Automated 3D PDM Construction Using Deformable Models

M.R. Kaus1, V. Pekar1,2, C. Lorenz1, R. Truyen3, S. Lobregt3, J. Richolt4, J. Weese1

1. Philips Research Laboratories, Division Technical Systems, Hamburg, Germany
2. Medical University of Lübeck, Institute for Signal Processing, Lübeck, Germany
3. Philips Medical Systems, MIMIT-AD, Best, The Netherlands
4. Orthopedic University Hospital Friedrichsheim, Frankfurt, Germany

Michael.Kaus@philips.com

Abstract

In recent years several methods have been proposed for constructing statistical shape models to aid image analysis tasks by providing a-priori knowledge. Examples include principal component analysis (PCA) of manually or semi-automatically placed corresponding landmarks on the learning shapes (point distribution models, PDM), which is time consuming and subjective. However, automatically establishing surface correspondences continues to be a difficult problem. This paper presents a novel method for the automated construction of 3D PDM from segmented images. Corresponding surface landmarks are established by adapting a triangulated learning shape to segmented volumetric images of the remaining shapes. The adaptation is based on a novel deformable model technique. We illustrate our approach using CT data of the vertebra and the femur. We demonstrate that our method accurately represents and predicts shapes.

1. Introduction

3D statistical shape models introduced by Cootes and Taylor [3] have been successfully employed to make image analysis tasks such as segmentation, registration and recognition more robust by providing a-priori shape knowledge [4, 10, 17]. Such models capture the variability of objects by projecting a set of high-dimensional object representations onto a lower-dimensional subspace of possible shapes, leading to a reduced set of shape parameters.

One of the most crucial aspects of shape model generation is to represent a set of learning shapes with vectors of corresponding elements. This paper describes a novel method for the automated generation of 3D statistical shape models from segmented volumetric images.

Object representation is one important aspect of shape modeling. Using spherical Fourier descriptors [10, 15] has the advantage that the statistical analysis can be applied directly to the weights of the harmonic basis functions. However, open objects or objects with holes are difficult to represent. Point distribution models (PDM) are a class of shape models where the parameter vectors are composed of the coordinates of surface points. This approach is less restrictive on topology, but requires a set of corresponding landmarks to be selected which can be consistently located from one shape to another. The number of necessary landmarks is substantial for 3D surfaces, making manual selection very tedious, time consuming and user dependent.

Recent work has investigated the possibility of automating the selection of landmarks by means of image registration. Brett et al. [2] use an ICP-based registration algorithm for object surfaces which are topological disks. Fleute et al. [6] estimate a free-form volumetric transformation to register shapes. Their algorithm involves the evaluation of the inverse transformation, which may not exist in areas with considerable shape variability or irregular volume deformations. Early work by this group [12] proposed a semi-automated landmarking method which required manual identification of a few anatomical landmarks to estimate a registration of triangulations based on thin-plate splines.

In this work we propose to automatically generate corresponding landmarks based on a novel technique for adapting a triangular mesh to a binary image. The adaptation technique is an elastically deformable model (see e.g. [13]). The mesh adaptation is governed by an external energy which drives the mesh towards the object boundaries in the binary image, and an internal energy which maintains the vertex distribution on the mesh surface. Our method has the advantage of taking the mesh connectivity into account, and improving the compactness (number of parameters) of the model by restricting artificial model variability due to a varying distribution of vertices on the mesh surface. Finally,
Our method does not require the inversion of a volumetric transformation.

The paper proceeds as follows. In Section 2 the landmarking procedure is described in detail, followed by a brief description of our implementation of the statistical model building process in Section 3. In Section 4 the method is applied to the vertebra of the spine and the femur. We demonstrate that the learning shapes are accurately and compactly represented by a PDM based on triangular meshes using a deformable model technique and that the shape models approximate unknown shapes well.

2. Establishing Surface Correspondence

The input to our method is a learning sample which consists of a set of learning shapes. Originally, the shapes are represented by volumetric image data sets where the objects of interest have been segmented. For each learning shape $i = 1 \ldots L$, we seek a triangular mesh with vertices $\{v_j\} j = 1 \ldots V$, where the $j$-th vertex corresponds to approximately the same surface location from one shape to another. A consistent triangulation of all objects in the learning set is constructed as follows:

1. Select an arbitrary template from the set of learning shapes and triangulate the object surface (Section 2.1).
2. For each of the remaining learning shapes, estimate a preliminary pose and scaling that approximately aligns the template mesh to the learning shape in the segmented volume data set (Section 2.2). This serves as an initialization for the subsequent local adaptation process.
3. Locally adapt the template mesh to the learning shape in the segmented volume data set using a deformable model based approach (Section 2.3).

This procedure yields a set of triangulated surfaces, together with point-wise corresponding vertices from one mesh to the other, which are necessary to build a PDM [3].

To simplify the notation, we omit the instance index $i$ for the remainder of this section.

2.1. Representation of Surface Models

Given a segmented object in a volumetric image, we wish to derive a triangulation with as few vertices and triangles as possible. This is important because a small number of vertices and triangles results in a faster adaptation process during automated landmarking and requires less computation for the approximation of unknown objects. For this purpose we used a triangulation method where the vertex density is proportional to the surface curvature [11], which we briefly outline below. First, the surface voxels are extracted by selecting those voxels which are 26-connected to the background. The set of surface voxels is then iteratively reduced to a subset with a dense voxel distribution in areas of high curvature (highly detailed regions) and a sparse voxel distribution in areas of low curvature (little detail). This subset is triangulated using Delaunay triangulation (e.g. [5]). Two examples are shown in Figure 1.

2.2. Rigid Alignment of a Triangular Mesh to a Volumetric Image

The approximate alignment of the template mesh with the target (i.e. learning shape) in the segmented data set is a preprocessing step for the automation of the subsequent local mesh adaptation. Since the mesh adaptation operates locally, large differences in pose and scaling could lead to false correspondence estimation. The purpose here is to estimate the $3 \times 3$ rotation matrix $R$, the translation $t$ and the scaling $s$ that aligns the template mesh with the object in the target volume. Initially, $t$ is set to translate the center of the template to the center of the bounding box of the target. $R$ and $s$ are set to identity (assuming identical orientation of the objects during image acquisition). The registration parameters are now estimated by minimization of the sum of directed gradients

$$D_t(R, t, s) = \sum_{i=1}^{T} F(sRx_i + t), \quad F(x_i) = n_i^T g(x_i), \quad (1)$$

where $\{x_i| i = 1 \ldots T\}$ is the set of triangle centers, $n_i$ is the triangle normal pointing outwards, and $g(x_i)$ is the image gradient at location $x_i$. Defining segmented objects to be bright on a dark background, the image gradients on the object boundary point from the outside to the inside. This results in large negative values of $F$ for strong image gradients in direction of the triangle normal (directed gradient) [18]. Minimization of the function is implemented with the downhill-simplex algorithm as in [14].
2.3. Local Adaptation of a Triangular Mesh to a Volumetric Image

Our approach is closely related to the deformable models introduced by Kass et al. [9] and reviewed for medical image analysis purposes in [13]. Given an approximate initial pose and scaling, the deformable model (i.e. the template mesh) is iteratively deformed to optimally adapt the surface of the target (i.e. the learning shape in the segmented volumetric data set). The deformation is guided by the minimization of an energy function

$$E = E_{ext} + \alpha E_{int}.$$  \hspace{1cm} (2)

The external energy $E_{ext}$ drives the template mesh towards the surface of the learning shape, the internal energy $E_{int}$ restricts the flexibility of the deforming mesh, and the parameter $\alpha$ weights the relative influence of both terms. The adaptation process consists of a two-step iteration: First, at each triangle center, find the point along the triangle normal that restricts the flexibility of the deforming mesh, and the parameter $\alpha$ weights the relative influence of both terms. The adaptation process consists of a two-step iteration: First, at each triangle center, find the point along the triangle normal that best matches a boundary criterion. Second, calculate the positions of the new triangle centers by minimizing Equation (2). In the following we will describe these steps in detail.

Local Surface Detection  Surface detection is carried out in the following way: For each triangle center $x_i$ of the mesh, find the point $\hat{x}_i$ along the triangle normal $n_i$ with the optimal combination of the boundary detection function $F$ (Equation (1)) and distance $j\delta$ to the triangle center according to

$$\hat{x}_i = x_i + \delta n_i \arg\min_{j=-l,...,l} \{ D j^2 \delta^2 - F(x_i + j\delta n_i) \},$$ \hspace{1cm} (3)

where $2l+1$ is the number of points investigated, $\delta$ specifies the distance between two points on the normal, and $D$ controls the tradeoff between the directed gradient and the distance. By combining the boundary value with the distance information, surface points next to the mesh are preferred and mesh adaptation becomes more stable. If we searched for the point with the maximum directed gradient ($D = 0$), the search length $l$ would become a crucial parameter. For small $l$, the mesh would require a rather accurate initialization, and for large $l$, there may be no unique solution to the search.

Internal and External Energy Computation  The external energy term drives the triangle centers $\{x_i\}$ towards potential surface points $\hat{x}_i$ in the target image. It is defined as

$$E_{ext}(x) = \sum_{i=1}^{T} w_i \left( \frac{g(x_i)}{\|g(x_i)\|} (\hat{x}_i - x_i) \right)^2,$$ \hspace{1cm} (4)

where $T$ is the number of triangles. An important property is that the energy is invariant to movements of the triangle centers parallel to the plane perpendicular to the gradient vector $g$. Since the surfaces of segmented objects are iso-contours, the vertices can move along the object surface without a change of external energy, which considerably reduces the typical problem of mesh points getting stuck at local minima and allows the mesh to relax according to the internal energy. Without this property, the detected surface points would directly attract the triangle centers of the mesh. As a consequence, a triangle center could almost not move anymore, once a surface point has attracted it. For that reason, the mesh may remain attached to the false (i.e. closest) position on the object boundary, which is frequently detected at the beginning of the adaptation process.

The weights $w_i = \max \{ 0, F(\hat{x}_i) - D j^2 \delta^2 \}$ (see Equation (3)) give the most promising surface points $\hat{x}_i$, the largest influence during mesh reconfiguration which slightly reduces the necessary number of iterations.

The internal energy restricts the movement of the mesh vertices. Since the distribution of vertices on the initial template mesh is optimal with respect to surface-curvature, we wish to maintain that distribution. Another important aspect is to keep the vertex displacement as small as possible. The goal is to avoid unnecessary vertex displacement resulting in higher model variance. For that purpose the internal energy is defined to minimize deviation from the original distances between neighboring vertices, i.e.

$$E_{int}(x) = \sum_{j=1}^{V} \sum_{k \in N(j)} (v_j - v_k - sR(v_j - v_k))^2,$$ \hspace{1cm} (5)

where $N(j)$ is the set of indices of the neighboring vertices $\{v_k\}$ of vertex $v_j$, and $\{v_j\}$ are the vertices of the template mesh. Explicit estimation of the scaling $s$ and the rotation $R$ that align the template with the deforming mesh makes the internal energy invariant to pose and scaling of the target.

Energy Minimization  The quadratic energy in Equation (2) is minimized using a two-step optimization procedure. First, estimate the orientation $R$ and the scaling $s$ that aligns the edges of the deforming model to the edges of the initial template mesh. This is achieved with a closed-form point-based registration method based on SVD decomposition (e.g. [8]). Second, estimate $\{v_j\}$ given $R$ and $s$. Since the energy in Equation (2) is a quadratic function with respect to $\{v_j\}$, the minimization problem can be reduced to the solution of a sparse system of linear equations (typically in the order of 2000–3000 parameters). This linear system can be efficiently solved with a conjugate gradient method (e.g. [7]).
3. Construction of a Point Distribution Model

Having established point-wise correspondences between the vertices \( \{v_i^j\} \) for all shapes, we wish to build a statistical model invariant to pose and scaling, since these parameters are not intrinsic shape characteristics, and are not described by a linear PCA model. A common approach to building a shape model excluding pose and scaling from the model is an Procrustes analysis [1], which we have implemented with an iterative approximation by a two-step minimization of

\[
D_p = \sum_{i=1}^{L} \sum_{j=1}^{V} \|s_i R_i v_i^j + t_i - m^0_i\|_2.
\]  

(6)

\( R_i, t_i, s_i \) are the pose and scaling parameters that map the vertices \( \{v_i^j\} \) of the \( i \)-th instance onto the model mean \( m^0 \).

Initializing \( R_i, t_i, s_i \) to the unity transform, Equation (6) is optimized by iteration of the following two steps: First, estimate the components of the eigenmodes using PCA (see e.g. [3]). Second, align the vertices \( \{v_i^j\} \) of each learning sample with the mean \( m^0 \). This is done by estimating \( R_i, t_i, s_i \) using a point-based registration method based on SVD [8]. The optimization is terminated when the distance between the corresponding eigenmodes of two iterations is smaller than a threshold \( \epsilon \).

4. Results

To validate our method, we applied the modeling procedure to two distinct anatomical structures, the vertebra of the spine and the femur. These objects were selected because they differ in the degree of shape detail and shape variability.

The two main modeling steps, i.e. establishing corresponding landmarks on the learning shapes and deriving a statistical model, were analyzed in separate experiments. The first experiment analyzed the landmarking procedure. For this purpose we assessed how close the landmarks (i.e. the vertices of the adapted template mesh) were adapted to the surface of a learning object in the segmented volumetric images. The second experiment analyzed how well the statistical model generalized, by measuring how close an unknown shape, which was not in the learning sample, can be approximated by the PDM.

Although the mesh adaptation was based on segmented images and triangular meshes, we present the surface meshes with grey value images for illustrative purposes. For better viewability, the vertebra images were masked with the segmented images.

4.1. Image Data

**Femur Study** The femur study, a medium detailed shape, consisted of volumetric CT scans (voxel size: \( 0.6 \times 0.6 \times 2.0 \) mm\(^3\) ) of 13 individuals suffering from slipped capital femoral epiphysys (SCFE). This affected the angle (120–150°) between the femur shaft and neck resulting in substantial shape variability. In order to increase the sample size, the left femur was mirrored across the mid-sagittal plane as suggested in [16], leading to 26 samples of the right femur.

**Vertebra Study** This study consisted of volumetric CT scans (voxel size: \( 0.5 \times 0.5 \times 2.0 \) mm\(^3\) ) of 8 normal patients from which 32 vertebrae (L1–L4) were extracted. The vertebra is a highly detailed structure with medium variability.

All images were segmented with simple semi-automated volume editing tools and slice-by-slice correction using out-

![Figure 2. Orthogonal grey-value cross-sections of the template mesh (white outline) adapted to the segmented image of a vertebra (top) and a femur (bottom).](image-url)
4.2. Accuracy of Sample Representation

Automated landmarking of object surfaces with our method is achieved by adapting a triangular mesh to the surface of learning shapes in segmented volumetric images.

For the adaptation procedure, we set $D = 2$, $\delta = 1 \text{ mm}$, $l = 10$, $\alpha = 33.33$, and 50 iterations. The mean computation time for one adaptation was 30 seconds (Sun UltraSPARC II, 433 MHz).

To assess how close the landmarks (i.e. the mesh vertices) are to the object surface, we calculated the Euclidean distance between the vertices of the adapted triangular mesh and the surface of the corresponding learning shape in the segmented volumetric image. This was done by (i) extracting the surface voxels of the object in the segmented image data by selecting those voxels which are 26-connected to the background, (ii) computing an Euclidean distance transform and (iii) evaluating the distances at the vertex coordinates of the adapted mesh.

Figure 3. Mean and maximum distance between the landmarks (i.e. the vertices of the adapted mesh) and the object surface in the segmented image.

In some areas, sharp angles between neighboring triangles (but not surface foldings) occur, which can be explained with the lack of a smoothness constraint in the internal energy. Representation of the femur shapes was achieved with a maximum distance of 2.7 mm and a mean distance of 0.8 mm. The mean approximation errors correspond to 1-2 voxels, which is in the order of the variability typical for manual segmentations.

4.3. Accuracy of Shape Model Generalization

In this section we assessed how well our model approximated an unknown shape, i.e. a shape which was not in the learning sample at the time of the model construction.

The modeling procedure converged after approximately 30 iterations of Equation (6), and the mean computation time was 70 seconds (Sun UltraSPARC II, 433 MHz).

The common approach to assessing the generalizability of a statistical shape model is the leave-one-out experiment.

Figure 4. Orthogonal grey-value cross-sections of an unknown femur and the mesh approximated by the shape model.
Given a learning set, a model is built from all shapes but one. The model is then adapted to the shape left out. The approximation error between the left out shape and the approximation using the shape model is calculated. This is repeated for all shapes in turn.

The approximation \( \{w^i_j\} \) to a shape is calculated by iteratively i) projecting the unknown shape vector \( \{v^i_j\} \) onto the model eigenmodes, i.e.

\[
\hat{w}^i_j \approx m^0 + \sum_{k=1}^{M} p^i_k m^k,
\]

and ii) aligning the unknown shape with the approximation using point based registration [8].

The approximation error was defined as the mean and maximum Euclidean distance over all learning shapes, i.e.

\[
D_{\text{mean}} = \frac{1}{VL} \sum_{i=1}^{L} \sum_{j=1}^{V} \|w^i_j - v^i_j\|,
\]

\[
D_{\text{max}} = \max \|w^i_j - v^i_j\|.
\]

Figure 4 illustrates the ability of the models to approximate an unknown shape. Figure 5 shows the approximation accuracy for the leave-one-out experiments. In our approach we established the corresponding landmarks prior to pose and scaling estimation during the statistical modeling. This assumes that the position of the landmarks does not change much with different pose and scaling, which is justified by the high accuracy of the model.

Both vertebra and femur are accurately approximated by the model with \( (D_{\text{mean}}=1.48 \text{ mm}; D_{\text{max}}=4.90 \text{ mm}) \) and \( (D_{\text{mean}}=1.63 \text{ mm}; D_{\text{max}}=5.32 \text{ mm}) \). The difference in accuracy can be explained with the slightly better in-plane resolution of the vertebra \((0.5 \times 0.5 \text{ mm}^2)\) compared with the femur data \((0.6 \times 0.6 \text{ mm}^2)\). The higher degree of detail of the object surface has no obvious effect on the approximation accuracy, but is reflected by the number of vertices and triangles (vertebra: 1048, 2096; femur: 701, 1369).

Figure 6 shows the portion of shape variability (the relative cumulative shape variance) within the learning sample captured by the first \( n \) eigenmodes. To explain 90 % shape variability, 10 eigenmodes are required for the vertebra and 5 eigenmodes for the femur. This can be explained with a higher variability of the vertebra due the higher detail of the object surface.

5. Conclusion

In this paper we have presented a method to automatically construct statistical 3D point distribution models from...
segmented volumetric images. Corresponding landmarks on the surface of the shape are established automatically by the adaptation of a triangular mesh to segmented volumetric images.

Results from data sets of anatomical structures (femur and vertebra) demonstrate accurate representation of the objects (mean surface distance 0.8 mm) for the landmarking procedure, accurate approximation of objects with the shape model (femur: $D_{\text{mean}}=1.63$; vertebra: $D_{\text{mean}}=1.48$ mm; 10 eigenmodes), and good compactness of the model. Also, our method is fast (approximately 15 minutes computation time for modeling the datasets presented here), which is important for large object samples.

The internal energy of the deformable model prevents artificial vertex variability. Extending the energy formulation to preserve angles between triangles during adaptation may further improve the quality of the model.

So far we have applied our method to bone structures. It is an open question for future research to investigate i) the capabilities of automated correspondence finding and ii) the suitability of point distribution modeling based on principal component analysis to model more flexible structures. In this context it is also interesting to evaluate how well corresponding points reach the same anatomical location in different images.

Our results are difficult to compare with those of other reports [2, 4, 6, 10], since other objects were modeled (e.g. brains, hands or faces) and no common validation criteria have yet been established. Nevertheless, this would be an interesting future investigation, since our approach is sufficiently general to be applied to modeling tasks from other 2D or 3D application domains.

Acknowledgments

We thank Prof. Dr. W. Mali, Prof. Dr. B.C. Eikelboom and Dr. J.D. Blankensteijn (University Hospital Utrecht) for providing the vertebrae CT data. We thank Dr. J. Kordelle, Brigham & Women’s Hospital for help with the CT femur data. The algorithm was implemented on an experimental version of the EasyVision workstation from Philips Medical Systems.

References