synthetic and real data. Sec. 5 provides a brief summary and describes future directions of research.

2 3-D Shape Recovery

Estimating 3-D shape from a set of 2-D images is something humans do every waking second, in particular from the two images produced by our eyes. A number of techniques have been developed in the computer vision community to do the same. A large body of literature exists focused on the problem with exactly two images taken from very similar viewpoints (like the human vision problem). However, in this section we survey approaches to the problem with a large number of images from disparate viewpoints, the so-called *multi-view stereo reconstruction* problem, as this is the problem we encounter in the knee imaging application. A thorough qualitative and quantitative comparison of various techniques may be found in [6]. In this section, a common notation, which is also used in the other sections of the report, is defined first, followed by descriptions of different techniques.

2.1 Notation

A set of 2-D images $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$ with corresponding 2-D domains Ω_1, Ω_2, ..., Ω_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). The space may also be described using cylindrical coordinates (r, θ, z) or spherical coordinates (ρ, θ, ϕ). The area element on the surface is taken to be $d\sigma$.

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 specific form for projections is discussed later. 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.

2.2 Hull-Based Techniques

A number of techniques for reconstructing 3-D shape from 2-D images are based on the concept of the visual hull. The visual hull of a 3-D shape S is the largest shape V for which shape silhouettes in $\pi_i(S)$ and $\pi_i(V)$ match for all i [7]. A silhouette is a binary image that is black for the subset of Ω_i that a 3-D shape maps to and white elsewhere. Thus, the visual hull is a property of S, the viewpoints of $\mathcal{I}$, and the type of projection used. The visual hull always contains the shape and can be tighter than the convex hull of the shape [8].

In [7] and [8], silhouettes of S in $\mathcal{I}$ are used to compute the visual hull; consequently they are effective for texture-less surfaces (silhouettes have no texture). The method of [7] discusses both perspective projection and parallel projection. A cone, in the case of perspective, or cylinder, in the case of parallel, with cross-sections in the shape of the silhouette in I_i is extended out along the w_i axis. Taking the intersection of all N such cylinders or cones produces the visual hull. The approach in [8] operates completely in image spaces Ω for computational reasons, but the underlying theory is along the same lines.

Whereas the visual hull is based on silhouettes that lack texture and shading, a different hull defined in [9], the photo hull, makes use of scene radiance and is a subset of the visual hull. Radiance is an indicator of how bright or what color a surface appears to be when viewed from a particular direction. Under the restriction that radiance is locally computable and thus does not include shadows, transparency, or inter-reflections, the photo hull is defined as the largest collection of points p in $\mathbb{R}^3$ that meet the photo-consistency criterion, to be stated next. Considering a single image, I_i, p is photo-consistent if it is visible, i.e. in the field of view and unoccluded, and the brightness or color $I_i(\pi_i(p))$ matches the radiance. The entire shape is photo-consistent in I_i if all visible points are photo-consistent, and the shape is photo-consistent overall if the radiance function can be assigned such that the shape

is photo-consistent in all images of $\mathcal{I}$. The largest photo-consistent shape is the photo hull. The basic idea is that the shape and radiances of the photo hull should project to match the images $\mathcal{I}$, in much the same way as the shape of the visual hull should project to match silhouettes in the images $\mathcal{I}$. The space carving algorithm of [9] finds the photo hull; a probabilistic version of space carving also exists [10].

Recapitulating, the true shape in $\mathbb{R}^3$ is a subset of the photo hull, which is a subset of the visual hull, where by subset, we do not necessarily mean proper subset. Of course, the goal is to determine the true shape, so taking either hull as an estimate for the true shape is conservative. Using either the visual hull or the photo hull, the recovered 3-D shape is the largest shape that meets the constraints set by $\mathcal{I}$. Thus, it takes no prior information, for example about smoothness or geometry, into account. A different class of methods reconstruct the 3-D shape by finding the shape that minimizes an energy functional which may incorporate prior information into its design. These methods are discussed in the next section.

2.3 Variational Techniques

The photo-consistency conception described above makes use of local information; there is a correspondence among all of the projections of the same global point. The correspondence is directly connected to the restrictions on the radiance being locally computable. A variational method proposed in [11], making similar assumptions about radiance, also looks at local correspondence, giving preference to solutions whose projections exhibit the correspondence property.

In variational techniques, a cost functional is written such that its minimum or maximum expresses the desired goals. The functional proposed in [11] is as follows:

$$E(S) = -\iint_S \sum_{i,j=1, i\neq j}^{N} \frac{\langle I_i, I_j \rangle}{|I_i| \cdot |I_j|} d\sigma, \tag{1}$$

where the $\langle I_i, I_j \rangle$ is a cross-correlation type measure, but specified in such a way that projective geometry and occlusions are taken into account. Thinking like Riemann, as the integral rasters through all of the area elements $d\sigma$, if the current area element corresponds to a global coordinate (x, y, z), then the pixel values chosen from the images are $I_i(\pi_i(x, y, z))$ and $I_j(\pi_j(x, y, z))$, but only if the area is not occluded in that image. There are other technical details in the work, but this description suffices for our purposes.

The point to note is that this correlation-based technique relies on the assumption that a point on the surface should appear the same in all of the images if it is visible. For effectiveness, there should be some texture on the surface because otherwise, all points look the same and the problem is ill-posed. We do not discuss the method of solution here, but it involves finding the first variation of the cost functional and performing an iterative gradient flow using level set methods.

In contrast to the cost functional of [11], which is based on local correspondence, in [12], the functional takes on a more region-based flavor. One key ingredient in the formulation is

that the background is treated like 'blue sky,' i.e. it is homogeneous, behind everything, and far away. Also, the surface of the object is assumed to be homogeneous. The method finds the radiance of the background $\mathbf{g}$, the radiance of the surface $\mathbf{f}$, and of course the surface itself S. The cost functional is the sum of three terms, a geometric term, a smoothness term, and a data term. The geometric term is:

$$E_{\text{geom}}(S) = \iint_S d\sigma, \tag{2}$$

the smoothness term is:

$$E_{\text{smooth}}(S, \mathbf{f}, \mathbf{g}) = \iint_S \|\nabla \mathbf{f}\|^2 d\sigma + \iint_B \|\nabla \mathbf{g}\|^2 d\Theta, \tag{3}$$

and the data term is:

$$E_{\text{data}}(S, \mathbf{f}, \mathbf{g}) = \sum_{i=1}^{N} \iint_{\pi_i(S)} \left(\mathbf{f}\left(\pi_i^{-1}(u_i, v_i)\right) - I_i(u_i, v_i)\right)^2 du_i\, dv_i + \iint_{\pi_i(S)^c} \left(\mathbf{g}\left(\Theta(u_i, v_i)\right) - I_i(u_i, v_i)\right)^2 du_i\, dv_i, \tag{4}$$

where B is the background and Θ is projection to the background. The geometric prior prefers surfaces with small surface area, i.e. smooth surfaces. The smoothness term relates to the radiances $\mathbf{f}$ and $\mathbf{g}$ and the desire that they not vary too much. Unlike the correlation-based method discussed earlier in this section, the data term here does not compare points between pairs of images for consistency, but instead integrates the match of the entire region in the reconstruction to the input images. Like the previous functional, this functional also accounts for occlusions, manifested in the region of integration $\pi_i(S)$. Also as before, the functional is minimized using level set methods, but the details are omitted here. The next section considers the assumptions of the various multi-view stereo methods discussed heretofore in the context of X-ray imaging.

2.4 Existing Techniques and X-Ray Imaging

The multi-view stereo reconstruction problem in computer vision has traditionally been posed in the context of optical imaging for objects such as statuettes. Consequently, assumptions 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. 4; a number of characteristics should be noticed. First of all, getting oriented, the two dark pieces are the two metal components of the prosthetic. The other leg can be seen on the periphery.

Focusing on the red rectangle, the tibia ought to occlude the fibula according to the geometry, but the X-ray modality, which sums values through the line of projection, shows

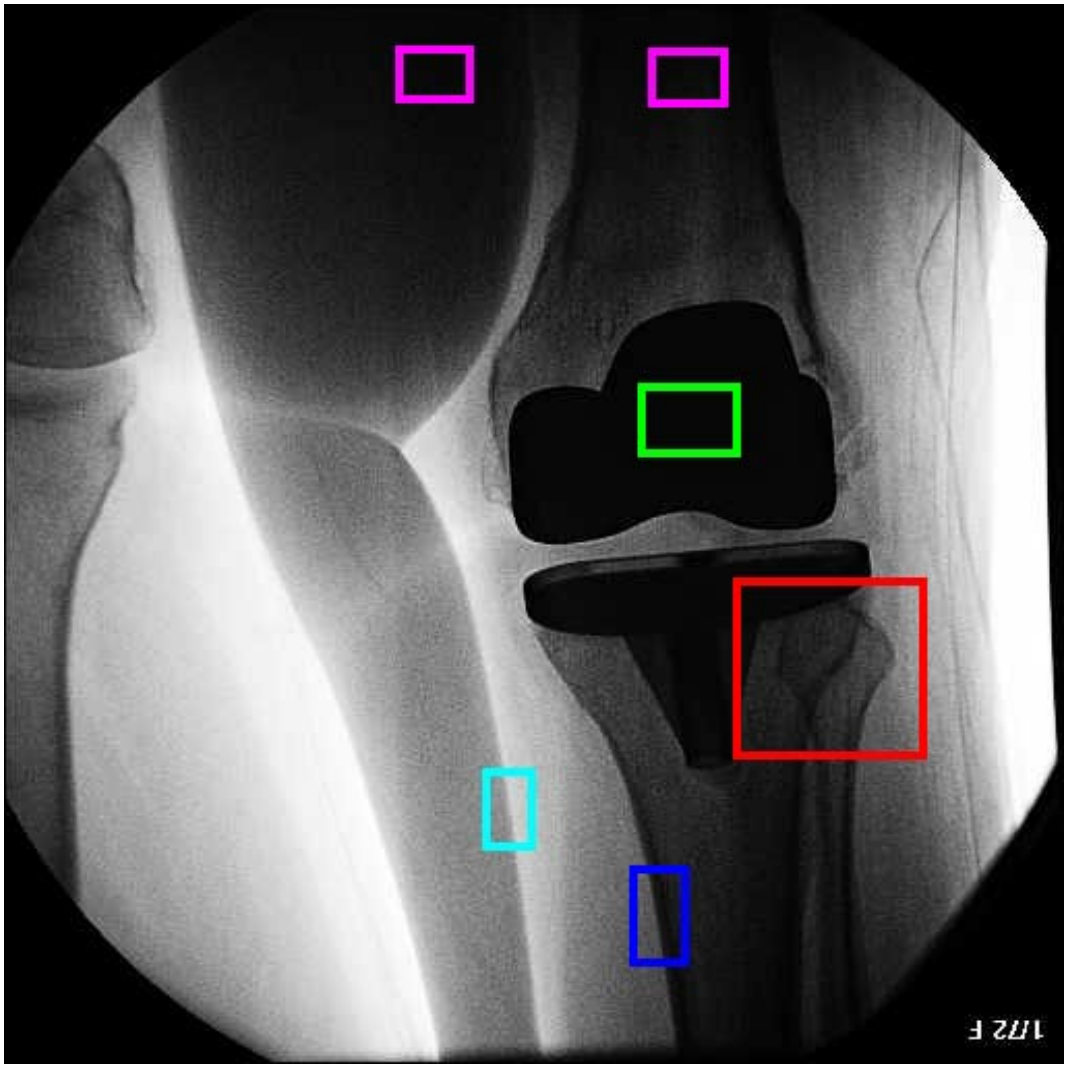


Figure 4: An X-ray image of a knee joint that has prosthetics. The femur (thigh bone) is on the top. Characteristics to notice are highlighted. In the red rectangle, one bone behind another does not result in occlusion. In the green rectangle, the prosthetic has no texture whatsoever. In the blue rectangle, a strong edge delimits the bone, but in the cyan rectangle, there is another strong edge in the background. The two magenta rectangles, one inside the object of interest and the other outside have almost identical intensities.

both bones with pixel intensities darkened in the overlap region. Clearly, radiance is not locally computable in the manner required for determination of the photo hull. The effect is neither transparency nor shadowing, but can be interpreted as something similar. 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, especially one 90° apart in θ. Due to these effects, local correspondence is not a reliable feature for 3-D reconstruction of bone, and space carving and the variational method of [11] are not suitable.

Now, looking at the prosthetic components and specifically the region outlined in green, it can be seen that there is no texture at all. Consequently, local correspondence once again cannot be used because with no texture it is ill-posed. The visual hull computed from silhouettes, however, is well matched to the prosthetics and can be applied readily. The only drawback with the visual hull is that it cannot take prior information into account. If there were a good way to determine the silhouette of the bones, then it would be possible to get the visual hull of the bones as well, but that task is not trivial.

The assumption made by the stereoscopic segmentation approach of [12] that the surface of the object be homogeneous applies to the prosthetic portion, but not to the bone. The background is clearly nothing like 'blue sky.' Further complicating matters, the image intensities inside and outside the object are very similar in places, for example within the two magenta rectangles.

Edges, such as seen inside the blue rectangle, are one feature that stands 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 rectangle is caused by the boundary of the overlapping region of soft tissue from the two legs. Nevertheless, the use of edge features seems like a possible avenue to pursue for this application along with an approach inspired by the silhouette-based visual hull.

3 Level Set Methods

The level set approach for representing evolving contours was pioneered by Osher and Sethian in [13]. Used to obtain solutions to partial differential equations (PDEs) numerically, level set methods have found application in many different fields including computer vision. This section discusses the main idea of the approach, how the calculation of certain geometric properties of the evolving contour is simplified within the framework, and finally technical details related to implementation. A nice introduction to the topic, with more details than we give here, may be found in [14]; for a comprehensive discussion, one is referred to [15].

3.1 Main Idea

A contour — a closed curve in the plane or surface in space — is often represented parametrically or by a collection of control points that lie on the contour. The level set approach,

however, breaks away from this paradigm and instead, constructs a function φ that is zero on the contour and only on the contour. The contour is the zero level set of φ. Given a particular level set function φ, the contour C it describes is unique. For contours that evolve as a function of time, $C(t)$, the corresponding level set functions also evolve with time; this is discussed in greater detail below.

To help understand the level set concept, an analogy with the Hawaiian islands is useful [16]. The elevation of the land, both underwater and above water can be considered as the level set function $\varphi(u, v)$, and sea level as the zero level. The coastlines of the landmasses that emerge out of the water are the curve C. Over time, the elevation can increase or decrease locally, changing the coastline. If the land rises up very high, two islands can merge and if there is sinking, one island may split into more than one island. Also, 'almost' islands which are peaks but not above sea level, can rise and create new islands. Level set methods naturally handle changes in topology to the contour C. The analogy is almost perfect except for one detail regarding sign, which is expounded as the level set function is formalized.

Let us consider an open region $R \in \mathbb{R}^n$ with closure $\bar{R}$[1] that may or may not be simply connected and has boundary $C = \bar{R} - R$. The region and its boundary change as a function of time t. The level set function $\varphi(\mathbf{x}, t)$, $\mathbf{x} \in \mathbb{R}^n$ satistfies the following properties:

$$\varphi(\mathbf{x}, t) < 0, \mathbf{x} \in R(t); \tag{5}$$

$$\varphi(\mathbf{x}, t) > 0, \mathbf{x} \notin \bar{R}(t); \tag{6}$$

$$\varphi(\mathbf{x}, t) = 0, \mathbf{x} \in C(t). \tag{7}$$

In the analogy, the sign of the level set function inside and outside the region R was reversed. Many different functions can be used for the level set function $\varphi(\mathbf{x})$, but a common choice is the signed distance function which measures Euclidean distance from the point $\mathbf{x}$ to the closest point of the boundary, but assigns a positive or negative sign to the value depending on whether $\mathbf{x}$ is inside or outside the region R.

3.2 Using Level Set Methods in Variational Problems

Now that the level set function has been defined, let us discuss how level set methods are applied to variational problems. Limiting ourselves to curves in the plane, common energy functionals come in two forms: those defined on the boundary and those defined on the region. The functional defined on the boundary is:

$$E_{\text{boundary}}(C) = \int_0^1 g(C(s))ds, \tag{8}$$

whereas the functional over the region is:

$$E_{\text{region}}(C) = \iint_R f\,(u, v)\,du\,dv. \tag{9}$$

[1] In later parts of the report, we are less precise and consider R to be the inside and R^c to be the outside.

The functions f and g are arbitrary and the above functionals are to be minimized. At all local minima, the calculus of variations tells us that the first variation of a functional is zero. The first variation of the boundary functional is:

$$\frac{\delta E_{\text{boundary}}(C)}{\delta C} = -\left(\kappa g - \langle \nabla g, \mathbf{n} \rangle\right) \mathbf{n}, \tag{10}$$

where $\mathbf{n}$ is the unit normal vector to C in the plane, κ is the curvature of C, and $\langle \cdot, \cdot \rangle$ is the standard inner product. When $\frac{\delta E}{\delta C}$ is set equal to zero, this is an Euler-Lagrange PDE. To approach a minimum starting from some initial curve, we move in the gradient direction. As an evolution or flow with a time parameter t, the change in the curve C is:

$$\frac{\partial C}{\partial t}^{(\text{boundary})} = \left(\kappa g - \langle \nabla g, \mathbf{n} \rangle\right) \mathbf{n}. \tag{11}$$

Similarly, the first variation and gradient flow for the region functional are:

$$\frac{\delta E_{\text{region}}(C)}{\delta C} = f \mathbf{n}, \tag{12}$$

and:

$$\frac{\partial C}{\partial t}^{(\text{region})} = -f \mathbf{n}, \tag{13}$$

respectively. We do not delve into first variation calculations here, but they may be found in many references including [17].

Now finally coming to level set implementation, the two above curve flows translate directly to signed distance function update equations. The two respective updates are:

$$\varphi_t^{(\text{boundary})} = \left(\kappa g - \langle \nabla g, \mathbf{n} \rangle\right) \mathbf{n}, \tag{14}$$

and:

$$\varphi_t^{(\text{region})} = f \mathbf{n}. \tag{15}$$

In the context of level set functions, the calculation of $\mathbf{n}$ and κ are simplified as discussed in the next section.

3.3 Calculation of Geometric Quantities

One appealing feature of level set functions is the ease with which geometric quantities of the contours they represent, such as the normal vector and curvature, may be calculated. The unit normal $\mathbf{n}$ to the contour in any dimension is the normalized gradient of the level set function:

$$\mathbf{n} = \frac{\nabla \varphi}{|\nabla \varphi|} \tag{16}$$

In three dimensions, the gradient of the level set function is:

$$\nabla\varphi = \begin{bmatrix} \frac{\partial\varphi}{\partial x} \\ \frac{\partial\varphi}{\partial y} \\ \frac{\partial\varphi}{\partial z} \end{bmatrix} \equiv \begin{bmatrix} \varphi_x \\ \varphi_y \\ \varphi_z \end{bmatrix}, \tag{17}$$

and can be approximated by finite differences:

$$\nabla\varphi(x,y,z)|_{(x_0,y_0,z_0)} \approx \begin{bmatrix} \frac{\varphi(x_0^+,y_0,z_0)-\varphi(x_0^-,y_0,z_0)}{2} \\ \frac{\varphi(x_0,y_0^+,z_0)-\varphi(x_0,y_0^-,z_0)}{2} \\ \frac{\varphi(x_0,y_0,z_0^+)-\varphi(x_0,y_0,z_0^-)}{2} \end{bmatrix}, \tag{18}$$

where we indicate the previous sample by a minus sign superscript and the next sample by a plus sign. Other types of discrete approximations to derivatives can also be applied, but in our work, we use the above finite differences.

Similarly, the curvature κ is also a simple function of the level set function:

$$\kappa = -\nabla \cdot \left(\frac{\nabla\varphi}{|\nabla\varphi|} \right). \tag{19}$$

If we approximate second partial derivatives as follows (example given for x direction):

$$\frac{\partial^2\varphi}{\partial x^2} \equiv \varphi_{xx}(x_0,y_0,z_0) \approx \varphi(x_0^+,y_0,z_0) - 2\varphi(x_0,y_0,z_0) + \varphi(x_0^-,y_0,z_0), \tag{20}$$

and mixed partial derivatives as:

$$\frac{\partial^2\varphi}{\partial x\partial y} \equiv \varphi_{xy}(x_0,y_0,z_0) \approx \frac{\varphi(x_0^+,y_0^+,z_0) - \varphi(x_0^+,y_0^-,z_0) - \varphi(x_0^-,y_0^+,z_0) + \varphi(x_0^-,y_0^-,z_0)}{4}, \tag{21}$$

(and similarly for the other two mixed partials), then κ is explicitly [18]:

$$\kappa = \frac{\varphi_x^2\varphi_{yy} + \varphi_x^2\varphi_{zz} + \varphi_y^2\varphi_{xx} + \varphi_y^2\varphi_{zz} + \varphi_z^2\varphi_{xx} + \varphi_z^2\varphi_{yy} - 2\varphi_x\varphi_y\varphi_{xy} - 2\varphi_x\varphi_z\varphi_{xz} - 2\varphi_y\varphi_z\varphi_{yz}}{2\left(\varphi_x^2+\varphi_y^2+\varphi_z^2\right)^{3/2}}. \tag{22}$$

These expressions can be further simplified by noting that with the choice of the signed distance function as the level set function, the Eikonal PDE $|\nabla\varphi| = 1$ applies.

3.4 Implementation Details

In this section, we note a few details concerned with the implementation of level set methods. An initial contour is required, which is then evolved to minimize some cost, but specifying

Figure 5: A few examples of the real image data from rotating sensor of human knee with attached prosthetic device.

the initial contour does not give an initial level set function. One way to get a level set function, specifically a signed distance function, is to make use of the Eikonal PDE. A map indicating the initial contour just through negative values inside and positive values outside can be made into a signed distance function through the following update:

$$\varphi_t = \text{sign}(\varphi)\,(1 - |\nabla\varphi|)\,, \tag{23}$$

which is derived from the Eikonal PDE. This is the approach we follow. During the contour evolution process, updates do not keep a signed distance function as a signed distance function and periodic reinitializations are necessary, using the same update (23).

Another detail in implementation relates to the fact that the level set function exists over the entire domain, even though it only comes into play near the zero level set. Updating the entire level set function is wasteful, especially when dealing with three or higher dimensions. In our work, we only update the level set function in a narrow band around the zero level set.

4 Shape Reconstruction of Knee Prosthetic and Bone

We have already seen one data image in Fig. 4, which looks straight on at a leg. In fact, all of the images we use have viewpoints that are in a ring outside the leg, having been acquired using a rotating sensor. Each Ω_i plane is parallel to the global z-axis and forms an angle θ_i to the x-z plane. The mental picture to have is of images pasted on a cylinder, where the knee is inside the cylinder and the cylinder extends along the z-axis. A few more of the images are shown in Fig. 5.

The θ_i are known approximately, but we assume in our work that these approximate θ_i are true. Parallel projection is a valid assumption in dealing with X-ray images [19]; hence, we assume π_i is such that $u_i = x\cos\theta_i + y\sin\theta_i$ and $v_i = z$.

In this section we state our surface evolution based approach, starting only with edge-based terms, and then building up to incorporate region-based terms. Results are shown within each subsection.

4.1 Edge-Based Functional

The approach we take is to construct a cost functional and iteratively evolve an initial surface with a gradient flow to minimize the functional, using a level set representation for the surface evolution. For now, the cost functional is composed of a sum of N two-dimensional edge-based terms; later, we also consider region-based terms. Edges are considered because of their presence and importance in X-ray images.

The geodesic active contour cost functional in two dimensions $E_{\mathrm{GAC}}(C)$ has minima where the curve C falls along strong edges in an image I [20]. The cost functional is:

$$E_{\mathrm{GAC}}(C) = \oint_C g\left(C(s)\right) ds, \tag{24}$$

where g is given by:

$$g\left(|\nabla I|\right) = \frac{1}{1+|\nabla I|^p},\ p \in [1,2]. \tag{25}$$

We use this geodesic active contour functional as a foundation for a functional in 3-D.

As seen in Sec. 3.2, this functional leads to the following level set update:

$$\varphi_t = \left(\kappa g - \left\langle \nabla g, \frac{\nabla \varphi}{|\nabla \varphi|} \right\rangle\right) |\nabla \varphi|. \tag{26}$$

In practice with geodesic active contours, a 'balloon' force is also included to prevent local minima and the curve evolution becomes:

$$\varphi_t = \left(\kappa (g+c) - \left\langle \nabla g, \frac{\nabla \varphi}{|\nabla \varphi|} \right\rangle\right) |\nabla \varphi|, \tag{27}$$

where c is a scalar constant.

Writing $\varphi_t = F|\nabla \varphi|$, the force function F in expanded form is:

$$F(u,v) = \kappa(u,v)\left(g(u,v)+c\right) - \frac{1}{\sqrt{\varphi_u(u,v)^2 + \varphi_v(u,v)^2}} \begin{bmatrix} g_u(u,v) \\ g_v(u,v) \end{bmatrix}^T \begin{bmatrix} \varphi_u(u,v) \\ \varphi_v(u,v) \end{bmatrix}. \tag{28}$$

Coming to the problem in three dimensions, we construct the following cost functional:

$$\sum_{i=1}^{N} \oint_{C_i} g\left(C_i(s_i)\right) ds_i, \tag{29}$$

where g is as above. Proceeding as above with a signed distance function $\varphi(x,y,z;t)$, we end up with the following level set update:

$$\varphi_t = \sum_{i=1}^{N} \left(\kappa (g_i + c) - \left\langle \nabla_i g_i, \frac{\nabla_i \varphi}{|\nabla_i \varphi|} \right\rangle\right) |\nabla \varphi|, \tag{30}$$

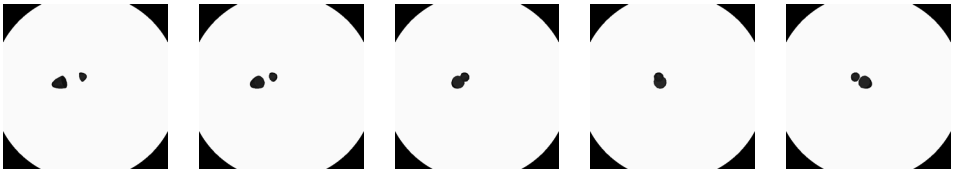


Figure 6: The first 5 images of the sliced ellipsoid example.

where 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. Then,

$$F(x,y,z) = \sum_{i=1}^{N} \kappa(x,y,z)\left(g(u_i,v_i)+c\right) - \begin{bmatrix} g_{u_i}(u_i,v_i) \\ g_{v_i}(u_i,v_i) \end{bmatrix}^T \begin{bmatrix} \varphi_{u_i}(x,y,z) \\ \varphi_{v_i}(x,y,z) \end{bmatrix}, \tag{31}$$

where we have omitted the normalization factors in front of the normal vectors for the purposes of clarity. Explicitly, $\varphi_{u_i}(x,y,z) = \varphi_x(x,y,z)\cos\theta_i + \varphi_y(x,y,z)\sin\theta_i$, and $\varphi_{v_i}(x,y,z) = \varphi_z(x,y,z)$.

In general (u_i, v_i) are not integers and we can either round or take the floor of the indices, or interpolate the value. Although not included here, it may make sense to weigh the different images differently if for example θ is not sampled uniformly (however that is not the case with our dataset).

Although the surface evolution couples information provided by each image in $\mathcal{I}$, one mental picture is of a cylinder extending back along w_i as the region in 3-D that image I_i has influence. It is in this way that the approach relates to visual hull calculation. Occlusions are not modeled for the reasons discussed in Sec. 2.4.

We now use the above surface evolution on some synthetic examples. The first example is an ellipsoid with a slice perpendicular to the major axis missing; it is not aligned to any of the global coordinate axes. We create $N = 16$ images spaced 0.4 radians apart, that are similar to a prosthetic appearing in an X-ray image. The first five images are shown in Fig. 6. We use a large cube completely outside of the object as an initial surface. The surface evolution is shown in Fig. 7. Different views of the last iteration of Fig. 7 are shown in Fig. 8. The extracted 3-D shape is fairly good in that it accurately represents the original synthetic shape.

The same thing is done for a different shape, a sphere with a cylindrical bite taken out. Fig. 9 through Fig. 11 correspond to Fig. 6 through Fig. 8 for this example. In this case, the reconstructed 3-D shape is still fairly good, but not as good as the previous example. There is leakage through some edges. We address this shortcoming by introducing region-based terms in Sec. 4.2.

In Fig. 12, a visualization of the implicit signed distance function $\varphi(x,y,z)$ at an intermediate iteration is shown, where the top left part of the figure looks along the z-axis, the top right part looks along the y-axis and the bottom left part looks along the x-axis. We see

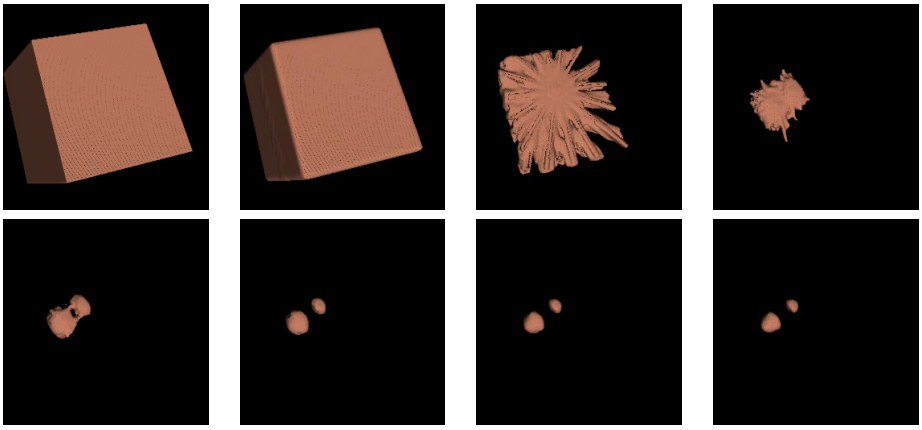


Figure 7: Iterations of the surface evolution in raster scan order.

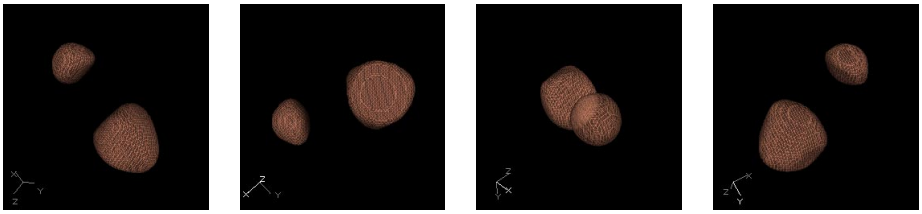


Figure 8: Different views of the 3-D reconstruction.

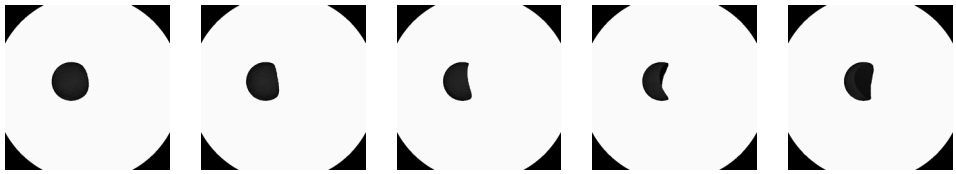


Figure 9: The first 5 images of the sphere with a cylindrical bite example.

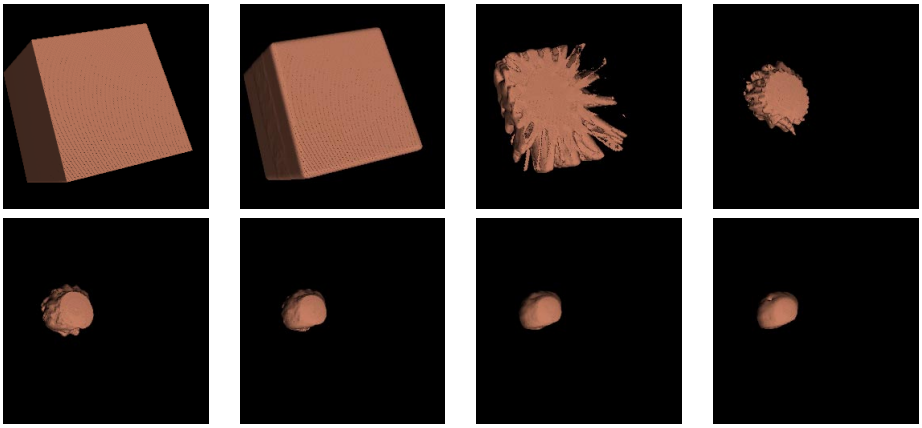


Figure 10: Iterations of the surface evolution in raster scan order.

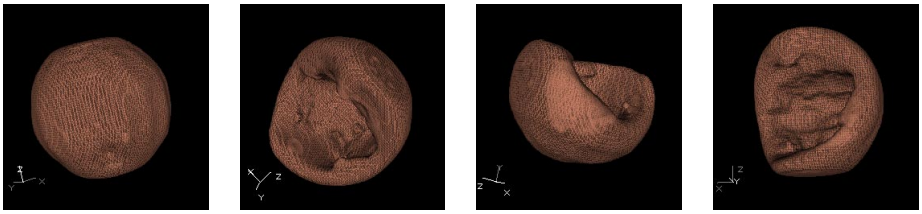


Figure 11: Different views of the 3-D reconstruction.

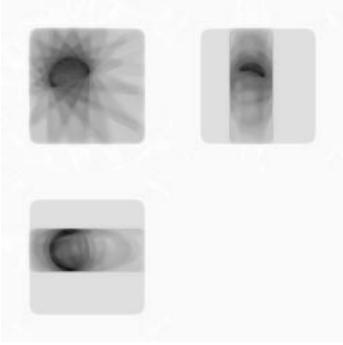


Figure 12: Visualization of signed distance function at intermediate iteration.

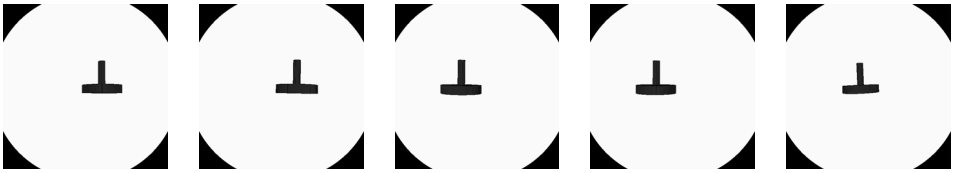


Figure 13: Some sample images of the two stacked cylinders.

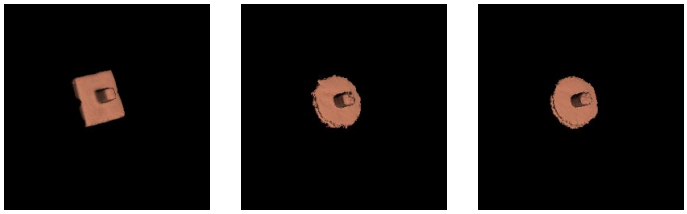


Figure 14: Surfaces with increasing N from left to right.

that the limited number of views, i.e. $N = 16$, create some artifacts, which are also visible in the surface evolutions shown in Fig. 7 and Fig. 10. The figure may remind some readers of CT reconstructions with undersampled measurements. In general, this is a similar effect, but the details are slightly different.

Using a different shape, a thin cylinder stacked on a wide cylinder, we show the effect of N, i.e. different numbers of images. Some sample data images are given in Fig. 13. We use three different values of N along with different θ spacings. The first has $N = 2$ images and the views are 1.52 radians apart; the second uses $N = 12$ images with a θ spacing of 0.56 radians; whereas the third has $N = 28$ images 0.24 radians apart. This is by no means a detailed study into the effect of N, but the resulting surfaces are shown in Fig. 14. As the number of views is increased, the resulting surface is smoother and more circular as one would expect.

Through these examples, we have seen both that our proposed surface evolution is a step in the right direction and that we need something more. The surface leaking through the true boundary on an example with a benign background is a strong indication that the method, as is, does not work with real data, and in fact it does not. Results from real data are difficult to even interpret. In the next section, region-based terms are introduced that allow shape extraction from real X-rays.

4.2 Region-Based Terms

In 2-D, problems with curves not stopping at weaker edges and situations in which regions are not separated by edges led to the development of the geodesic active regions functional

[21]. In the same way as the geodesic active contour functional was the foundation for our 3-D edge-based functional, we use the geodesic active regions functional to build a 3-D edge- and region-based functional.

The geodesic active region approach assumes that some knowledge regarding the image intensities of the two regions, the inside R and outside R^c, is known a priori. The assumption is that pixel values are independent given the region and have probability distribution function either $p_R(I(u,v))$ or $p_{R^c}(I(u,v))$ depending on the region. The geodesic active regions energy functional includes an edge term as well as a region term. Taking the edge term to be E_{GAC} of equation (24), the functional is:

$$E_{\mathrm{GAR}}(C) = \alpha E_{\mathrm{GAC}} + (1-\alpha)\left(-\iint_R \log\left(p_R(I(u,v))\right) du\, dv - \iint_{R^c} \log\left(p_{R^c}(I(u,v))\right) du\, dv\right). \tag{32}$$

The parameter $\alpha \in [0,1]$ trades off the contributions of the edge and region terms. Based on Sec. 3.2 and Sec. 4.1, the force function F is:

$$F(u,v) = \alpha\left(\kappa(u,v)\left(g(u,v)+c\right) - \frac{1}{\sqrt{\varphi_u(u,v)^2+\varphi_v(u,v)^2}}\begin{bmatrix} g_u(u,v) \\ g_v(u,v)\end{bmatrix}^T\begin{bmatrix}\varphi_u(u,v)\\ \varphi_v(u,v)\end{bmatrix}\right) + (1-\alpha)\left(-\log\left(p_R(I(u,v))\right) + \log\left(p_{R_c}(I(u,v))\right)\right). \tag{33}$$

The sign of the outside part of the region term is reversed because the inward normal of R^c is the outward normal of R.

Continuing on to the 3-D extension, in the same manner as the previous section, we compose an overall functional as the sum of individual 2-D functionals for each image in $\mathcal{I}$ that extend back in a cylinder of influence. Here, the assumption is of the shape S having an intensity distribution p_R and the background having an intensity distribution p_{R^c}.

$$\alpha\sum_{i=1}^{N}\oint_{C_i} g\left(C_i(s_i)\right) ds_i + (1-\alpha)\sum_{i=1}^{N} -\iint_{\pi_i(S)} \log\left(p_R(I_i(u_i,v_i))\right) du_i\, dv_i - \iint_{\pi_i(S)^c} \log\left(p_{R^c}(I_i(u_i,v_i))\right) du_i\, dv_i. \tag{34}$$

The force function in this case is:

$$F(x,y,z) = \alpha\sum_{i=1}^{N}\kappa(x,y,z)\left(g(u_i,v_i)+c\right) - \begin{bmatrix} g_{u_i}(u_i,v_i) \\ g_{v_i}(u_i,v_i)\end{bmatrix}^T\begin{bmatrix}\varphi_{u_i}(x,y,z)\\ \varphi_{v_i}(x,y,z)\end{bmatrix} + (1-\alpha)\sum_{i=1}^{N} -\log\left(p_R(I_i(u_i,v_i))\right) + \log\left(p_{R_c}(I_i(u_i,v_i))\right). \tag{35}$$

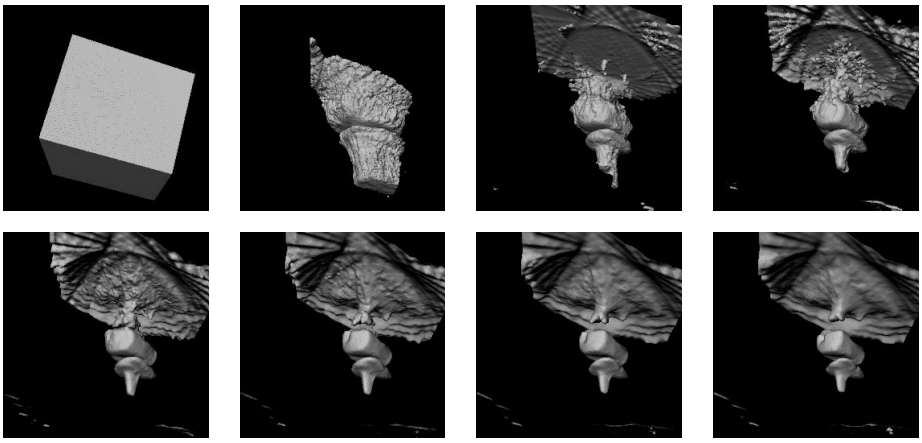


Figure 15: Iterations of the surface evolution in raster scan order.

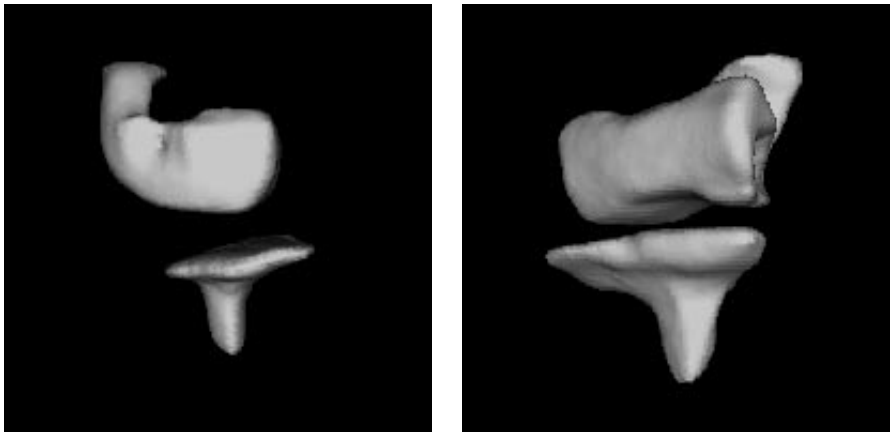
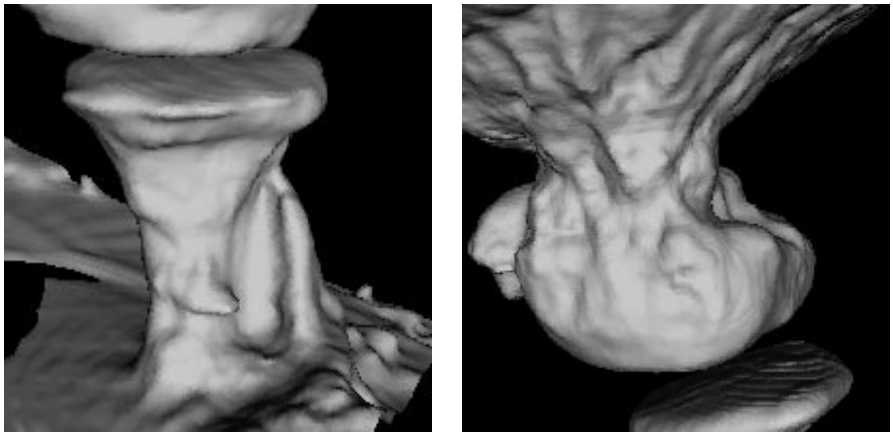


Figure 16: Different views of the 3-D reconstruction.

To obtain the 3-D shape of the prosthetics from the real X-ray data, although not accurate, we assume Gaussian distributions for p_R and p_{R^c} with different means μ_R and μ_{R^c}. $N = 16$ images with an approximate spacing of 0.038 radians in θ are used and the images are slightly blurred beforehand. The surface evolution is shown in Fig. 15 and different views of the final surface are shown in Fig. 16. A very small piece of the femur along with part of the black boundary is apparent in Fig. 15, but these pieces are disjoint from the prosthetics and easily discarded in Fig. 16. In the final 3-D reconstruction, the two



(a) (b)

Figure 17: 3-D reconstructions of (a) tibia and fibula, and (b) femur.

pieces of the prosthetic have been recovered; they are correct visually. We should note that with this edge-region surface evolution, the final reconstructions of the synthetic examples of the previous section are also improved. By lifting the geodesic active regions functional to 3-D, homogeneous objects are reconstructed.

4.3 Intensity Distribution In Region

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.

A spatially varying mean function, $\mu_R(d_i)$, is applied to the Gaussian probability distribution for the inside region R in the edge-region surface evolution. Since the signed distance function in the 3-D surface evolution $\varphi(x, y, z)$ is in 3-D, to get a distance d_i for the argument of μ_R, the 3-D signed distance is projected down to the plane Ω_i for each i:

$$d_i = \varphi(x, y, z) \sin\left(\frac{\frac{\partial\varphi(x,y,z)}{\partial x}\cos\theta_i + \frac{\partial\varphi(x,y,z)}{\partial y}\sin\theta_i}{|\nabla\varphi(x, y, z)|}\right). \tag{36}$$

Results using a spatially varying mean are shown in Fig. 17. The recovered tibia and fibula are shown in Fig. 17a, and the femur in Fig. 17b. 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.

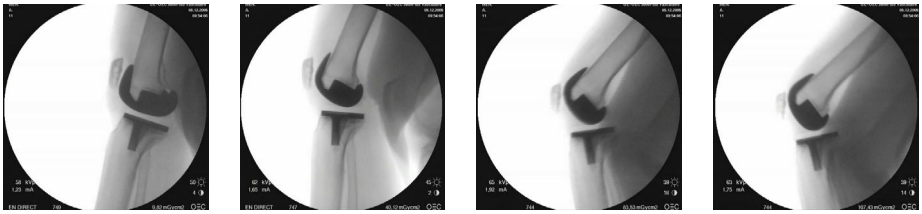


Figure 18: Movie sequence of flexing knee joint.

5 Conclusion

In this report, 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.

Results acquired thus far have been fairly successful at reconstructing the 3-D shape of the metal prosthetics. Results with bones are promising but can be improved through more sophisticated modeling of the intensity distribution in bones, building upon the simple formulation that has been attempted; one way to do so is by learning from data [22]. Another direction for improving the formulation is to include shape priors. Also, multi-phase segmentation, i.e. partitioning space into more than two regions, may be considered. The actual surface evolution may be carried out using Sobolev active contours [23].

Beyond 3-D shape reconstruction, the medical application suggests another problem, namely tracking the movement of the bones and prosthetics in a movie sequence, also from the X-ray modality, and inferring their kinematics. As an example of such a sequence, some frames of a knee that is bending are shown in Fig. 18. 3D-2D rigid body registration and tracking are difficult problems which will be complicated by the same imaging issues discussed in this report, as well as others such as motion blur.

X-ray radiography is an exciting, medically relevant, and challenging domain that brings forth issues for computer vision methods not encountered with other imaging modalities. Some of these issues have been looked at in this report, but the surface has just been scratched; there are many opportunities for advances within X-ray imaging.

Acknowledgment

The problem was suggested by and X-ray data of the knee were provided by Alain Kulski, Rémy Raymond, Phillipe Hernigou, and Alain Rahmouni of Service de Radiologie et d'Imagerie Médicale and Service d'Orthopédie, Hôpital Henri Mondor, Créteil, France.

References

[1] F. Rodriguez y Baena, "Man and the machine," *IET Computing & Control Engineering*, pp. 28–31, Oct./Nov. 2006.

[2] O. Glasser, "The hand of Mrs. Wilhelm Roentgen: The first X-ray image, 1895," in *Wilhelm Conrad Röntgen and the Early History of the Roentgen Rays.* London: Bale & Danielsson, 1933.

[3] 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.

[4] B. De Man, J. Nuyts, P. Dupont, G. Marachal, and P. Suetens, "Metal streak artifacts in X-ray computed tomography: A simulation study," *IEEE Transactions on Nuclear Science*, vol. 46, no. 3, pp. 691–696, June 1999.

[5] J. Williamson, B. Whiting, J. Benac, R. Murphy, G. Blaine, J. O'Sullivan, D. Politte, and D. Snyder, "Prospects for quantitative CT imaging in the presence of foreign metal bodies using statistical image reconstruction," in *Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging*, July 2002, pp. 691–696.

[6] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, New York, June 2006, pp. 519–526.

[7] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150–162, Feb. 1994.

[8] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-based visual hulls," in *Proceedings of the 27th International Conference on Computer Graphics and Interactive Techniques*, New Orleans, July 2000.

[9] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.

[10] A. Broadhurst, T. W. Drummond, and R. Cipolla, "A probabilistic framework for space carving," in *Proceedings of the IEEE International Conference on Computer Vision*, Vancouver, July 2001.

[11] 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.

[12] A. Yezzi and S. Soatto, "Stereoscopic segmentation," *International Journal of Computer Vision*, vol. 53, no. 1, pp. 31–43, 2003.

[13] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on the Hamilton-Jacobi formulation," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[14] S. Osher, "Level set methods," in *Geometric Level Set Methods in Imaging, Vision, and Graphics*, S. Osher and N. Paragios, Eds. New York: Springer-Verlag, 2003, pp. 3–20.

[15] J. Sethian, *Level Set Methods*. Camridge University Press, 1996.

[16] A. S. Willsky, personal communication.

[17] S. Osher and N. Paragios, Eds., *Geometric Level Set Methods in Imaging, Vision, and Graphics*. New York: Springer-Verlag, 2003.

[18] R. Goldman, "Curvature formulas for implicit curves and surfaces," *Computer Aided Geometric Design*, vol. 22, no. 7, pp. 632–658, Oct. 2005.

[19] X. Hu and N. Ahuja, "Motion estimation under orthographic projection," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 6, pp. 848–853, Dec. 1991.

[20] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, pp. 61–79, 1997.

[21] N. Paragios and R. Deriche, "Geodesic active regions: A new framework to deal with frame partition problems in computer vision," *Journal of Visual Communication and Image Representation*, vol. 13, pp. 249–268, 2002.

[22] M. E. Leventon, O. Faugeras, W. E. L. Grimson, and W. M. Wells, III, "Level set based segmentation with intensity and curvature priors," in *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, Hilton Head, South Carolina, June 2000.

[23] G. Sundaramoorthi, A. Yezzi, and A. C. Mennucci, "Sobolev active contours," *International Journal of Computer Vision*, p. to appear.

Laboratoire MAS − Ecole Centrale Paris

Grande Voie des Vignes

92290 Chatenay Malabry Cedex (France)

http://www.mas.ecp.fr