

A 2D Fourier Approach to Deformable Model Segmentation of 3D Medical Images

Eric Berg¹, Mohamed Mahfouz^{2,3}, Christian Debrunner¹, William Hoff¹

¹ Colorado School of Mines, Golden, Colorado

² University of Tennessee, Knoxville, Tennessee

³ Oak Ridge National Laboratory, Oak Ridge, Tennessee

Abstract. Anatomical shapes present a unique problem in terms of accurate representation and medical image segmentation. Three-dimensional statistical shape models have been extensively researched as a means of autonomously segmenting and representing models. We present a segmentation method driven by a statistical shape model based on *a priori* shape information from manually segmented training image sets. Our model is comprised of a stack of two-dimensional Fourier descriptors computed from the perimeters of the segmented training image sets after a transformation into a canonical coordinate frame. We apply our shape model to the segmentation of CT and MRI images of the distal femur via an original iterative method based on active contours. The results from the application of our novel method demonstrate its ability to accurately capture anatomical shape variations and guide segmentation. Our quantitative results are unique in that most similar previous work presents only qualitative results.

1 Introduction

Current methods in three-dimensional image segmentation typically employ statistical shape models, first developed by Cootes and Taylor [1] as a means to incorporate *a priori* shape information. A statistical shape model is trained by considering the canonical parameterization of a set of similar shape instances, or training shapes. Performing a principal components analysis (PCA) on the parameterized shapes highlights the statistical modes of variation in the shape, allowing for a possible reduction in the dimension of the model shape space, discussed in more detail in section 2.3. A well-constructed statistical model will provide a shape constraint on subsequent segmentations of new images.

The primary original contribution of this work is in our segmentation algorithm, consisting of two key steps: Development of a shape parameterization for a set of training instances and; autonomous segmentation of similar shapes in three-dimensional (3D) medical images, guided by the statistical shape model. For the purpose of developing and testing our method, we modeled the human distal femur from 19 sets of 3D images generated by both CT and MRI image modalities (15 CT, 4 MRI); the datasets include both left and right femurs, so we mirrored the left femur across the midsagittal plane as was done in [2, 3], creating additional samples of the right femur.

A particular instance of the distal femur is constructed as a stack of closed two-dimensional (2D) contours representing the outer surface of the bone on each slice. Each contour is parameterized by a Fourier descriptor (FD) decomposition after the stack of contours is transformed into a canonical coordinate frame. A bone instance is then represented as a single vector whose components are the FDs from all the slices comprising the instance, ordered identically from one bone instance to the next. The statistical 3D model of a class of bones (e.g., distal femurs) is captured by a PCA of such vectors from many bone instances, forming a model shape space.

Our segmentation algorithm employs an iterative process of 1) independently adjusting each boundary contour via a traditional active contours technique [4], followed by 2) a projection into the shape space defined by the statistical model. As we will show, this iterative process provides for local deformations of the shape model during segmentation, while partially constraining the global shape to be consistent with the prior shapes. Prior methods (see e.g., [5]) have explored combined optimization functions that include terms for both the local (image) constraints and the global (statistical model) constraints. The advantage of this combined optimization function is its ability to optimize a model's shape and its relation to the images simultaneously during segmentation. The inability of the combined function to capture shapes outside of the model shape space limits solutions to be fully within the shape space. Our segmentation technique is geared toward solving this problem. As we will explain in section 2.4, by separating the local and global optimizations, we allow for objects slightly outside of the shape space to be segmented, thus permitting the use of these new shapes as training instances to further refine future PCA results.

The iterative segmentation procedure yields a stack of 2D contours, each of which is simply a permutation of the contours initialized on each slice; the contours are then stacked to allow formation of a 3D surface model. Our results will demonstrate the ability of our method to accurately segment new datasets.

2 Approach

The statistical shape model consists of the average of the training shapes in addition to the primary modes of variation among the training shapes as determined by PCA. Model-guided segmentation produces a new model whose shape will conform globally to the training shapes used to form the statistical model, and locally to the images being segmented. In this section, we discuss the development of the statistical model from images and its application to our segmentation approach.

2.1 Image Preprocessing

Each 3D image volume to be used as a training set is manually segmented to extract the distal section of the femur. The binary (black and white) images resulting from the segmentation are rigidly registered to a coordinate system as shown in Fig. 1. The z -axis passes approximately through the center of the condyles and intersects the center of the shaft. We define the center of the condyles as the centroid of the segmented region in the image containing the proximal end of the intercondyloid fossa, the junction of the medial and lateral condyles. The center of the shaft is defined as the

centroid of the segmented region in the image approximately 120 mm above the tip of the medial condyle. The x -axis coincides with a line drawn between the center of the medial and lateral condyles and intersects the distal end of the z -axis. The y -axis is positive in the posterior direction. Due to potentially large initial misalignment with the coordinate frame, new training image sets may be iteratively registered to ensure that proper alignment is achieved. We leave a detailed analysis of this registration process as future work.

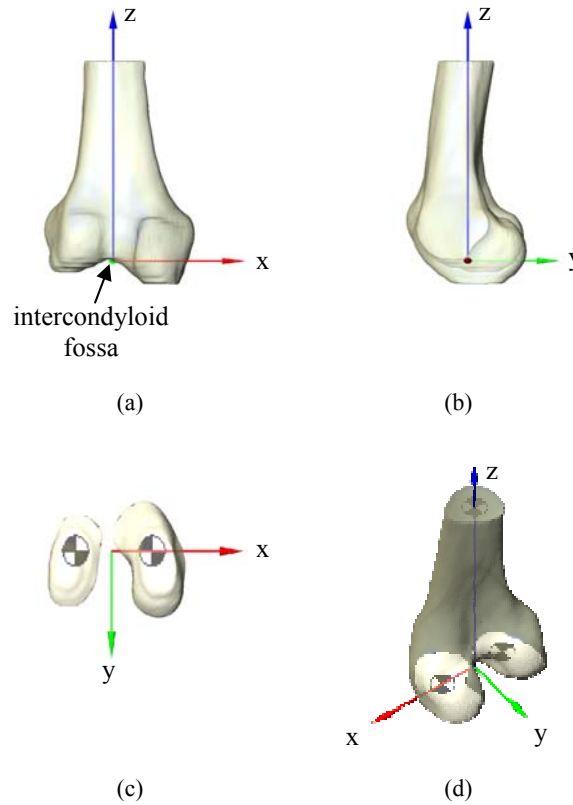


Fig. 1. Coordinate frame definition showing (a) coronal view, (b) sagittal view, (c) axial view, and (d) isometric view

2.2 Fourier Descriptors

The binary images resulting from the manual segmentations each contain at least one region representing the bone cross-section; in the case of images intersecting the condyles, two distinct regions may be present. The regions on each slice can be reduced to closed contours, allowing for a FD representation to be computed. FDs were originally developed by [6] for use in shape discrimination and identification as

in [7]. A more recent application of FDs involves medical image segmentation [5], where the combined optimization approach discussed in the introduction is used.

Each contour is defined by a set of N perimeter pixels, whose indices, after multiplication by the pixel size in millimeters, represent x and y coordinates. Conversion to FDs normalizes contours of unequal length and shape and reduces the amount of data required to store each model. The pixel locations are represented as a vector of the form,

$$s = (s_1 \ s_2 \ \dots \ s_N)^T, \quad (1)$$

where $s_i = x_i + jy_i$ and j is $\sqrt{-1}$. The 2D FDs are computed via the discrete Fourier transform (DFT) as follows:

$$a_k = \frac{1}{N} \sum_{q=0}^{N-1} s_q \exp\left(\frac{-j2\pi qk}{N}\right), \quad (2)$$

where the values of k are the indices $0, 1, \dots, N-1$.

The most significant features of the contour are captured in the lower frequency terms of the Fourier transform, so we can produce a more concise description of the contour by eliminating the high-frequency terms. Due to the periodicity of the DFT, the low-frequency terms are located at the ends of the coefficient sequence, so we remove coefficients from the middle of the sequence. We found through experiments that the 32 lowest frequency coefficients sufficiently describe the shape of all distal femur cross-sections. To reconstruct a contour with N' points from this reduced form, the FD vector a' must be constructed as follows,

$$a' = (a_1 \ \dots \ a_{16} \ 0 \ \dots \ 0 \ a_{N'-15} \ \dots \ a_{N'})^T, \quad (3)$$

where the number of zeros to be inserted is $N' - 32$, assuming $N > 32$. The vector is padded with zeros to create a vector of length N' , while maintaining the exact shape of the contour as described by the 32 coefficients. The effect of the additional terms in a' is to increase the number of points to be reconstructed from the FDs. The reconstruction of the contour points is achieved through the inverse discrete Fourier transform (IDFT),

$$s_q = N' \sum_{k=0}^{N'-1} a'_k \exp\left(\frac{j2\pi kq}{N'}\right). \quad (4)$$

After rounding the reconstructed coordinates to the nearest pixel, it is important to note that if $N' > N$, s_q will occasionally have repeated values representing the same pixel location. If the goal is to reconstruct images from the FDs, then this is not a problem; however, if the points are to be processed by an algorithm that assigns weights to each point, such as iterative closest point (ICP), the duplicated points should be removed to avoid improper bias. In our case, we are reconstructing images, so we assign a large value to N' , ensuring that no gaps occur in the reconstructed contour.

2.3 Shape Model Representation and Construction

Each femur model is composed of M contour slices parallel to the axial plane and stacked in the z -direction (Fig. 2), where the previously defined z -axis is normal to the axial plane. Each of the M contours is described by Q FDs (in our case, $Q = 32$), which are complex pairs as prescribed by equations 1 and 2. A contour is represented by a single vector formed by concatenating the real and imaginary parts of the Q FD coefficients that represent it. The model vector m is then formed by concatenating the M contour vectors to form a vector of length $n = 2QM$. The multiple of 2 is due to the real and imaginary part of each FD. The vector formation is illustrated in Fig. 3.

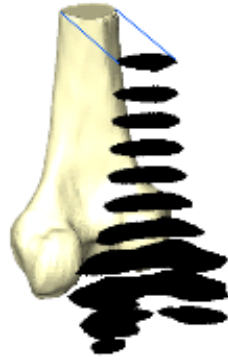


Fig. 2. Model of the distal femur demonstrating the slice structure. Note that the slices shown are a select few of the total number of slices

For a set of T models, $m_1, m_2, \dots, m_T$, the $n \times 1$ mean vector, $\bar{m}$, is calculated by,

$$\bar{m} = \frac{1}{T} \sum_{i=1}^T m_i . \quad (5)$$

The modes of variation among the training models are determined by a PCA. First, the mean model vector, $\bar{m}$, is subtracted from each m_i , forming a deviation matrix whose columns are the deviation vectors,

$$dm_i = m_i - \bar{m} . \quad (6)$$

The PCA is computed using an eigendecomposition of the $n \times n$ covariance matrix formed from the deviation matrix, dm . The resulting orthonormal eigenvectors, p_k , correspond to eigenvalues, λ_k . Arranging the eigenvalues in descending order and reordering the eigenvectors accordingly, we can determine the eigenvectors that account for the most significant variance.

Any shape instance in the training set can be represented by the mean model and a linear combination of the eigenvectors [8]. Typically, a model can be accurately described by $t < T$ eigenvectors, or principal components (PCs), corresponding to the t largest eigenvalues. As Cootes *et al.* [9] and Hutton *et al.* [10] point out, the PCs

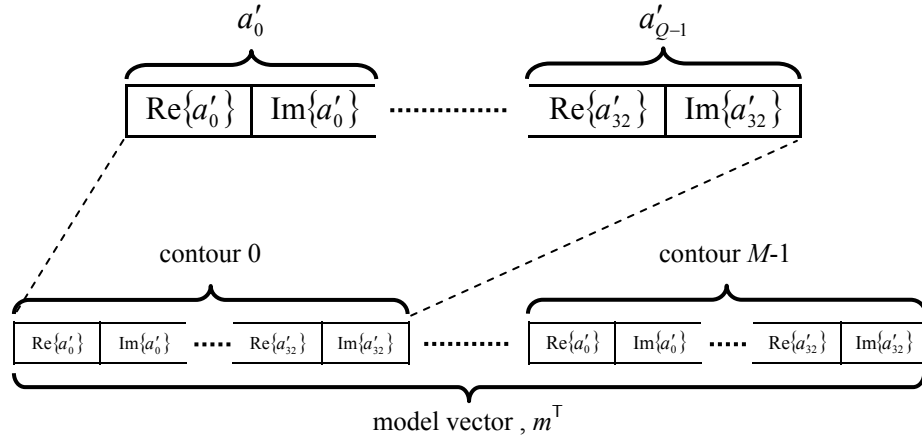


Fig. 3. Model vector formation. The upper vector shows a single image represented as Q FDs; the lower vector represents a single model, m as M sets of FDs. Note that m is shown transposed

representing 98% of the variance should be sufficient to capture the shape of anatomical models, while the remaining 2% is considered noise. A model is then approximated as,

$$m = \bar{m} + \sum_{i=1}^t b_i p_i, \quad (7)$$

where b_i is the coefficient multiplying the i^{th} PC. The orthonormal column vectors, p_i form the principal basis, or shape space of the model. Thus, our representation of the statistical model consists of the mean model vector and the vectors corresponding to the first t PCs. Section 3 will provide more insight into the effects of model reduction.

2.4 3D Image Segmentation

A 3D image volume that is not included in the training set can be segmented by iteratively deforming the FD statistical model until it closely matches the correct features in the 3D image set. The statistical model ensures a segmentation consistent with the training models. Prior to segmentation, the new image set must be registered to the coordinate frame described in section 2.1. Note that only three points must be manually selected for registration; the Fourier descriptor representation provides the feature correspondences required for the PCA. A search for the optimal solution occurs as separate local and global optimizations; a local search (unconstrained by the statistical model) is performed independently on each image using an active contours algorithm, and a global search takes the form of a projection of the active contours solution onto the principal basis of the statistical model. Since the two searches are forced to intersect occasionally, the constraints from both optimization functions are imposed. The approach is outlined in Algorithm 1.

Most previous work in this area involves minimizing a global objective function that combines terms driven by the image data and terms driven by the prior shape information. This approach will often result in a solution that is not the true global minimum, but a spurious local minimum. Our algorithm also accounts for the case where the optimal solution may not actually occur at a minimum, either local or global. For instance, if a dataset outside of the training set is segmented, the optimal solution will most likely fall slightly beyond the shape space defined by the PCs. By separating the search spaces, we can find solutions that the combined optimization cannot find.

Algorithm 1. Segmentation procedure

- 1) Initialize active contours (see *Local Search*) based on the mean model, $\bar{m}$
- 2) Allow active contours to deform independently on each image for h iterations
- 3) Convert active contours to Fourier descriptors
- 4) Project FDs onto principal basis
- 5) Form a new model from projected FDs
- 6) Check equation 8 for convergence (see below)
 - a) if $\Delta < \text{threshold}$, repeat steps 2 and 3 only
 - b) if $\Delta > \text{threshold}$, repeat steps 2-6 with new model in step 5 as new initialization for active contours

The alternating process of active contour deformation and shape space projection continues until the following convergence function reaches some empirical threshold,

$$\Delta = \sum_{i=1}^t (b_{ni} - b_{(n-1)i})^2 \quad \text{for } n > 1, \quad (8)$$

where Δ represents the change in the model parameters from one iteration to the next. When this squared sum is less than the specified threshold, the model is assumed to have converged on a shape space solution. An additional optimization of the active contours is performed after this convergence.

Local Search

The local search for a solution occurs in the individual 2D image spaces in the form of active contours, or snakes. Active contours, originally developed by Kass *et al.* [4] employ an energy minimization approach to detect features in images such as lines and edges. The technique typically works best with images that have clearly defined edges with minimal noise; otherwise a good initialization is required for an active contour to converge on an acceptable solution. In the case of shape modeling, we have a good initialization provided by the model information from the previous iteration, provided by steps 1 or 6b in Algorithm 1. The active contour is constructed

as a spline, parametrically defined as $\mathbf{v}(r) = [x(r), y(r)]$, where $x(r)$ and $y(r)$ are the x and y positions, in pixel coordinates, of the nodes of the contour to be optimized. As in [4, 11], the energy of the spline can be represented as,

$$E_s = \int_0^1 [E_i(\mathbf{v}(r)) + E_e(\mathbf{v}(r))] ds, \quad (9)$$

where E_i and E_e are the internal and external spline energies respectively. E_i can be expanded as,

$$E_i = \frac{1}{2} [\alpha |v'(r)|^2 + \beta |v''(r)|^2], \quad (10)$$

where the first order term is analogous to elasticity between the nodes and the second order term controls rigidity of the spline. The coefficients α and β are used to control the extent to which each term constrains the deformation of the contour. E_e represents the effect on the energy functional of the image gradient, computed as,

$$E_e = -|\nabla I(x, y)|^2, \quad (11)$$

where ∇I is the image gradient computed via a Sobel operator. The combined internal and external energies tend to drive the contour toward the image gradients, while maintaining control over the shape of the contour.

Global Search

The global search for a solution occurs in the principal space by computing the parameters b_i in equation 7 by projecting the FDs from step 3 in Algorithm 1 onto the principal basis. This has the effect of constraining the active contours solution to be in the shape space, thus from one iteration to the next, the local solution is projected into shape space to find a global solution. After several iterations, the parameters b_i will vary negligibly, indicating that a global solution has been found. As previously discussed, this may not represent the optimal solution since the actual solution likely falls slightly beyond the shape space. With this optimal global solution as an initialization for one final local optimization, ie. an additional active contours adjustment, a solution that is close to the shape space, but optimized to fit the 3D image information can be found.

3 Results

In order to verify the ability of our modeling and segmentation techniques to extract anatomical features from medical images, we performed experiments with the 19 CT and MRI datasets. The datasets are each segmented manually and the resulting binary images are transformed and resampled as described in section 2.1. For testing purposes we sample the image volumes to a voxel size of 0.75 x 0.75 x 3.00 mm, resulting in 40 images spanning the distal 120 mm of the femur. The original grayscale images are transformed and resampled by the same transformation as the

binary images so that subsequent model-based segmentations will be performed in the model domain.

Each of the 19 image sets is autonomously segmented according to Algorithm 1 with a model defined via a leave-one-out approach (see e.g., [12]), where the remaining 18 manual segmentations are used as training shapes. To illustrate the effect of the number of principal components on model-based segmentation, each of the 19 image sets is segmented multiple times, each time with a shape model defined by the remaining 18 datasets with a varying number of PCs ($t = 1, 2, \dots, 18$). This results in $19 \times 18 = 342$ total segmentations of the 19 datasets.

We compare the autonomous model-based segmentation results (actual results) to the manual segmentation results (expected results) to obtain a measure of the quality of the segmentation. This comparison is performed in the 3D image domain by computing the shortest 3D Euclidean distance between the two voxel “surfaces” for each voxel on the autonomously segmented set. For each of the 342 segmentations, we obtain a mean separation distance and a maximum, or worst case, distance. Ideally, these distances are zero for a perfect fit, assuming that the solution we seek is the same as the manual segmentation.

Two experiments are performed to verify that the iterative optimization with a final local optimization provides a better final result than a solution constrained to be fully within the shape space. Figs. 4 and 5 show, respectively, the mean and maximum distance averaged over all segmentations as a function of the number of PCs in the statistical model. The two curves in each plot represent the different final optimizations, one for the locally optimized solution and one for the shape space constrained solution.

In section 2.3, we noted that a model should capture approximately 98% of the variance in the training model data. Fig. 6 shows the cumulative variance as a function of the number of PCs for a typical model. The cumulative variance for the distal femur model approaches 98% at approximately 11 PCs.

The mean separation distance graphically shown in Fig. 4 demonstrates that an increase in the number of PCs yields an autonomous segmentation solution closer to the manual segmentation. Additionally, the final active contours adjustment pushes the mean distance between the expected surface and the actual surface to approximately half of the mean distance found by constraining the solution to be fully within the shape space. We can see the same trend for the maximum separation. Note that the mean distance levels off between 11 and 12 PCs, indicating that an increase in the number of PCs beyond 11 or 12 may not improve the segmentation.

Our best mean and maximum distance measures compare well to the best values presented in [3] and are summarized in Table 1. In fact, our best mean value is five times less than the best reported value by Kaus *et al.* [3]. We must note that [3] does not solve a segmentation problem, rather they are fitting a triangular mesh to manually segmented noiseless binary 3D images to form a unique surface model. Our method is applied to segment noisy 3D medical images and compare the results to a manual segmentation of the same sets. Their measure of deviation between the autonomously adjusted mesh (vertices) and the manually segmented binary 3D images (voxels) is similar to our measure of deviation between the autonomously and manually segmented 3D images (voxel to voxel), thus a comparison to their results is valid.

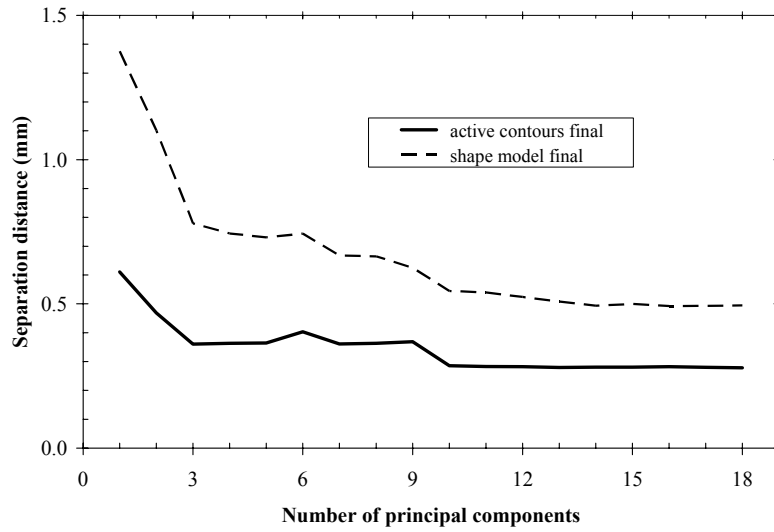


Fig. 4. Mean separation distance between autonomously segmented 3D image volumes and corresponding manually segmented volume. The solid line represents the solution where the active contours are adjusted after the final shape space projection; the dashed line is the solution resulting from the final shape space projection

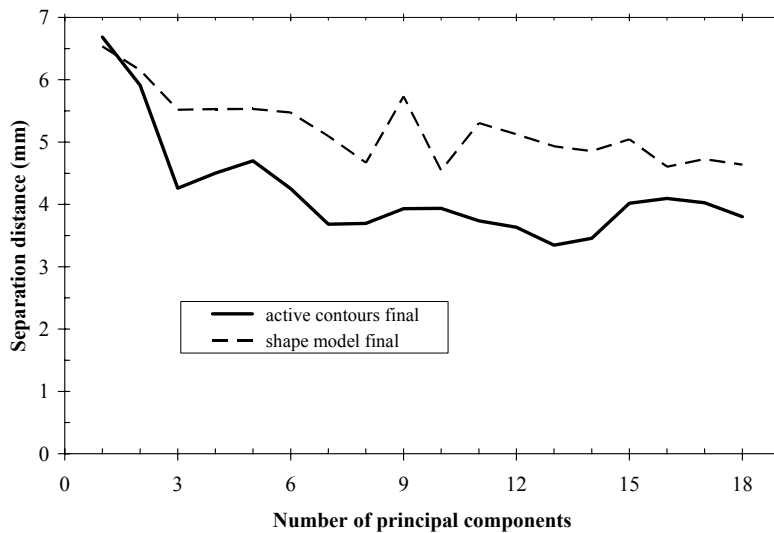


Fig. 5. Maximum separation distance between autonomously segmented 3D image volumes and corresponding manually segmented volume. The solid line represents the solution where the active contours are adjusted after the final shape space projection; the dashed line is the solution resulting from the final shape space projection

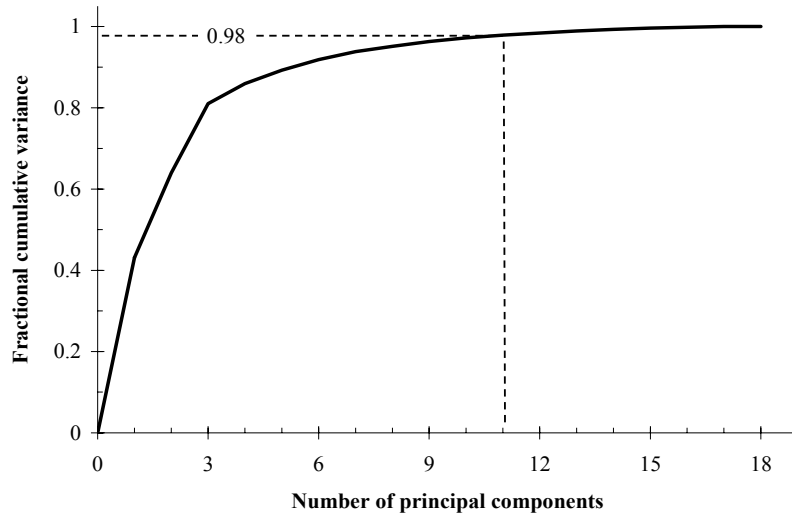


Fig. 6. Average cumulative variance as a function of the number of principal components

Table 1. Comparison of results with Kaus *et al.* Value in parenthesis is the lowest number of PCs at which the value occurs

	max Euclidean distance (mm)	mean Euclidean distance (mm)
Our results	3.35 (13)	0.28 (12)
Kaus <i>et al.</i>	~ 4.50 (20)	~ 1.40 (20)

4 Conclusion and Future Work

The described experiments have demonstrated that our model building method accurately and compactly captures the shape of the distal femur and our segmentation procedure successfully segments new datasets (Fig. 7). A comparison of our results to previously published values indicates that our segmentation method is accurate. With only 12 PCs we were able to segment a 3D image volume with an average deviation of only 0.28 mm, a deviation of less than a voxel in most 3D medical images.

Future work in this area will involve extending the size of the training set for more thorough testing and to explore a potential correlation between the number of training models and the number of PCs required to capture the statistics of the distal femur. Additional work will also include modeling more complex structures, including soft tissue.

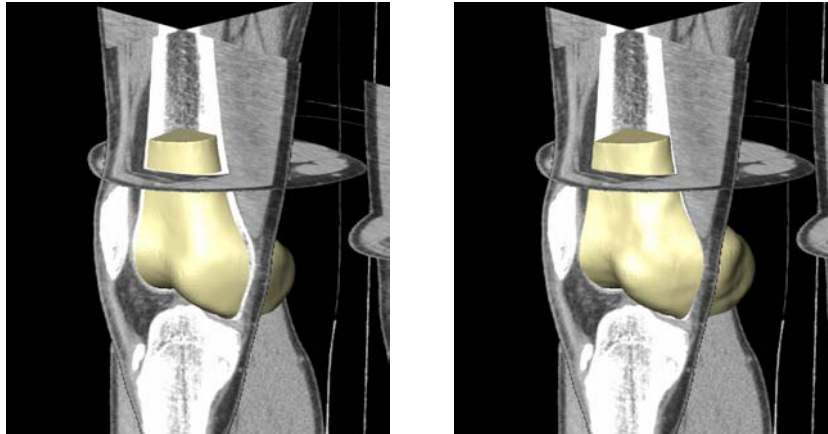


Fig. 7. Segmentation result. Initial mean model (left) and final, autonomously segmented model (right)

References

1. T. Cootes and C. Taylor, "Active shape models—'Smart snakes'," *Proc. British Mach. Vision Conf.*, pp. 266-275, 1992.
2. C. Taylor, T. Cootes, A. Hill, and J. Haslam, "Medical Image Segmentation Using Active Shape Models," *Proc. Medical Imaging Workshop, Bruxelles, Belgium*, pp. 121-143, 1995.
3. M. Kaus, V. Pekar, C. Lorenz, R. Truyen, S. Lobregt, and J. Weese, "Automated 3-D PDM Construction from Segmented Images Using Deformable Models," *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, pp. 1005-1013, Aug 2003.
4. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, Vol. 1, pp. 321-331, 1987.
5. L. Staib and J. Duncan, "Boundary Finding with Parametrically Deformable Models," *IEEE PAMI*, Vol. 14, No. 11, pp. 1061-1075, Nov 1992.
6. C. Zahn and R. Roskies, "Fourier Descriptors for Plane Closed Curves," *IEEE Transactions on Computers*, Vol. 21, No. 3, pp. 269-281, Mar 1972.
7. E. Persoon and K. Fu, "Shape Discrimination Using Fourier Descriptors," *IEEE Trans. on Sys., Man, and Cyber.*, Vol. SMC-7, No. 3, pp. 629-639, Mar 1977.
8. T. Cootes, A. Hill, C. Taylor, and J. Haslam, "The Use of Active Shape Models for Locating Structures in Medical Images," *Image and Vision Computing*, Vol. 12, No. 6, pp. 355-365, Jul 1994.
9. T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," in *Proc. Eur. Conf. Computer Vision*, Vol. 2, H. Burkhardt and B. Neumann, Eds., pp. 484-498, 1998.
10. T. Hutton, B. Buxton, P. Hammond, and H. Potts, "Estimating Average Growth trajectories in Shape-Space Using Kernel Smoothing," *IEEE Transactions on Medical Imaging*, Vol. 22, No. 6, pp. 747-753, Jun 2003.
11. C. Xu and J. Prince, "Gradient vector flow: A new external force for snakes," in *IEEE Proc. Conf. on Computer Vision and Pattern Recognition*, pp. 66-71, 1997.
12. A. Kelemen, G. Székely, and G. Gerig, "Elastic Model-Based Segmentation of 3-D Neuroradiological Data Sets," *IEEE Transactions on Medical Imaging*, Vol. 18, No. 10, pp. 828-839, Oct 1999.