Statistical shape models for 3D medical image segmentation: A review

Tobias Heimann *, Hans-Peter Meinzer

Div. of Medical and Biological Informatics, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

1. Introduction

In the last two decades, model-based segmentation approaches have been established as one of the most successful methods for image analysis. By matching a model which contains information about the expected shape and appearance of the structure of interest to new images, the segmentation is conducted in a top-down fashion. Due to the inherent a-priori information, this approach is more stable against local image artifacts and perturbations than conventional low-level algorithms. While a single template shape is an adequate model for industrial applications where mass-produced, rigid objects need to be detected, this method is prone to fail in case of biological objects due to their considerable natural variability. Information about common variations thus has to be included in the model. A straight-forward approach to gather this information is to examine a number of training shapes by statistical means, leading to statistical shape models (SSMs).

In this paper, we will review methods and procedures for generating, training and employing statistical models of shape and appearance for 3D medical image segmentation. Specifically, we will discuss work on discrete, parametric models which can be trained from a set of example data. Probably the best-known methods in that area are the Active Shape models (Cootes et al., 1995) and Active Appearance models (Cootes et al., 2001) by Cootes et al. In addition, we will discuss related concepts and alternative approaches, all within the context of statistical shape models. Due to the constantly increasing importance of 3D imaging and the urgent need for segmentation in that particular area, we will concentrate on methods for volumetric images. However, many modeling methods have only been applied to 2D so far. To the extent that such methods can be generalized or extended to the 3D case, we have included them in the present review. We have furthermore included some methods which have shown to be very successful in 2D, but which are technically not feasible in 3D – simply to emphasize the difference.

The objective of this article is to provide the reader with a summary of the current state of the art with regard to 3D statistical shape models, to demonstrate what has been done until now, but also to present some ideas of what might yet be done. To ensure comprehensive coverage, we have screened all publications included in IEEE Transactions on Medical Imaging and Medical Image Analysis during the last 10 years for articles related to shape models. In addition, we have included a large number of articles from other international journals, but also numerous conference and workshop papers which present good ideas, but which have not been published in any journal yet. Our main source of references was the Internet; we have searched for the terms shape model and statistical model on PubMed, IEEE-Xplore, Citeseer and Google. We have also followed the references encountered in papers from these sites, until we had collected a comprehensive library of more than 400 articles on the topic. In case we encountered several papers from one author about the same subject, we generally picked the most detailed one for this review.

1.1. Related work

Before reviewing statistical shape models, let us first define what we regard as related work that will not be discussed further in this article.
1.1.1. Freely deformable models
Kass et al. started the use of deformable models for image segmentation in their seminal snakes paper (Kass et al., 1988). Their main idea is that model evolution is driven by two energies: an external energy that adapts the model to the image data and an internal energy that stabilizes its shape based on general smoothness constraints. Shortly afterwards, Terzopoulos et al. (1988) generalized the concept which initially had only been applied to 2D examples to 3D shapes. In (Delingette et al., 1994), Delingette introduced the deformable simplex mesh, which features a stable internal energy that can easily be customized to deform toward a specific template shape. Using a different approach, McInerney and Terzopoulos (1999) presented a method of how to implement topology changes for deformable surfaces. After almost two decades of deformable models, several review articles on the topic have been published, notably by McInerney and Terzopoulos (1996), Jain et al. (1998), and Montagnat et al. (2001). We disregarded these methods in this review because the underlying deformation algorithms do not incorporate learned constraints of shape variability. Although freely-deformable models can be customized to represent specific shapes (and often are), the stabilizing forces or energies are based on general smoothness properties and are not driven by statistical information.

1.1.2. Level-sets
Level-sets were introduced by Osher and Sethian (1988) and made popular for computer vision and image analysis by Malladi et al. (1995). They feature an implicit shape representation and can be employed with regional or edge-based features. Leventon et al. (2000) extended the original energy formulation by an additional term which deforms the contour towards a previously learned shape model. A frequent criticism is that the signed distance maps which the shape model is based on, do not form a linear space, which can lead to invalid shapes if training samples vary too much. Nevertheless, the approach quickly gained popularity and was extended in several directions, among others by Tsai et al. (2003) who employ Leventon’s modeling method with a region-based energy functional. Recently, Pohl et al. (2006) presented a method of embedding the signed distance maps into the linear LogOdds space, which could solve the modeling problems. To keep this review at a reasonable length, we had to ignore level-set theory and techniques: The conceptual differences between the implicit representation and the discrete models we intend to focus on would have required a special treatment for all following sections. For an overview of statistical approaches to level-set segmentation – including prior shape knowledge – we refer the reader to the recent survey by Cremers et al. (2007).

1.2. Structure of this article
In order to present a systematic overview of the topic, we have divided this article into several parts, each highlighting a specific aspect of statistical shape models: In Section 2, we will start with presenting different possibilities of how to represent shapes for statistical analysis. Subsequently, we will explain how to extract the principal modes of variation from a set of training shapes in Section 3. A general requirement for this step is that the correspondences between all shapes of the training set are known, a topic which will be discussed in Section 4. After that, in Section 5, we will present techniques to model the appearance of the examined object. The different algorithms that employ shape and appearance models for image analysis and segmentation will be discussed in Section 6. Subsequently, we will present an overview of the areas of application in medical imaging, which have been tackled with three-dimensional SSMs in Section 7. Before concluding the review, we will recapitulate the main points and predict future developments in Section 8.

2. Shape representation
Training data for SSMs in the medical field will most likely consist of segmented volumetric images. Depending on the segmentation method used, the initial representation might be binary voxel data, fuzzy voxel data (e.g. from probabilistic methods), or surface meshes. Data originating from other sources of acquisition, e.g. surface scanning, might be represented differently. In any case, all shape representations can be converted into each other, and the choice of shape representation is the first fundamental decision when designing statistical shape models. Most of the subsequent steps (described in the following sections) depend on this initial decision, and many methods are technically limited to certain representations.

2.1. Landmarks and meshes
Probably the simplest and at the same time the most generic method used to represent shapes is a set of points distributed across the surface (which can be extracted from volume data by, e.g. the Marching Cubes algorithm). Coordinates for all k points are concatenated to one vector x that describes the shape:

\[ x = (x_1, y_1, z_1, \ldots, x_k, y_k, z_k)^T \] (1)

In SSM literature, the involved points are often also referred to as landmarks, although they do not need to be located at salient feature points such as the common definition for anatomical landmarks. To point out this contradiction, some authors also call them semi-landmarks. Often, additional connectivity information between the points is stored to allow for reconstruction of the surface and calculation of normal vectors, which is important for many search algorithms (see Section 6). A point set with connectivity information is called a mesh. Landmarks have been used extensively for the statistical study of biological shape by Kendall (1989), Bookstein (2003) and others. For the use of landmarks as basis of an SSM, Cootes et al. (1992) have coined the name Point Distribution Models (PDMs), which has become quite popular in literature and is often used synonymously to landmarks. The majority of current shape models are based on PDMs, and they will form a strong focus for the remainder of this article.

2.2. Medial models
Medial models or skeletons have been used by Blum (1973) to describe biological shapes back in the 1970s and are commonly utilized in image analysis. They represent objects by their centerlines and the corresponding radii, often leading to a more compact description than landmarks. Pizer et al. present a medial model with a coarse-to-fine representation for two dimensions in (Pizer et al., 1999): Basically, it consists of a collection of points on the centerlines and vectors pointing from there toward the boundary. This approach was later extended to 3D in (Pizer et al., 2003) and termed m-rep. Recently, Yushkevich et al. (2006) presented a continuous version of m-reps. Since their introduction, m-reps have been used successfully for a number of medical image processing tasks, including segmentation, registration and shape discrimination.

2.3. Other representations
A multitude of other possibilities to represent shape exists, which we will only briefly mention here: Staib and Duncan...
align a group of shapes to their unknown mean, this procedure is run iteratively, resulting in the GPA. In addition to the references provided above, details of the method can also be found in (Dryden and Mardia, 1998). It should be pointed out that GPA is not resistant to outlier points, and there have been some experiments of replacing the Euclidean distance metric in the scheme with the $L_1$ or $L_\infty$ norm (see Larsen and Eriksen, 2001). Moreover, Ericsson and Karlsson (2005) proposed aligning the rotation of all training shapes using Minimum Description Length optimization, similar to the population-based optimization of correspondences (more on this topic in Section 4.5).

Standard GPA scales all training samples $x_i$ to minimize their Euclidean distance to the mean $\bar{x}$. This procedure leads to non-linearities between the individual shapes which (depending on the amount of variation) will result in inaccuracies in the following decomposition into linear modes of variation. The solution is to project all shapes into the tangent space of the mean by scaling with $1/(\bar{x}^Tx)$ for more information on tangent space, refer to Dryden and Mardia (1998); a short description of the implications for model building is also provided in (Cootes and Taylor, 2004).

3. Shape model construction

Constructing a statistical shape model basically consists of extracting the mean shape and a number of modes of variation from a collection of training samples. Obviously, the methods employed strongly depend on the chosen shape representation. Due to the dominant role of landmark-based point distribution models, in this section we will concentrate on PDMs and only briefly deal with the corresponding procedures for other representations. An essential requirement for building shape models with PDMs is that landmarks on all training samples are located at corresponding positions. We will assume this requirement as fulfilled for this section and will discuss the correspondence issue in detail in Section 4.

3.1. Alignment

Shape is defined as a property which does not change under similarity transformations, i.e. it is invariant to translation, rotation and scaling. In general, shape changes induced by these global transformations should not be modeled by an SSM in order to keep the model as specific as possible. For some applications in medical image analysis, however, changes e.g. in size may be treated as part of the biological variation (as done in Neumann and Lorenz, 1998; and Mardia, 1998). It should be pointed out that GPA is not resistant to outlier points, and there have been some experiments of replacing the Euclidean distance metric in the scheme with the $L_1$ or $L_\infty$ norm (see Larsen and Eriksen, 2001). Moreover, Ericsson and Karlsson (2005) proposed aligning the rotation of all training shapes using Minimum Description Length optimization, similar to the population-based optimization of correspondences (more on this topic in Section 4.5).

Standard GPA scales all training samples $x_i$ to minimize their Euclidean distance to the mean $\bar{x}$. This procedure leads to non-linearities between the individual shapes which (depending on the amount of variation) will result in inaccuracies in the following decomposition into linear modes of variation. The solution is to project all shapes into the tangent space of the mean by scaling with $1/(\bar{x}^Tx)$ for more information on tangent space, refer to Dryden and Mardia (1998); a short description of the implications for model building is also provided in (Cootes and Taylor, 2004).

3.2. Dimensionality reduction

After alignment, the next step is to reduce the dimensionality of the training set, i.e. to find a small set of modes that best describes the observed variation. This is usually accomplished using principal component analysis (PCA) (Jolliffe, 2002). Following Eq. (1), every aligned training shape is described by $3k$ point coordinates in the vector $x_i$. The mean shape can then be formed by simply averaging over all $s$ samples:

$$\bar{x} = \frac{1}{s} \sum_{i=1}^{s} x_i$$

(2)

The corresponding covariance matrix $S$ is given by:

$$S = \frac{1}{s-1} \sum_{i=1}^{s} (x_i - \bar{x})(x_i - \bar{x})^T$$

(3)

An eigendecomposition on $S$ delivers the max($s-1, 3k$) principal modes of variation $\phi_m$ (eigenvectors) and their respective variances $\lambda_m$ (eigenvalues) – Fig. 1 visualizes the three largest modes for the example of a liver SSM. Instead of using eigenanalysis on the covariance matrix, it is also possible to calculate $\phi_m$ and $\lambda_m$ by a singular value decomposition (SVD) on the aligned landmark matrix $L = (x_1 - \bar{x} \cdots (x_s - \bar{x}))$. From a computational point of view, this method is to be preferred due to the higher numerical stability. As both methods set up their principal modes according to least-squares estimation, they are susceptible to outliers – sometimes an entire sample might be outside the expected distribution, sometimes only a small number of landmarks for a given sample. One algorithm that is able to handle both types of outliers is the robust PCA presented by De la Torre and Black (2003).

In any case, the resulting modes of variation are ordered by their variances so that $\lambda_1 > \lambda_2 \cdots \lambda_{3k-1}$. In the next step, it is possible to approximate every valid shape by a linear combination of the first $c$ modes:

$$x = \bar{x} + \sum_{m=1}^{c} b_m \phi_m$$

(4)

In many cases, $c$ is chosen so that the accumulated variance $\sum_{m=1}^{c} \lambda_m$ reaches a certain ratio $r$ of the total variance $\sum_{m=1}^{3k} \lambda_m$. Common values for $r$ are 0.9–0.98. Alternatively, it is possible to determine $c$ by looking for a characteristic drop in variance (the so-called elbow) in the screen graph of eigenvalues (Jolliffe, 2002). Another option is the use of parallel analysis (Horn, 1965), in which resulting eigenvalues are compared to a PCA of random data.
Vector $b$ in Eq. (4) holds the shape parameters. To constrain the allowed variation to plausible shapes, $b$ has to be limited to a certain interval. A common method is to treat all modes as independent distributions and constrain each $b_m$ to lie inside $[-3\lambda_m, 3\lambda_m]$. Alternatively, viewing all modes as one multi-variate distribution, $b$ can also be constrained to lie inside a hyperellipsoid, i.e.

$$
\left( \sum_{m=1}^{\infty} \frac{b_m^2}{\lambda_m^2} \right) \leq M_t
$$

where $M_t$ is a threshold chosen from the $\chi^2$ distribution. If the shape variation does not follow a Gaussian distribution, the presented constraints for $b$ will possibly result in invalid shapes. A more accurate estimate of valid parameter constellations can be obtained by Gaussian mixture models, as proposed by Cootes and Taylor (1996). Recently, Li and Ito (2005) presented a method of parameter restriction which is based on multi-dimensional tables that are constructed from the training data.

In general, PCA results in global modes which influence all variables simultaneously, i.e. varying one mode will affect all landmarks of the shape model. For models used for shape analysis or diagnostic purposes, it is usually beneficial to have more isolated effects, which are more intuitive to interpret: Each mode should only affect a limited, preferably locally clustered number of landmarks. This property is called sparsity: a straight-forward method to obtain sparser modes is to employ the Orthomax as proposed by Stegmann et al. (2006). As the name indicates, Orthomax rotates the PCA modes to increase sparsity, while maintaining the orthogonality of components and thus Euclidean space. If the latter property is not essential, sparse PCA by Sjöstrand et al. (2007) is an alternative approach which still produces near-orthogonal components. Other decomposition methods do not incorporate any orthogonality criteria at all but directly strive to maximize sparsity. The most popular one is probably the independent component analysis (ICA) (Hyvärinen et al., 2001), which does not assume a Gaussian data distribution and delivers statistically independent projections. Among others, it was applied to shape modeling by Üzümcü et al. (2003) and Suinesiaputra et al. (2004) for the characterization of myocardial diseases. Another technique to obtain sparse modes of variation is the maximum autocorrelation factor (MAF) analysis as used by Hilger et al. (2003). An extensive comparison between PCA, MAF and minimum noise fraction (MNF – another non-Euclidean decomposition) is presented by Larsen and Hilger (2003). They also prove that MAF is actually equivalent to ICA.

As the non-PCA approaches for dimensionality reduction deliver modes that are not directly specified by variance, there is no natural ordering for these modes. In order to still be able to make a sensible selection of modes for the SSM, a number of different techniques can be employed (Stegmann et al., 2006; Sjöstrand et al., 2006; Üzümcü et al., 2003). When using non-orthogonal decompositions, it is important to realize that standard methods for shape fitting by least squares projection do no longer work, thus impacting on the model search algorithms (Section 6).

Apart from exploring different linear decompositions, there has also been work on building non-linear SSMs. Non-linear models allow for natural representation of variations based on bending and rotations, which can only inadequately be approximated by linear models. Sozou et al. suggested non-linear PCA based on polynomial regression (Sozou et al., 1994) and multi-layer perceptrons (Sozou et al., 1995) for this purpose. Later, Twinning and Taylor (2001) proposed the use of Kernel PCA instead, which they claim to be more general than other methods. Recently, there is a growing interest in using Kernel PCA for implicit shape analysis (Cremers et al., 2003; Dambreville et al., 2008; Rathi et al., 2006). Major reason for this development is that non-linear decompositions can cope with the level-set specific problem mentioned in Section 1.1.2, i.e. that signed distance maps do not form a linear vector space.

3.3. Enlarging variations

The power of a statistical model rises and falls with the quantity of available training data. In case of 3D SSMs, this quantity is almost always too low, as in practice there are rarely enough training images available and their required manual segmentation is very cumbersome and time-consuming. Even if enough data exists, it might not be fully used due to a poor estimate of the required sample size. This results in models that are over-constrained, i.e. the restrictions imposed on the deformations do not enable them to adapt accurately to new data. Evidence for insufficient training data is better segmentation accuracy when additional training samples are used (as demonstrated e.g. for liver segmentation (Lamecker et al., 2004)), and a considerable error for leave-one-out reconstruction even for large training databases (reported e.g. for hippocampus (Davies, 2002), liver, lung and prostate shapes (Heimann, 2007)).

Cootes and Taylor (1995) use finite element methods to calculate vibrational modes for each training shape, which are used to generate a number of modified shape instances. All variants are subsequently included in the construction of the SSM, which leads to a model featuring original and synthetic variations. Depending on the amount of original training data, it is possible to adapt the number of generated synthetic shapes. In a subsequent publication (Cootes and Taylor, 1996), the authors present a simpler technique for the constraint of valid parameter constellations.
for the same purpose: The synthetic variation is added directly to the covariance matrix $S$, coupling the movement of neighboring points. The same method is used by Wang and Staib (2000) to build more flexible models. Lötjönen et al. (2005) propose several additional methods of how to synthetically modify training shapes. From their set, the best results are delivered by a non-rigid movement strategy which deforms shapes randomly by local warping.

A different approach to increase model flexibility is to divide the SSM into several, independently modeled parts. The rationale behind this is that smaller parts exhibit less variation, which can be captured with fewer training samples than the variation for the full shape. For the segmentation of aortic aneurysms, de Bruijne et al. (2003) model cross-sections of the vessel by one SSM, while variations along the axis are captured by a second model. Davatzikos et al. (2003) employ the wavelet transform to organize their model into a hierarchy of several parts: The lower bands of the transform correspond to more global shape changes, the higher bands to more local ones. Each band is modeled independently from the other ones. Another modeling scheme based on mesh partitioning is subsequently presented by Zhao et al. (2005): Again, each part of the mesh is modeled separately, but parameters for the individual parts are connected by curves in a combined shape space. The authors claim that limiting these curves to similar patterns as encountered in the training set helps to prevent invalid shapes of the SSM.

### 4. Shape correspondence

Modeling the statistics of a class of shapes requires a set of appropriate training shapes with well-defined correspondences. Depending on the chosen representation, the methods of how to best define these correspondences vary. In any case, establishing dense point correspondences between all shapes of the training set is generally the most challenging part of 3D model construction, and at the same time one of the major factors influencing model quality (the other one being the local gray-value appearances). Even for 2D models, manual landmarking is getting increasingly unpopular, not only due to the tedious and time-consuming expert work required, but also due to the lack of reproducibility of the results. In 3D, these arguments weight even heavier, since typically, a much larger amount of landmarks is needed and the correspondences are much harder to pinpoint on 3D shapes – even for experts.

In principle, all algorithms that automatically compute correspondences actually perform a registration between the involved shapes. In this section, we have categorized all methods according to the type of registration process involved, i.e. matching a mesh to another mesh is one category, matching a mesh to an image volume another etc.

#### 4.1. Mesh-to-mesh registration

The straightforward solution to landmark creation in 3D is to work directly on the training meshes. Fig. 2 shows the general outline of this scheme. A number of established algorithms exists for surface matching, popular ones being the Iterative Closest Point (ICP) algorithm by Besl and McKay (1992) and the Softassign Procrustes by Rangarajan et al. (1997): Both accept two surfaces with potentially different number of vertices as input and deliver the optimal similarity transformation from one surface to the other as a result. Many of the simpler methods of determining correspondences for SSMs directly rely on these algorithms, e.g. by choosing an arbitrary shape as reference and registering it to all others, thus creating all necessary landmarks (Heinze et al., 2002; Josephson et al., 2005; Vos et al., 2004). One of the shortcomings of this approach is the bias induced by the choice of reference shape – even if the procedure is repeated a second time with the mean shape as reference, as suggested by Vos et al. (2004). A solution to this problem is the method described by Brett and Taylor (1999), based on the ideas published by Hill et al. (2000): A symmetric version of the ICP is used to build a binary tree of well-matching shapes, and landmarks are propagated from the root to the leaves. As pointed out by Frangi et al. (2001), this method requires that all possible combinations of matches are scored in advance, but the extra work seems well worth in light of the gained independence from reference shapes. Another improvement to this class of correspondence detection is modification of the similarity metric of the point matching algorithm to include additional features like crest lines (Subsol et al., 1998), normal vectors (Brett and Taylor, 2000) or local shape and pattern information (Cauce and Taylor, 2001).

The largest draw-back of using a standard point matching algorithm like ICP for landmark creation is the restriction to similarity transformations. While this may be sufficient to match relatively similar training shapes such as bones, it is not adequate for a population of shapes featuring larger variations: In such cases, determination of corresponding points by proximity alone can not only lead to obviously wrong correspondences, but also to non-homeomorphic mappings and thus to flipping triangles in the mesh. Among the first to propose non-rigid registration of training meshes for building SSMs were Subsol et al. (1998) using B-Splines and Fleute et al. (1999) using multi-resolution Octree splines (Szeliski and Lavallée, 1996). An alternative to spline-based registration is finding a certain number of matching shape features on all meshes and mapping the other points by methods assuring a homeomorphism: Shelton (2000) developed a multi-resolution approach to match a number of surfaces by minimizing a cost function based on similarity, structure and prior information. In (Wang et al., 2003), Wang et al. match a limited number of landmarks by local surface geometry and determine the other points by geodesic interpolation. Yet another possibility is to use pattern recognition techniques to determine correspondences: Ferrarini et al. (2007) determine interesting parts of a surface by way of unsupervised clustering and use a classifier similar to self-organizing maps to detect the corresponding parts on other shapes. The approach presented by Pitiot et al. (2007) uses dynamic programming and pattern matching to find corresponding points according to features learned from a set of sparsely annotated training examples. The work looks very promising, but while the authors claim that it works in arbitrary dimensions, examples are only given for 2D and it is not clear how the method generalizes to surfaces.

![Fig. 2. Basic scheme for mesh-to-mesh registration to determine point correspondences: First, surfaces are extracted from all training images. In a next step, a landmarked template mesh is fitted to all these surfaces, propagating its landmarks to all samples.](image-url)
If a sparse set of user-generated landmarks is available, these known correspondences can be used to guide a Thin-Plate-Spline deformation as initial registration (Hutton et al., 2001; Lorenz and Krahnstöver, 2000; Paulsen et al., 2002; Paulsen and Hilger, 2003). However, in order to match the surfaces exactly, an additional step is needed: The simplest solution is to match the spatially closest points (Hutton et al., 2001; Paulsen et al., 2002, but this can lead to non-homeomorphic mappings and again to flipping triangles. For this reason, correspondences should be regularized somehow: Possibilities include mesh relaxation with mass-spring-models as proposed by Lorenz and Krahnstöver (2000) or using Markov Random Fields to regularize the deformation field as suggested by Paulsen and Hilger (2003). In any case, some 10–20 landmarks have to be specified manually on each training shape beforehand – depending on the complexity of shapes and number of samples this might either be relatively easy to accomplish or next to unfeasible.

4.2. Mesh-to-volume registration

The 3D SSMs in the medical imaging area are almost exclusively based on imaging modalities such as CT, MRI, or 3D-US, i.e. the original data representation of the training shapes is not a mesh but rather a segmented volume. Therefore, a different approach to landmark generation is adapting a deformable surface model to these volumes. The correspondences are then defined by the vertex locations of the deformable template after the surface evolution has converged. Refer to Fig. 3 for a schematic illustration of the method. Mostly, the template is fitted to the segmented binary volumes (Kaus et al., 2003; Shang and Dossel, 2004; Shen et al., 2001; Zhao and Teoh, 2005), but Fleute et al. (2002) and Heitz et al. (2005) also use the template to segment the original volume while determining correspondences. This has the advantage that training images do not have to be segmented manually in advance, while it is obviously limited to structures that can be segmented reliably using templates, like e.g. certain bone structures. The deformable surface model is usually extracted from an arbitrary training sample, which can lead to biased results. To minimize this bias, Zhao and Teoh (2005) test beforehand which sample is best-suited as a template and use additional intermediate meshes to segment the data sets for which the standard template did not work adequately. The key to a successful landmark generation using mesh-to-volume registration is the robustness of the deformable template algorithm. Techniques used to ensure the necessary robustness include multi-resolution approaches (Fleute et al., 2002), gradient vector flow (Shang and Dossel, 2004) and regularizing internal energies (Shen et al., 2001; Kaus et al., 2003; Zhao and Teoh, 2005). Recently, Dam et al. (2008) presented a bootstrapping approach, in which an SSM is used to segment training data and iteratively refine correspondences. For all mesh-to-volume approaches a homeomorphic mapping between the input shapes is guaranteed if the template does not fold itself in the adaptation process.

4.3. Volume-to-volume registration

Instead of adapting a template mesh to the training data, it is also possible to register a volumetric atlas. Using the resulting deformation field, landmarks placed on the atlas (e.g. by decimating a dense surface mesh) can be propagated to the training data to define the correspondences for the SSM. A visualization of the basic scheme is presented in Fig. 4. Frangi et al. describe how to build an atlas of binary volumes using quasi-affine registration with the normalized mutual information (NMI) metric in (Frangi et al., 2001). Subsequently, a quasi-affine registration followed by a multi-resolution B-Splines deformation is used to warp the atlas to the binary training volumes. In (Frangi et al., 2002), the authors change the underlying NMI to label consistency and Kappa metric to apply the method for multiple objects. In a later publication (Ordás et al., 2004), this version of the algorithm has been tested on 450 input shapes from cardiac MRI with good results – although the volumetric registration does not guarantee a homeomorphic mapping, no triangle flipping was observed.

As with the mesh-to-volume registration, there has also been work on registering a volumetric atlas to the original gray-value images, i.e without the need for manual segmentations. Rueckert et al. (2003) use multi-resolution B-Splines and the NMI metric for that purpose, although they do not propagate landmarks to the training data but perform the statistics on the deformation field (the control points of the B-Splines transform) directly. To remove the bias inherent in the atlas construction scheme as per (Frangi et al., 2001), the authors also present a “natural coordinate system” that minimizes deformations from the atlas to the training images. Similar work to establish correspondences is presented by Park et al. (2003) using Thin-Plate-Splines and the mutual information metric and Heitz et al. (2004) using B-Splines and NMI. The latter two simply use a single representative shape as atlas – obviously with a larger bias than the method by Rueckert et al. (2003). Again, it needs to be emphasized that results of methods working on the original gray volumes strongly depend on the type of objects to be modeled and on the quality of input images. Generally, the more similar the different training samples are and the stronger

![Fig. 3. Basic scheme for mesh-to-volume registration to determine point correspondences: A landmarked deformable surface is adapted to all training images, either to the segmented binary versions or to the original gray-level data. The final positions of landmarks in the images then determine the correspondences.](image1)

![Fig. 4. Basic scheme for volume-to-volume registration to determine point correspondences: A volumetric atlas is matched to all training images, either to the segmented binary versions or to the original gray-level data. The resulting deformation field is used to transform a landmarked template mesh (extracted from the atlas) to all samples, propagating landmarks and defining correspondences.](image2)
the objects of interest stand out from the background, the more accurate the resulting statistical models will be.

4.4. Parameterization-to-parameterization registration

Parameterization is a bijective mapping between a mesh and an appropriate base domain. In 2D, the common base domain for closed contours is the circle, and registering two contours by their parameterization is equivalent to determining correspondences by the relative arc-length. In 3D, parameterization of surfaces is far more complex and usually dependant on the topology of the shape. Most of the methods described below are limited to orientable closed 2D-manifolds of genus zero (i.e. objects without holes and self-intersections), whose common base domain is the sphere. For an introduction to mesh parameterization, refer to Floater and Hormann (2002, 2005). Fig. 5 shows the general outline of the approach for building shape models.

In order to determine correspondences for an SSM, Kelemen et al. (1999) propose to generate a spherical harmonics (SPHARM) mapping for each training shape. After aligning all shapes by their first order ellipsoid, the surface points that map to the same position on the sphere are considered to be corresponding. Brett and Taylor (2000) use a similar method for 3D topological disks. In their approach, all shapes are mapped to 2D disks and aligned by optimizing the disk rotation to minimize distances between corresponding points as found by a preceding ICP. While these methods guarantee a diffeomorphism between all shapes, the obtained correspondences are mostly arbitrary and the quality of the resulting SSM will depend strongly on the input shapes.

For this reason, other authors have tried to control the parameterizations through a limited set of known or at least assumed correspondences: In order to construct a statistical model of the brain, Thompson et al. (1996) manually outline all sulci and transform each resulting parallel cross-section of the surface to a separate parametric mesh. In a similar way, Lamecker et al. (2002) let the user specify corresponding patches across different liver shapes, which are again parameterized independently from each other. Praun et al. (2001) use harmonic maps to interpolate between a set of corresponding points, either detected by automatic feature comparison or created manually. While these methods explicitly build the parameterizations to conform to the given features, it is also possible to start with standard parameterizations and to subsequently modify them for a better match – a technique also known as re-parameterization. For this purpose, Thompson et al. (1997) propose a warping mechanism for spherical maps that matches user-drawn feature lines. Meier and Fisher (2002) employ a similarity criterion based on Euclidean point-to-point distances, normal vectors and curvature to guide their warps. Last but not least, Wang et al. (2005) calculate curvature features over all training surfaces and optimize a mutual information metric on the corresponding parameterizations. All shapes are automatically divided into the required number of patches, which makes the method independent of the topological genus.

4.5. Population-based optimization

The first method of determining a population-based (or group-wise) correspondence between shapes was published by Kotcheff and Taylor (1998). Since there was (and still is) no general definition of "good correspondence" between shapes that could unconditionally be used for pair-wise matching, they focused on the desired properties of the resulting statistical model instead, which they wanted to be as compact as possible, i.e. featuring low eigenvalues concentrated into few modes. To quantify this property, they developed an objective function $F$ based on the determinant of the covariance matrix $S$ (from Eq. (3)):

$$F = \ln |S + \delta I| - n \ln \delta,$$

where $I$ is the identity matrix, $n$ the number of training samples and $\delta$ the amount of additional Gaussian noise. The latter is necessary to prevent problems with zero eigenvalues of $S$: Since $|S| = 1 + \lambda_r$, a single zero leads to an undefined value of $F$. In their work, Kotcheff and Taylor use a genetic algorithm to modify landmark positions and to minimize $F$. Although their approach to landmark modification only works in 2D, the core of the approach (i.e. optimization of the objective function) is independent of the number of dimensions. While the method produced good results (in some cases better than manually landmarked models), there were a couple of problems: The genetic optimization scheme did not converge in all cases, the landmark locations had to be chosen from a limited set of points, and there was no sound theoretical foundation of the objective function.

To remedy the latter issue, Davies et al. proposed a new objective function based on the minimum description length (MDL) of the statistical model (Davies et al., 2002). MDL is a principle from information theory which states that the best model should describe the (training) data as efficiently as possible. It runs in line with the principle of Occam’s Razor that the simplest solution should be preferred over more complex ones. The full objective function for this approach is a rather complex formula and best described in (Davies, 2002). Initially, Davies et al. used a piece-wise linear, recursive parameterization scheme to place landmarks and a genetic algorithm to optimize correspondences (Davies et al., 2002). Although it was used for 2D only, there was already an outline of how this scheme could be ported to 3D. Later, they switched to a continuous re-parameterization with Cauchy kernels and an optimization using the Nelder-Mead algorithm (Davies et al., 2003). The presented results were excellent, and an extension to the case of 3D objects was finally presented in (Davies et al., 2002). The main contribution here was a re-parameterization scheme for closed surfaces of spherical topology (genus-zero), which was implemented using Theta transformations with wrapped Cauchy kernels. Tests on small-sized training sets of brain ventricles and rat kidneys delivered good results and generated further interest in the method.

While the MDL cost function has a sound theoretical foundation, the actual calculation is not trivial and computationally expensive. In (Thodberg, 2003), Thodberg proposes a simplified version of the cost function:

$$F = \sum_m S_m \quad \text{with} \quad S_m = \begin{cases} 1 + \log(\lambda_m/\lambda_{cut}) & \text{for } \lambda_m \geq \lambda_{cut} \\ \lambda_m/\lambda_{cut} & \text{for } \lambda_m < \lambda_{cut} \end{cases} \quad (7)$$

Fig. 5. General outline of the parameterization approach to determine point correspondences: After extracting surfaces from all training images, a one-to-one mapping to a common base domain is created for each surface. By defining landmarks in the base domain, the corresponding points can directly be transferred to all training surfaces using the inverse of the mapping function.
Although strictly speaking, this version is no MDL and indeed similar to Kotcheff’s original formulation (see Eq. (6)), its performance is comparable to the full MDL and it has gained considerable popularity – probably also due to the included open source Matlab code for the 2D case. In another paper, Thodberg also suggests to add curvature costs to the objective function (Thodberg and Olafsdottir, 2003), but the method is evaluated only on one dataset of 2D face contours.

Whereas the objective function of Eq. (7) is simple to compute, the optimization scheme using the Nelder–Mead algorithm is slow and can take hours or days to converge, depending on the number of used landmarks and dimension. While in theory, model-building is only performed once, in practice, tests, optimization and evaluation require a whole number of runs. In (Ericsson and Aström, 2003), Ericsson and Aström derive the gradients of the simplified MDL and present a gradient descent optimization of 2D shapes. As a comparison with the algorithm from Thodberg (2003) reveals, this leads not only to faster convergence but also to better final results. Heimann et al. (2005) use this work to develop a gradient descent optimization for 3D data, employing locally constrained Gaussian kernels (instead of wrapped Cauchy kernels) for reparameterization. As in the 2D case, this version is both faster and delivers better results than the Nelder–Mead optimization from Davies et al. (2002). An open source version of the algorithm was later presented in (Heimann et al., 2006).

Whereas re-parameterization on the sphere currently is the most popular method for group-wise correspondence, there are also some alternative approaches: Horkaew and Yang (2003) parameterize genus-zero shapes to a disc and use piecewise linear maps to optimize the MDL criterion. In an evaluation with 38 human left ventricles and a B-Spline shape representation, the method works well, but no comparison with Davies et al. (2002) is conducted. In (Cates et al., 2007), Cates et al. present an approach that optimizes correspondences based on the entropy of shapes (generally equivalent to MDL). They use a non-parametric particle system to place landmarks, which makes the method independent from the topology of the surfaces. Results in 2D surpass the simplified MDL optimization from Thodberg (2003), the 3D results look very promising but have not been compared to other methods yet. Last but not least, there is work towards a group-wise registration of images by Twinning et al. (2005), Cootes et al. and the method is evaluated only on one dataset of 2D face contours.

General outline of an optimization scheme to determine point correspondences: As in the parameterization approach, mappings to a common base domain are created for all training meshes, and landmarks are defined in the base domain. In an iterative optimization the mapping functions are modified, resulting in changing correspondences. This process is continued until a cost function describing the quality of the model has converged to the best value.

4.6. Evaluation of correspondence methods

Given the wide variety of correspondence methods, an objective comparison between them should be capable of differentiating between suitable and unsuitable approaches. However, this evaluation is hampered by the fact that the true correspondences of biological shapes are generally not known. Davies (2002) therefore introduced three measures to evaluate the general quality of an SSM: Generalization ability, specificity and compactness. Generalization ability quantifies the capability of the model to represent new shapes. It can be estimated by performing a series of leave-one-out tests on the training set, measuring the distance of the omitted training shape to the closest match of the reduced model. Specificity describes the validity of the shapes produced by the model. The value is estimated by generating random parameter values from a normal distribution with zero mean and the respective standard deviation from the PCA. The distance of the generated shape to the closest match of the training set is averaged over a large number of runs. Compactness simply measures the accumulative variance of the model. All measures are commonly defined as functions of the number of modes or parameters used by the model and displayed as piecwise linear graphs. Smaller values indicate better models.

Stynert et al. (2003) employ these three measures to compare models built by manually initialized subdivision surfaces, SPHARM parameterization, and optimization of the determinant of the covariance matrix (Eq. (6)) and of the MDL. In this comparison, using lateral brain ventricle and femoral head data, the two optimization approaches generate the best results. Another comparison between different correspondence methods is conducted by Assen (2006): A heart ventricle model is built using a slice-wise ray-shooting technique and volume-to-volume registration (from Frangi et al., 2002), the latter with two different landmark distributions on the atlas. In addition to the three theoretical measures, the segmentation accuracy of the resulting model on test data is also evaluated. The model built with one of the registration methods reaches the best results for the segmentation task (with only slightly better accuracy than the other two). Remarkably, the synthetic measures for this method result in worse values than for any of the other two.

In another set of experiments, Ericsson and Karlsson (2006) also encounter problems with the three measures: In a test with synthetic shapes, models with optimal correspondences score worse results for specificity than those with slightly altered correspondences. The authors therefore propose to use a ground truth correspondence measure, which describes the differences to manual reference landmarks (or the optimal landmarks for synthetic shapes). However, all conducted experiments are performed with 2D shapes and the method seems difficult to transfer to 3D because of excessive manual landmarking. Heimann et al. (2006) also analyze the generalization and specificity measures and stress their dependency from the underlying landmark distribution, which can lead to better scores for worse models. By changing the internal shape difference metric from landmark distance to volumetric overlap, they remedy this issue and generalize the measures to be able to compare models with different numbers of landmarks and different distributions.

A first evaluation of different correspondence methods for shape analysis is presented in a recent publication by Stynert et al. (2007): Using volumetric registration, SPHARM and MDL optimization, they generate hippocampus models for schizophrenia patients and for a control group. An analysis of the resulting group differences reveals that the utilized correspondence method noticeably influences the results, with the volumetric registration producing the least significant group differences.
5. Appearance models

The majority of published works uses shape models for image segmentation, i.e. after construction the model is fitted to new, previously unseen data. For this purpose, a model of the appearance of the structure of interest is required. Although the first version of shape models simply adapted to the strongest edges in the image (Cootes et al., 1995), the state of the art quickly developed towards specialized, statistical models of appearance. As the shape model, these appearance models have to be trained from sample data. Below, we will present a number of different appearance models based on boundary and region features. In addition, we will highlight some clustering methods applicable to make the appearance models more robust.

5.1. Boundary-based features

The first statistical appearance models are described by Cootes and Taylor (1993, 1994): Their basic idea is to sample profiles perpendicular to the surface in all training images and – similar to the shape model construction – extract the mean profile and principal modes of variation for each landmark. During search, the fit of a profile model at a certain position is evaluated by measuring the Mahalanobis distance between the sample profile and the model – in case the complete covariance matrix is not stored, the Mahalanobis distance can be approximated by the largest eigenmodes. Typically, the used profiles contain the plain pixel intensity values, the corresponding derivatives, or the normalized versions thereof. An evaluation of different profile types for face recognition showed that normalized derivative profiles deliver the best results (Cootes and Taylor, 1993). During another, more extensive evaluation for bone detection in radiographs, the normalized intensity profile generated the most accurate segmentations (Behiels et al., 2002). While the latter study also reports some experiments with different cost functions (including the correlation coefficient), the Mahalanobis distance generated the best results overall.

It is also possible to combine several different profile types into one large feature vector, as e.g. shown by Brejl and Sonka (2000): Intensity and gradient profiles are joined with the average inside and outside gray value, and the weighting between different features is handled by the covariance matrix. Another innovative aspect of that study is that intensity profiles are constructed by averaging a number of pixels in the neighborhood, increasing the robustness of the method.

One feature that has been used successfully in face recognition with 2D SSMs is that of Gabor wavelets (Daugman, 1988): This method measures the local image response to a set of Gabor filters, resulting in a feature vector (also called jet) containing amplitude and phase for each individual filter. McKenna et al. (1997) were the first to use this Gabor jet in conjunction with a PDM for a video tracking application. The training, however, was limited to the first video frame, where landmarks had to be placed manually. Jiao et al. (2003) present a real training method for a set of images, employing the Gabor jet as a Gaussian mixture model that is created using the EM algorithm (Dempster et al., 1977). Like McKenna et al., they make use of the fact that the optimal displacement for a landmark can be estimated from the phase information in the corresponding jet. Shen et al. (2003) utilize rotation-invariant Gabor filters as smoothing edge detectors for the segmentation of 2D ultrasound images. Since wavelets can also be used in 3D volumes (Muraki, 1993), this method seems promising for the analysis of medical data as well. Another type of local descriptor – steerable filters (Freeman and Adelson, 1991) – is utilized by Langs et al. (2006): Like wavelets, this method too can be applied to 3D and might become of interest in medical image segmentation.

Relatively early-on in the development of SSMs and appearance models, Haslam et al. (1994) presented a Bayesian approach to modeling the boundary appearance, concentrating all landmark profiles from a training image into a single feature vector. Two appearance models are built using the classical method of PCA on the covariance matrix of these feature vectors: The first is a boundary model of the true landmark positions, and the second is a background model created by varying the pose parameters of the training shapes, thereby moving the landmarks to incorrect positions. A downside to this method of modeling the appearance is that due to the concatenation of profiles, it is not possible to independently evaluate the goodness of fit for each landmark, but for the entire shape only. Therefore, the method requires a special form of search algorithms (see also Section 6) and has not become very popular.

One problem common to all appearance models based on the covariance matrix is that training data (i.e. the profiles) should ideally have a Gaussian distribution. As noted by de Bruijne et al. (2002), this is not the case with many medical imaging tasks. As an alternative, a non-linear method to model the object boundary using a kNN-classifier has been proposed: During training, profiles are sampled at the true boundary positions and at a number of additional positions shifted towards the inside and the outside of the object. During search, the amount of true profiles among the k nearest neighbors of a sample profile serves as an estimate for the goodness of fit and assumes the role of the Mahalanobis distance. In general, any classifier that calculates a probability value for class membership can be used for that purpose, and not just kNN. Boosting methods, which combine several weak classifiers into a single strong one, are becoming more and more popular, especially the AdaBoost algorithm presented by Freund and Schapire (1997). This algorithm is also used by Li et al. (2004) who have trained it on more than 6,000 features calculated by steerable filters. During model search, 200 features per landmark are used at most in order to estimate the probability of points along the normal vector forming part of the boundary. Other approaches employing AdaBoost are presented by Li and Ito (2005) who work with a histogram classifier trained on quantized intensity profiles and by Zhang et al. (2005) who use Haar wavelets.

Instead of a complete training from sample images, some authors prefer to explicitly specify the boundary appearance: Bailleul et al. (2004) calculate four features along each profile, which are assumed to accurately describe the border (of the putamen in MRI datasets): maximum intensity difference between inside and outside voxels, intensity means difference, inner voxels regularity and inner-outer voxels regularity difference. All features are normalized to [0..1] and added to give a final estimate of the boundary fit. In the work of Lamecker et al. (2004), all profile intensities on the inside of the modeled shape (a liver in CT datasets) are required to lie within a fixed interval. A preceding diffusion filtering ensures that all intensities are as homogeneous as possible. Similarly, Peters et al. (2005) also require certain features (e.g. central gray value, gradient magnitude) to be part of a specific interval. The innovative part of their work is that the interval borders are optimized for each appearance model during a simulated search which estimates the performance of different models.

5.2. Region-based features

The most popular method using region-based features is the use of Active Appearance models (see Section 6.3 for details). Here, the entire inner region of the shape model is employed to build a feature vector \( \mathbf{g} \). In the simplest case by storing the plain intensity values of all pixels of the region, i.e. the image texture. To build \( \mathbf{g} \) for different shapes, the respective region has to be transformed into a standard shape first – usually the mean shape of the SSM. The
appearance model is then created by performing a PCA on the covariance matrix of textures. As with the profile models, normalization of the intensities is generally useful: This can be accomplished by adding an offset and by scaling to transform all gray-values to zero-mean and unit-variance (Cootes and Taylor, 2001), or by a non-linear procedure as described by Bosch et al. (2002). The latter method proved especially useful in images with a strongly non-Gaussian distribution as encountered, e.g. in ultrasound images. As an alternative to working with intensities, Cootes and Taylor (2001) propose storing gradient direction and strength, the latter mapped non-linearly. In an evaluation on face images, this method proved to be better than the normalized intensity. Scott et al. (2003) use gradient orientation, corner and edge strength for the detection of vertebrae in dual energy X-ray images. For segmentation of colored face images, Stegmann and Larsen (2003) employ HSV color space and model appearance as value (i.e. intensity), hue and edge (gradient magnitude).

Obviously, memory requirements have to be taken into account when using data from the entire inner region of the model: Even in 2D, the involved matrices can get fairly large, and for 3D volumes or time series the problem is much more acute. Nonetheless, complete region-based appearance models can be built for these cases by reducing the resolution of the employed texture, as several studies demonstrate (Beichel et al., 2002; Bosch et al., 2002; Mitchell et al., 2002). Another option is to use only certain parts of the region, e.g. rectangular regions around specific feature points, as proposed by Cristinacce and Cootes (2008). In this scheme, all local regions are independently intensity-normalized and then concatenated into one feature vector. Still, the most efficient way of using region information is to extract specific features from the region of interest. As an extreme example, Freedman et al. (2005) propose to model appearance by the gray-value distribution (i.e. a histogram) of all voxels inside the examined object. To be able to better discriminate between the modeled object and the background, a widely used enhancement is to model both inner and outer regions of the SSM. Below, we will present several works in this direction, which can all be used in the classical Active Shape model framework (Section 6.2).

Broadhurst et al. (2005) create histograms for inside and outside regions and map these to Euclidean space using the Earth mover’s distance, where they can apply PCA to the data. They also evaluate different methods of how to best define the used regions (e.g. only use pixels within a certain distance from the boundary) and compare their method to a boundary-based profile approach. In a subsequent publication (Broadhurst et al., 2006), the authors enhance their approach by weighting the influence of each pixel and introducing a-priori information about the expected variance. van Ginneken et al. (2002) model the inside and outside appearance around each landmark using a number of derivatives at different scales (also called the local jet (Florack et al., 1996)). An optimal feature selection picks the best set of features for each landmark. During search, these best sets are employed to estimate the inside probability for a number of pixels on a profile using a kNN-classifier. The method has been tested on X-ray and MRI images and delivered better results than the Mahalanobis distance on gray value profiles. Later, the approach was modified by Ordás et al. (2003) to use Cartesian Differential Invariants instead of the local jet, making it invariant to orientation. Zhan and Shen (2006) employ Gabor filter banks to extract features from 3D ultrasound images. The features extracted by two Gabor filters in orthogonal planes are forwarded to a kernel support vector machine that estimates the likelihood of a voxel belonging to prostate tissue or not.

Recently, Larsen et al. (2007) presented the first approach to employing feature extraction to the Active Appearance model framework (Section 6.3): They represent object texture by wavelets and wedgelets, thus not only reducing the dimensionality of texture vectors, but also reducing noise. Due to the inherent frequency separation, emphasis on edges can be incorporated without additional computational costs.

5.3. Clustering techniques

A common approach to building appearance models is to train one model for every landmark of the SSM. However, almost all of the local methods described above can benefit from clustering: In this case, several landmarks with similar boundary appearance are combined to train a shared appearance model. The obvious advantage is that clustered models are based on more training samples and thus offer improved generalization ability.

Brejl and Sonka (2000) describe the first use of clustering for appearance models, employing the c-means algorithm to generate a user-specified number of n clusters from all landmarks in all training images. During search, they compare the sampled features with all n models and select the boundary fitting costs of the most similar one. To enable utilization of different models for different parts of the boundary, they suggest a manual definition of the respective border segments and storage of those models that are allowed for each one. Only the allowed models are then compared with the sampled features. Kaus et al. present a more practical solution in (Kaus et al., 2004): After clustering all features into n classes with the k-means algorithm, each landmark is assigned to the cluster containing the largest number of training features from that point. During search, the features sampled at a certain landmark only need to be compared with the assigned cluster. Different appearance models for different boundary segments are thus supported automatically. Stough et al. (2004) reach the same result with a different method: After the clustering process, they assign each landmark to the one cluster that returns the best accumulated fitting costs for all training features at that point. In their work, these fitting costs are calculated by the normalized correlation between the sample feature and a prototype generated for each cluster.

While the presented approaches cluster boundary appearance with respect to the feature values, Ho and Giger (2004) suggest a spatial clustering. Working with profile models, they define a profile scale-space and combine profiles hierarchically based on their neighborhood relation. In this case, the neighborhood relation is defined on a sphere and thus requires objects of spherical topology to work. The results are profiles that are blurred along the boundary, like in case of directional smoothing.

6. Search algorithms

Due to the large size of the search space in 3D, most methods applied to locate an SSM in new image data use local search algorithms that require an initial estimate of the model pose. In the first subsection, we will review several approaches for this initialization, including some global search algorithms that deliver a complete solution for shape and pose. Subsequently, we will deal with the popular Active Shape models and Active Appearance models with some of their variants. In a next step, we will discuss various alternative search algorithms and finally present methods for the refinement of detected shapes and the incorporation of user interaction.

6.1. Initialization

The easiest solution to initializing an SSM is user interaction, as implemented, e.g. in (Kelemen et al., 1999; Lamecker et al., 2004; Pizer et al., 2003; Rao et al., 2005; Weese et al., 2001). In general,
it is sufficient to roughly align the position and rotation of the mean shape to the data, which can be accomplished in negligible time (less than half a minute). For cases requiring a more specific initialization, Hug et al. (2000) propose to manually specify a small mean shape to the data, which can be accomplished in negligible time. It is sufficient to roughly align the position and rotation of the mean shape to the data, which can be accomplished in negligible time. However, the necessary accumulator array seems unsuitable for 3D since the number of array dimensions increases from four to seven if using the proposed similarity transformation. Ecabert et al. (2008) thus employ a more constrained transformation (restricted to translation and isotropic scaling) when using the Generalized Hough transform for initialization of a 3D model.

Last but not least, there is the possibility of conducting a global search on the entire image. Due to the large search space and numerous local minima, this task is predisposed for the use of evolutionary algorithms (EA). Already in the 1970s, Holland (1975) developed one of the most popular variants of EA, the genetic algorithms: A large population of solutions (each one encoded as a binary chromosome) is initially spread across the entire search space and evaluated simultaneously. According to these results, the best set of chromosomes is selected for reproduction to produce the next generation of solutions. Mimicking biological evolution, there are mutation (random bit changes) and cross-over operators which influence the process. After several iterations/generations, the best-rated chromosome in the population is regarded as the final solution. Hill and Taylor (1992) were the first to employ genetic algorithms for the model-based segmentation of echocardiogram images, optimizing four pose and six shape parameters (all encoded in one chromosome) of a custom-built model. They also presented an extension called niches and species to the algorithm in order to detect several candidate shapes at once. Subsequently (Hill et al., 1996), this approach was used in the automatic detection of PDMS with one extension: The mutation part was not constrained (restricted to translation and isotropic scaling) when using the Generalized Hough transform for initialization of a 3D model.

6.2. Active Shape model search

Since their introduction by Cootes et al. (1992), Cootes et al., 1995, Active Shape models (ASMs) have become one of the most popular model-based segmentation techniques in medical imaging. ASM is a local search algorithm based on a point distribution model; an instance of the model $y$ in an image is defined by a similarity transformation $T$ and shape parameters $b$:

$$ y = T(x + \phi b) $$

where $\phi = (\phi_1 \cdots \phi_k)$ is the matrix of eigenvectors. From an initial state, adjustments are calculated individually for each landmark by evaluating the fit of the appearance model at different positions along the normal vector to the surface. This leads to a vector of optimal displacements $dy$. In a first step, the pose $T$ of the model is adjusted by a Procrustes match of the model to $y + dy$, leading to a new transformation $T$ and new residual displacements $dy$. In a second step, $dy$ is transformed into model space and then projected into parameter space to give the optimal parameter updates:

$$ db = p^T T^{-1}(dy) $$

where $T^{-1}$ is equal to $T$ but without the translational part. After updating $b$ and applying appropriate parameter limits (see Section 3.2), we have an updated valid instance of the model. These two steps are conducted iteratively, until a specified convergence criterion is hit, e.g. the maximum or average landmark movement is below a given threshold. The most detailed description of the algorithm can be found in a technical report by Cootes and Taylor (2004). Fig. 7 shows a simple 2D example of the update procedure.

One of the first extensions to the basic ASM algorithm was the introduction of a coarse-to-fine search strategy in (Cootes et al., 1994): For this variant, the image data is organized as a multi-resolution pyramid and the SSM has separate appearance models for each level. The search starts at the coarsest level and switches to the next resolution when a predefined criterion is met, e.g. when the optimal displacement positions are located in the central part of the search profile for 95% of all landmarks. This procedure leads to a considerable increase in speed, robustness and quality of fit; it is commonly employed in all recent publications using ASMs. Another extension by Hill et al. (1996) introduces correctional constraints for shape approximation: In both Procrustes matching and shape parameter calculation, the error term based on $dy$ is split into two weighted parts, one normal to the surface and one tangential. By weighting the tangential part lower than the normal, landmarks can slide along the surface more easily. In experiments with 2D images this leads to faster convergence and better accuracy, for 3D images the algorithm is described but not evaluated.

Fig. 7. One iteration of an ASM search: At the beginning, the model (contour with landmarks) is located at the lower left of the true position (solid gray object). Local appearance models for all landmarks are evaluated at different positions perpendicular to the model surface. The best positions are shown as small circles in the center image. Finally, model parameters are updated to minimize the squared distances to the found best positions, bringing the model closer to the correct solution.
The largest part of proposed modifications to the original ASM algorithm tries to improve the stability against outliers: The default least-squares approach is only optimal for a Gaussian distribution of residuals (Barnett and Lewis, 1994), and outliers in the appearance model match can disturb the search considerably, especially if they occur clustered. Dutta and Sonka (1998) propose an appearance model match can disturb the search considerably, and outliers in the distribution of residuals (Barnett and Lewis, 1994), and outliers in the least-squares approach is only optimal for a Gaussian distribution of residuals (Barnett and Lewis, 1994). For the second run, weights are set according to the remaining residuals after the first optimization. If that failed to decrease the residual for a specific landmark below a given threshold, the relevant landmark is weighted down or not considered at all. Zhao et al. (2004) combine values from the fit of the current appearance model, the stability of point positions during the last iteration and the stability of appearance model into one weight matrix. In an evaluation on face images, this procedure leads to more accurate results than those achieved without weighting.

Yet another approach to treat outliers is proposed by Nahed et al. (2006): Firstly, a number of candidate feature points are extracted from the image around the current model position. In a next step, a robust point matching algorithm which rejects outliers (Gold et al., 1998) is used to find the best-fitting model. A method which is not applicable to 3D but which is very efficient in 2D is to employ a minimum cost search for the boundary as proposed by Behiels et al. (2002). In this approach, the difference of neighboring candidate positions (i.e. the shift perpendicular to the surface) is penalized and weighted with the fit of the appearance model, leading to an optimal solution with respect to the defined cost function.

6.3. Active Appearance model search

Active Appearance models (AAMs), introduced by Cootes et al. (1998), belong to the class of generative models, i.e. they can synthesize realistic images of the modeled data. This is made possible by storing the complete appearance (or texture) of the object, including its variations, in addition to the expected shape. However, AAMs are more than shape models with region-based features; they also feature a proprietary search method which differs completely from ASMs. Shape and appearance variation are combined into one linear system, allowing for shape \( \mathbf{x} \) and appearance \( \mathbf{g} \) to be described by a common parameter vector \( \mathbf{c} \):

\[
\mathbf{x} = \mathbf{x} + \Phi_s \mathbf{W}_s Q_s \mathbf{c} \\
\mathbf{g} = \mathbf{g} + \Phi_g Q_g \mathbf{c}
\]

(10)

In this formulation, \( \Phi_s \) and \( \Phi_g \) are the independent eigenvector matrices of shape and gray-value model, respectively, and \( \mathbf{W}_s \) is a diagonal weight matrix for the shape parameters. \( Q_s \) is the eigenvector matrix of the combined shape and gray-value parameters and obtained by a PCA on the independent parameters \( \mathbf{b} = \left( \begin{array}{c} \mathbf{b}_s \\ \mathbf{b}_g \end{array} \right) \). An instance of an AAM in an image is thus defined by a similarity transformation \( \mathbf{T} \) and the combined shape-appearance parameters \( \mathbf{c} \) – below, we will denote both together as parameters \( \mathbf{p} \). To evaluate the goodness of fit, the image texture is warped to the mean shape and normalized, resulting in \( \mathbf{g}_s \). Using the modeled gray-value appearance \( \mathbf{g}_s = \mathbf{g} \) from Eq. (10), the residuals are then given by \( \mathbf{r}(\mathbf{p}) = \mathbf{g}_s - \mathbf{g}_s \) and the error by \( E = \mathbf{r}^T \mathbf{r} \).

The key idea of AAMs search is to assume a constant relationship between texture residuals \( \mathbf{r}(\mathbf{p}) \) and parameter updates \( \mathbf{dp} \) over the entire search:

\[
\mathbf{dp} = -\mathbf{R}(\mathbf{r}(\mathbf{p}))
\]

(11)

The success of this optimization scheme largely depends on the derivative matrix \( \mathbf{R} \). In the first presentation of AAMs (Cootes et al., 1998), \( \mathbf{R} \) was computed using multivariate linear regression on a large number of simulated disturbances of the training images. Later, regression was replaced by numeric differentiation, claimed to be both faster and more reliable (Cootes et al., 2001). Donner et al. (2006) point out a more practical advantage of the latter approach, i.e. it does not require the loading of all training images into memory at once for training.

AAMs quickly gained popularity in the image analysis community, and several modifications to the original scheme were proposed. A review of the early variants is given by Cootes and Kittipanya-ngam (2002), who compare Shape AAMs (Cootes et al., 1998) (which only update pose and shape by residuals, while texture is updated directly), compositional updating (Baker and Matthews, 2001) (using thin-plate splines for the pose update) and Direct AAMs (Hou et al., 2001) (which predict shape by texture). More recent modifications include the work of Matthews and Baker (2004), who suggest an inverse compositional warp update instead of the standard additive update and present an analytical derivation for a gradient descent search. Originally presented only for 2D AAMs, the algorithm was later extended to 3D by Andreopoulos and Tsotsos (2008). Beichel et al. (2005) partition the residuals \( \mathbf{r}(\mathbf{p}) \) into different modes using the mean-shift algorithm. During search, they match the image to all different combinations of modes and calculate parameter updates for every single one. For each iteration they accept the combination leading to the best result of the objective function, thus rejecting up to 50% of outliers. Donner et al. (2006) employ canonical correlation analysis to calculate the derivative matrix \( \mathbf{R} \). This method provides more accurate estimates for parameter updates than the standard numeric differentiation, resulting in a search that is up to four times faster. In addition, it requires fewer examples in the training phase, thus also speeding up the model construction considerably.

Although AAMs are mainly described for 2D (a prominent application is face recognition), the method itself is by no means limited in dimension. Mitchell et al. (2002) presented the first true 3D model, using volumetric textures to represent appearance. More 3D medical applications are described in Section 7. A good starting point to experiment with AAMs is the open source software FAME by Stegmann et al. (2003) – although it currently only supports 2D models. As with ASMs, the most detailed description of the basic algorithm is given in the technical report by Cootes and Taylor (2004). Fig. 8 shows a simple 2D example of the update procedure.
6.4. Other algorithms with strict shape constraints

Though ASM and AAM are the most frequently employed constrained search algorithms, there exist several other possibilities to fit an SSM to new data. After initialization of the model, standard optimization techniques can be used to directly optimize the shape parameters. Successfully employed algorithms include the Nelder–Mead simplex approach (used by Heinze et al. (2002) and Tang and Ellis (2005)), the Quasi-Newton method (used by Haslam et al. (1994) and Székely et al. (1996)), the conjugate gradient (CG) algorithm (used by Staib and Duncan (1996) and Löfjöden et al. (2004)) and gradient descent optimization (used by Freedman et al. (2005) and Nain et al. (2007)). In general though, these methods are slower than the approaches presented above: Haslam et al. (1994) report a runtime for the Quasi-Newtonian optimizer that is 50 times longer than ASM. Therefore, the standard optimization techniques are mainly used in cases where ASM is not applicable, either because the SSM uses a different shape representation (no PDM), or because the appearance model does not allow the individual evaluation of fit for each landmark. An efficient use of the Nelder–Mead simplex optimizer is demonstrated by Cristinacce and Cootes (2008) for constrained local models, which reach similar speed as AAM.

For volumetric shape models, Rueckert et al. (2003) present a method called statistical deformation models. Here, the statistical analysis is not conducted by using point coordinates, but rather by using volumetric deformation fields between a mean image (the atlas) and training samples. The results are used to constrain the B-splines registration of the atlas to new images, reducing the search space considerably.

Klim et al. (2003) present a mixture of the described ASM and AAM methods: They use a standard PDM with edge-based appearance models for all landmarks, which determine the individual best fits along the respective normal vectors. But instead of using the ASM procedure to update parameters, they employ a regression matrix as in AAM optimization. The method is only evaluated for 2D (with better performance than ASM on hand X-rays), but could also be implemented for the 3D case.

6.5. Algorithms with relaxed shape constraints

In the preceding sections (Sections 6.1–6.4), we have discussed algorithms which strictly adhere to the underlying SSM. Below, we will describe various methods to relax the hard shape constraints and to allow for additional flexibility during search.

One approach to solving this problem is to run a normal SSM until convergence and to include an additional refinement step afterwards. The rationale behind this is that the model surface should be close enough to the real contour at that stage to be attracted with edge-based features. Shang and Dossel (2004) simply search for the best fit of each individual vertex within a confined local neighborhood, guided by the local appearance models. However, in the absence of shape constraints, some kind of regularization is generally required to deal with noise and image clutter. To this end, Cootes and Taylor (1996) examine the candidate positions suggested by the individual local appearance models and conduct a pruned tree search to determine the subset of points that best matches the current shape model. Li and Reinhardt (2001) employ a slice-wise contour refinement for the segmented volume: In each slice, the 2D contour is adjusted and regularized by a snake, which draws its external energy from the local appearance models. Pekar et al. (2001) utilize a freely deformable triangular mesh that is initialized with the converged SSM and regularized by its internal smoothness constraints.

Another class of methods introduces additional flexibility already from the start of the search: Such approaches are typically based on a deformable mesh that bases its internal energies on the discrepancy to the closest allowed shape model. Weese et al. (2001) use the differences in corresponding edge lengths between mesh and model for this purpose and employ a conjugate gradient algorithm to optimize all point positions. Heimann et al. (2007) use the same internal energy but a different optimization approach: an optimal surface search determines the best-fitting positions (according to the local appearance models), and forces towards these positions are then combined with internal forces. Shen et al. (2001) use the difference between learned attribute vectors (characterizing the geometric structure of the model) and the current mesh as basis for internal energy. They employ local affine transformations, modulated by a Gaussian envelop function, to adjust the deformable model. Due to the ability to weight different parts of the model differently during search, this algorithm is termed Adaptive Focus Deformable Shape model (AFDSM). Tsagaan et al. (2002) model shape variation based on local curvature feature vectors, which are used to calculate covariance matrices during training. The internal energy for the search is then based on the Mahalanobis distances from the features of the current mesh to the stored data. Seghers et al. (2007) also use local shape statistics to obtain a more flexible model. They employ dynamic programming to find a set of optimal 1D paths over the deformable surface. By combining votes for different random paths, they chose the vertices that form the final 3D segmentation. For the m-rep shape representation by Pizer et al. (2003), a special deformation scheme has been developed: In a hierarchical procedure, the entire SSM is first fitted globally, then individual subfigures and subsequently the medial atoms are adjusted (all using a conjugate gradient optimizer). The last step is a refinement of the boundary vertices along the corresponding normal vectors.

6.6. Incorporation of user interaction

A topic which has received relatively little attention until now is that of incorporation of user interaction into SSM search algorithms. In case of a failure of automatic segmentation, the user should be able to correct the result by moving a limited number of incorrectly placed landmarks to their true positions. The most elegant solution to this problem is presented by Cootes and Taylor (2001) as constrained AAM: They define a probabilistic framework for AAMs, in which the error function is enhanced to include priors on parameters and point positions. In this way, the additional data is integrated seamlessly into the segmentation process. Neumann and Lorenz (1998), working with an ASM-like search algorithm, suggest adding a new error term to the least-squares procedure which fits model points to candidate locations in each iteration. The resulting linear system of equations can be solved analytically for 2D shapes, but for 3D an iterative procedure is required. Ginneken et al. (2003) propose running the standard ASM
shape parameter update repeatedly: Every time, landmarks that have been corrected by the user are set to their specified positions again, until the shape model converges towards these positions. An evaluation in the same paper reveals that too many manually corrected landmarks actually result in a deterioration of the performance of the SSM, because the deformation becomes too restricted.

7. Applications

In this section, we will present applications of 3D SSMs for medical image analysis. The focus is on the medical application, i.e. we will only list those papers which state a real-world problem solvable with the approach. Studies where a shape model is built only to demonstrate the feasibility of the modeling approach will not be considered.

7.1. Segmentation

According to our survey, 3D segmentation – i.e. the analysis and labeling of volumetric images – is the major application of statistical shape models. With the arrival of automatic landmarking methods, this field has experienced an enormous boost, and publication numbers continue to rise. SSMs have been used for the segmentation of a variety of structures in the human body. To provide the reader with an overview, we sorted all publications in Table 1.

Table 1

<table>
<thead>
<tr>
<th>Authors</th>
<th>Object(s) of interest</th>
<th>Modality</th>
<th>Correspondences</th>
<th>Appearance model</th>
<th>Search algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhan and Shen (2006)</td>
<td>Prostate</td>
<td>US</td>
<td>Mesh-to-volume</td>
<td>F-Region</td>
<td>Class</td>
</tr>
<tr>
<td>Davatzikos et al. (2002)</td>
<td>Spine</td>
<td>MRI</td>
<td>Mesh-to-volume</td>
<td>G-Profile</td>
<td>Untr</td>
</tr>
<tr>
<td>Heinzle et al. (2002)</td>
<td>Knee joint</td>
<td>CT</td>
<td>Point matching</td>
<td>Profile</td>
<td>Expl</td>
</tr>
<tr>
<td>Frapp et al. (2006)</td>
<td>Knee bones</td>
<td>MRI</td>
<td>Optimization</td>
<td>?</td>
<td>ASM</td>
</tr>
<tr>
<td>Josephson et al. (2005)</td>
<td>Knee (femur)</td>
<td>MRI</td>
<td>Point matching</td>
<td>I-Profile</td>
<td>Mah</td>
</tr>
<tr>
<td>Tang and Ellis (2005)</td>
<td>Femur</td>
<td>Multi X-ray</td>
<td>Point matching</td>
<td>G-Profile</td>
<td>Cor</td>
</tr>
<tr>
<td>Lamer et al. (2004)</td>
<td>Pelvic bone</td>
<td>CT</td>
<td>Man + Param</td>
<td>Profile</td>
<td>Expl</td>
</tr>
<tr>
<td>Heimann et al. (2007)</td>
<td>Liver</td>
<td>CT</td>
<td>Mesh-to-volume</td>
<td>G-Profile</td>
<td>Untr</td>
</tr>
<tr>
<td>Pekar et al. (2004)</td>
<td>Bladder, rectum, fem.h.</td>
<td>CT</td>
<td>Mesh-to-volume</td>
<td>F-Profile</td>
<td>Expl</td>
</tr>
<tr>
<td>Zhou and Shen (2006)</td>
<td>Prostate</td>
<td>CT</td>
<td>Mesh-to-volume</td>
<td>F-Region</td>
<td>Class</td>
</tr>
<tr>
<td>Hodge et al. (2006)</td>
<td>Prostate</td>
<td>US</td>
<td>Optimization</td>
<td>cl.I-Profile</td>
<td>Class</td>
</tr>
<tr>
<td>de Bruin et al. (2003)</td>
<td>Aortic aneurysms</td>
<td>CT</td>
<td>Mesh-to-volume</td>
<td>I-Profile</td>
<td>Expl</td>
</tr>
<tr>
<td>de Bruin et al. (2004)</td>
<td>Aortic aneurysms</td>
<td>CT</td>
<td>Mesh-to-volume</td>
<td>I-Profile</td>
<td>Expl</td>
</tr>
</tbody>
</table>

In Table 1, the authors, object(s) of interest, modality, correspondences, appearance model, and search algorithm are listed. The table will only list those papers which state a real-world problem solvable with the approach. Studies where a shape model is built only to demonstrate the feasibility of the modeling approach will not be considered.

Cardiac structures

Staub and Duncan (1996) | Left ventricle | MRI | Parameterization | G-Profile | Untr |
| Mitchell et al. (2002) | Left ventricle | MRI + CT | Volume-to-volume | NI-Region | Res |
| Lapp et al. (2004) | Ventrices, myocardium | MRI | Mesh-to-volume | Profile | Expl |
| Kaus et al. (2004) | Left ventricle | MRI | Mesh-to-volume | NI-Region | Res |
| van Assen et al. (2003) | Left ventricle | MSCT | Arc-length | I-Profile | Class |
| Fritz et al. (2005) | Left ventricle | MSCT | Parameterization | I-Profile | Mah |
| Stegmann and Pedersen (2005) | Left ventricle | MRI | Arc-length | NI-Region | Res |
| Lötjönen et al. (2004) | Heart | MRI | Mesh-to-volume | I-Region | Res |
| Shang and Dossel (2004) | Heart | MRI | Mesh-to-volume | G-Profile | Untr |
| Fritz et al. (2006) | Cardiac chambers | MSCT | Mesh-to-volume | Profile | Expl |
| Lorenz and von Berg (2006) | Heart (incl. vessels) | MSCT | Mesh-to-volume | Profile | Expl |
| van Assen et al. (2006) | Left ventricle | Sparse MRI | Volume-to-volume | I-Profile | Class |
| Zambal et al. (2006) | Left ventricle | MRI | Parameterization | NI-Region | Res |
| Hansegård et al. (2007) | Left ventricle | Triplane US | Arc-length | NI-Region | Res |
| Andreopoulos and Tsotsos (2008) | Left ventricle | MRI | Arc-length | NI-Region | Res |
| Nahed et al. (2006) | Right ventricle | MRI | Optimization | G-Profile | Untr |
| Tölli et al. (2006) | Heart | MRI | Volume-to-volume | var. Profiles | var. PF & Global |

Soft tissue structures in the abdominal and pelvic area

Dickens et al. (2002) | Kidney | MRI | Mesh-to-volume | NI-Profile | Class |
| Tsaagaa et al. (2002) | Kidney | CT | Manual | G-Profile | Untr |
| Pizer et al. (2003) | Kidney | CT | Mesh-to-volume | NI-Profile | Cor |
| Stough et al. (2004), Rao et al., 2005 | Kidney | CT | Mesh-to-volume | cl.I-Profile | Cor |
| Beischel et al. (2002) | Diaphragm dome | CT | Mesh-to-volume | NI-Region | Res |
| Lamer et al. (2003, 2004) | Liver | CT | Man + Param | I-Profile | Expl |
| Lim and Ho (2006) | Liver | CT | Mesh-to-volume | F-Profile | Mah |
| Heimann et al. (2007) | Liver | CT | Optimization | cl.I-Profile | Class |
| Pekar et al. (2004) | Bladder, rectum, fem.h. | CT | Mesh-to-volume | I-Profile | Expl |
| Shang and Shen (2006) | Prostate | US | Optimization | I-Profile | Untr |
| Hodge et al. (2006) | Prostate | US | Optimization | I-Profile | Expl |
| de Bruin et al. (2003) | Aortic aneurysms | CT | Parameterization | I-Profile | Class |
grouped by the analyzed body parts: Bone structures, brain structures, cardiac structures and soft tissue structures in the abdominal and pelvic area. Each entry in the table contains information about the modeled object of interest, the image modality, and the techniques used for the work. Specifically, the method used to acquire the required point correspondences is listed as:

- Point matching, mesh-to-mesh: see Section 4.1.
- Mesh-to-volume: see Section 4.2.
- Volume-to-volume: see Section 4.3.
- Arc-length, SPHARM, parameterization: see Section 4.4.
- Optimization: see Section 4.5.
- Man + X: manual landmarking for a limited set of points, followed by an automatic propagation method.

The employed appearance model is categorized as Profile (Section 5.1) or Region (Section 5.2). If applicable, the type of feature used is given as:

- I, NI: (normalized) intensity.
- G, NG: (normalized) gradient.
- F: extracted features like wavelet coefficients, etc.
- C15: feature developments in the field.

A cl. in front indicates that the appearance model is clustered. The selection strategy or rating method for the fit of an appearance model at a new position is given at the end in parenthesis. Possible values are:

- Untr: untrained model, usually search for the strongest gradient, etc.
- Expl: explicit rules defining the best appearance.
- Mah: mahalanobis distance to Gaussian appearance model.
- Cor: correlation to one or several learned templates.
- Class: classification into foreground/background or edge/non-edge.
- Res: residuals to Gaussian appearance model, used for AAMs.

Finally, the search algorithm is listed as:

- ASM: Active Shape model, see Section 6.2.
- AAM: Active Appearance model, see Section 6.3.
- PF: parameter fitting by Nelder–Mead simplex, quasi-Newton, or conjugate gradient (CG) optimization, see Section 6.4.
- Deformable, AFDSM, m-rep: Algorithms based on deformable models with extended variability, see Section 6.5.

### 7.2. Other applications

In addition to segmentation, SSMs are employed to a variety of other applications in medical image analysis. One common task is that of shape analysis, i.e. the correlation of the shape of certain structures to patient-specific attributes in clinical studies. In the simplest case, this is accomplished by classifying the shape parameter vector of the object of interest into one of several groups. Another common task is the extrapolation of shapes from sparse 3D data to detailed representation, as required for some intra-operative navigation scenarios. Hawkes et al. (2005) have published a discussion paper about models in image-guided interventions, which closer reviews some works in that area. An overview of some applications other than segmentation is given in Table 2. The terms used for correspondences are the same as in Table 1, refer to Section 7.1 for details.

### 8. Discussion

In the preceding sections, we have reviewed the current state of the art in statistical shape modeling. To conclude our survey, we will recapitulate the main points and make some predictions on future developments in the field.

#### 8.1. Shape representation

While there is a huge variety of different shape representations, landmark-based systems are by far the most popular ones. Their

<table>
<thead>
<tr>
<th>Authors</th>
<th>Object(s) of interest</th>
<th>Correspondences</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shape analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paulsen et al. (2002, 2003)</td>
<td>Ear canal</td>
<td>Mesh-to-mesh</td>
<td>Correlation to gender</td>
</tr>
<tr>
<td>Styner et al. (2003, 2004)</td>
<td>Hippocampus</td>
<td>SPFARM</td>
<td>Correlation to schizophrenia</td>
</tr>
<tr>
<td>Davies et al. (2003)</td>
<td>Thalamus</td>
<td>Optimization</td>
<td>Correlation to schizophrenia</td>
</tr>
<tr>
<td>Ciemansky et al. (2004)</td>
<td>Hippocampus</td>
<td>Volume-to-volume</td>
<td>Correlation to schizophrenia</td>
</tr>
<tr>
<td>Kim et al. (2005)</td>
<td>Hippocampus, amygdala</td>
<td>Volume-to-volume</td>
<td>Correlation to attention deficit/ hyperactivity disorder</td>
</tr>
<tr>
<td>Bansal et al. (2007)</td>
<td>Hippocampus, ventricles</td>
<td>SPFARM</td>
<td>Correlation to Alzheimer disease</td>
</tr>
<tr>
<td>Thompson et al. (2004)</td>
<td>Brain ventricles</td>
<td>Mesh-to-mesh</td>
<td>Correlation to twins</td>
</tr>
<tr>
<td>Styner et al. (2003)</td>
<td>Cortical surface</td>
<td>Mesh-to-mesh</td>
<td>Correlation to aging</td>
</tr>
<tr>
<td>Ferrarinini et al. (2007)</td>
<td>Femur</td>
<td>Mesh-to-mesh</td>
<td>Correlation to neurologic and psychiatric disorders</td>
</tr>
<tr>
<td>Wang et al. (2003)</td>
<td>Femoral heads</td>
<td>Optimization</td>
<td>Extrapolate geometry from sparse 3D data for surgery</td>
</tr>
<tr>
<td><strong>Shape extrapolation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fleute et al. (1999)</td>
<td>Femur</td>
<td>Mesh-to-mesh</td>
<td>Extrapolate geometry from sparse 3D data for surgery</td>
</tr>
<tr>
<td>Chant et al. (2003)</td>
<td>Femur</td>
<td>Volume-to-volume</td>
<td>Extrapolate geometry from sparse 3D data for surgery</td>
</tr>
<tr>
<td>Zheng et al. (2005), Rajamani et al. (2007)</td>
<td>Femur and pelvis</td>
<td>Volume-to-volume</td>
<td>Extrapolate geometry from US point data for surgery</td>
</tr>
<tr>
<td><strong>Other applications</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsol et al. (1998)</td>
<td>Crest lines on skull</td>
<td>Mesh-to-mesh</td>
<td>Assess mandible deformation under craniofacial disease</td>
</tr>
<tr>
<td>Andresen et al. (2000)</td>
<td>Mandible</td>
<td>Mesh-to-mesh</td>
<td>Model mandible growth</td>
</tr>
<tr>
<td>Blackall et al. (2001)</td>
<td>Liver</td>
<td>Volume-to-volume</td>
<td>Model deformation caused by breathing</td>
</tr>
<tr>
<td>Rueckert et al. (2003)</td>
<td>Brain</td>
<td>Volume-to-volume</td>
<td>Register landmarks to new subjects</td>
</tr>
<tr>
<td>Lee et al. (2005)</td>
<td>Levator ani</td>
<td>Optimization</td>
<td>Model breathing motion for bronchoscopy simulation</td>
</tr>
<tr>
<td>Deligianni et al. (2006)</td>
<td>Bronchial tree</td>
<td>Feature points</td>
<td>Generate variable scenes for surgery simulator</td>
</tr>
<tr>
<td>Sierra et al. (2006)</td>
<td>Uterus</td>
<td>Man + Subdivision</td>
<td></td>
</tr>
</tbody>
</table>
strongest point is their simplicity: Landmarks are intuitive to understand and painless to implement – a decisive difference in comparison to methods like SPHARM or spherical wavelets. We therefore expect landmarks to remain the dominant representation of SSMs for some time to come. This is not to say that they are the optimal choice for all applications: If intuitive shape descriptions, e.g., of diameters and bending angles, are important for diagnostic purposes, a technique like the m-rep representation seems more suitable. Fourier surfaces, SPHARMs and surface harmonics feature a compact shape description by a limited number of coefficients and have the advantage of an inherent coarse-to-fine characteristic: The more coefficients are used, the more accurately these methods approximate a given surface.

8.2. Shape model construction

Regarding the alignment of training shapes, generalized Procrustes analysis and tangent space scaling are the current methods of choice. Both are reasonably simple to implement and backed by well-elaborated theory. In any case, extensions or replacements for these procedures have not met with much resonance until now.

The choice between the methods of dimensionality reduction and modeling is more open. On the one hand, the PCA approach employed by the vast majority of studies unquestionably has its weaknesses: A number of modeled shapes will certainly not feature Gaussian distributions and the linear approximation model will be suboptimal. On the other hand, PCA is fairly robust to the input data distribution and generally just works. ICA seems like a reasonable alternative, but it has yet to be shown in each specific application if the additional complexity is worth the effort. Nonlinear decompositions as Kernel PCA have hardly attracted attention for landmark-based shape modeling so far. Looking at the popularity these methods have gained in the machine learning community, this seems surprising. However, a decisive difference is that typical applications in machine learning and pattern recognition have a far larger amount of training data available. For 3D SSMs, problems as overfitting and and lacking robustness of nonlinear methods are thus more present. To gain more insight into the question of dimensionality reduction for landmark-based shape modeling, a detailed study comparing different decomposition methods on a variety of training data is indispensable.

Enlarging variations by creating artificial training shapes or modifying the covariance matrix directly is a possibility to enhance the flexibility of a model. However, enhancing flexibility during training cannot guarantee that an ASM or AAM search will fit all new data exactly, as the allowed variation is still limited by hard constraints. Therefore, additional flexibility should better be incorporated by the search algorithm.

8.3. Shape correspondence

The issue of correspondence is one of the key points in shape modeling and has drawn a lot of attention to it. Currently, the main problem in the area is the lack of reliable measures to quantify model quality: Several authors have reported inconsistencies in the generalization, specificity and compactness measures, which have been used for this purpose on the few occasions where comparisons actually have taken place. A comparison to manually placed landmarks is prohibitive in 3D, both due to the required manual work and the inter- and intra-observer variances. Right now, the only way to reliably estimate the quality of an SSM is to test it in the specific application. To determine the usefulness of the plethora of approaches and to evaluate the impact that point correspondences have on an SSM, a comparison between different correspondence methods seems inevitable and overdue.

As long as this has not taken place, we would venture to predict that population-based correspondence approaches (based on MDL) will come out somewhere at the top of the ranking: First of all, MDL optimization is the only method with a convincing theory about why it should deliver good correspondences. Moreover, it is also the method that makes best use of the training data, considering not only pairs of shapes, but all samples at the same time.

8.4. Local appearance models

In our opinion, local appearance models are the technical part with the highest potential to improve SSM-based image segmentation. Until now, the majority of studies has employed standard profile sampling techniques and estimates the goodness of fit using the Mahalanobis distance. However, more complex image features like Gabor wavelets (which are employed in the computer vision community for years) are likely to improve the object description for many applications. Moreover, training classifiers to distinguish boundary from non-boundary or object from background makes better use of the available data than simply modeling the feature of interest. In all reported cases, classifiers reached significantly better results than standard appearance models. Last but not least, clustering data for similar regions can again improve the performance. As segmentation with an SSM can only be as good as the initial feature detection, improvements in the local appearance model will have a direct impact on the final segmentation accuracy.

8.5. Search algorithms

Besides the correspondence issue and local appearance models, the employed search algorithm is the third major component in an SSM-based segmentation framework. We have divided the available local search algorithms for SSMs into four categories: ASM, AAM, other methods with strict shape constraints and methods with relaxed shape constraints. AAM, while very successful for tasks like face recognition and video tracking, is rarely employed for 3D image segmentation until now. The main problem of the approach is the excessive memory usage of the 3D texture model: To be able to run the method on a normal desktop computer, texture resolution for medium to large structures needs to be scaled down radically. In addition, it appears that the mostly homogeneous texture encountered in typical biological shapes does not deliver the same quality of features as, e.g., face recognition.

Currently, ASM is the most popular method for searching 3D SSMs in volumetric images. The basic algorithm is straightforward to implement, reasonably robust, and fast. A wide range of variants has been presented in recent years, many trying to improve outlier handling. Similar to the situation for the correspondence methods, we observe a lack of evaluation of new methods: At best, the performance of a new variant is compared against the original plain algorithm, without considering any other improvements. For this reason, it is difficult to recommend any specific variant at the moment.

A considerable number of alternative algorithms with strict shape constraints is still used for image segmentation, based on standard optimization techniques like Simplex-Downhill, CG, etc. Although the standard optimization algorithms have been analyzed much more thoroughly than ASM or AAM and feature desirable theoretic properties like guaranteed convergence, their relatively slow convergence rates make them inferior to the specialized searches when it comes to model-based image segmentation.

In our view, algorithms with relaxed shape constraints are the most promising method of all local search approaches. If implemented correctly, they can adapt their behavior from strictly
following the given constraints to a completely free deformation. Especially for soft tissue segmentation, the additional freedom can provide considerable benefits, as the large amount of natural variation in these structures often cannot be captured adequately with a limited amount of modes of variation.

A problem with all local search algorithms is that they always detect a local minimum of the cost function, not the global optimum. If the search is initialized reasonably well and the data similar to that used for training, the found local minimum will be close to the optimal solution and the segmentation successful. However, for initializations far from the true shape or for largely different data (e.g. different acquisition parameters or severe pathologies), the segmentation can result in a complete failure. In this case, the found contour may not show any similarities with the region of interest, which in a clinical setting will severely undermine a user’s trust in the method (we assume here that all segmentation results are verified before use and the error is detected). Two conclusions can be drawn: First, more efforts have to be undertaken to increase stability of the segmentation and reduce the number of complete failures. Second, a software relying on statistical shape models for segmentation should always offer possibilities for manual corrections. These interactive search algorithms are not a hot topic for now. However, once SSMs are actually applied in clinical practice, the possibility of correcting points manually will become a required feature rather than an option which is nice to have.

8.6 Applications

As our survey demonstrates, there is already a considerable amount of 3D SSMs employed in medical image analysis. Segmentation is the driving application for now, with noticeable focus on brain and cardiac structures. The striking aspect here is that the majority of applications do not make full use of the techniques available: A considerable amount of applications, e.g. still use generic appearance models, although the possibility to train appearance is a key aspect of SSMs.

Statistical shape models achieve their best performance for objects with systematic variations that can be captured by a reasonable number of modes – in our experience, a good model should be able to capture 90% of the total variance in the training set with less than a dozen modes. Medical objects of interest with a relatively stable shape over the population like the above-mentioned deep brain and cardiac structures – but also most bones – belong to this category and are ideal for shape modeling. Highly varying soft-tissue structures like the liver or vessel-systems are much harder to model and will most likely need some kind of additional flexibility in the search algorithm. Predominantly random shapes like tumors are unsuitable for statistical shape analysis with the presented methods.

It is interesting to note that all applications described in this review are developed and used exclusively for research purposes. We are not aware of any currently available commercial products that employ 3D statistical shape models. However, as technical details of commercial programs are often kept secret, it might still be that some existing products do already use this technique. In any case, we are certain that these products are under development and will appear on the market in the near future. The challenges in designing a commercial, i.e. clinically applicable product to segment medical images with SSMs are numerous. First, training data has to be sufficiently general to be useful for all clients’ sites. In a research environment, methods are generally trained and tested on images from a single source. In practice, different scanners generate images with different appearance. Thus, to assure adequate performance for a commercial product, a far larger number of training images are needed. Second, robustness is critical for clinically applied software. Radiologists and physicians will not trust a program that frequently fails to detect the region of interest. Although we do not believe that a fully automatic procedure will be able to segment 100% of images successfully in the foreseeable future, efforts to lower the number of complete failures have to be intensified. This will increase acceptance of statistical shape models in the clinic and lead to wider application and benefit.

Acknowledgements

Tobias Heimann was supported by the German Research Foundation DFG under grant WO 2188/2-1. We would like to thank Marleen de Bruijne from the DIKU group at the University of Copenhagen for many helpful comments and suggestions regarding this manuscript. In addition, Klaus Fritzschke, Tobias Schwarz, Ivo Wolf and Sascha Zelzer from our own group have contributed valuable comments.

References


