Hierarchical Vibrations for Part–based Recognition of Complex Objects

Karin Engel, Klaus Tönnies

Arbeitsgruppe Bildverarbeitung / Bildverstehen

Technical Report

Nr.: FIN-03-2009

Hierarchical Vibrations for Part–based Recognition of Complex Objects

Karin Engel, Klaus Tönnies

Arbeitsgruppe Bildverarbeitung / Bildverstehen

Hierarchical Vibrations for Part–based Recognition of Complex Objects

Karin Engel and Klaus Toennies

Department of Simulation and Graphics,
Otto von Guericke University, Magdeburg, Germany
engel@isg.cs.uni-magdeburg.de

Abstract We propose a technique for localization of complex shapes in images using a novel part–based deformable shape representation based on finite element vibration modes. Here, our method gives an extension for Finite Element Models to represent elastic co–variations of discrete variable shapes. It avoids misregistration by resolving several drawbacks inherent to standard shape–based approaches, which cannot detect structural variations and occlusions. Our algorithm uses a hierarchical shape model, involving an evolutionary deformable shape search strategy. The different levels of the shape hierarchy can influence each other, which can be exploited in top–down part–based recognition. It overcomes drawbacks of existing structural approaches, which cannot uniformly encode shape variation and co–variation, or rely on exhaustive prior training.
We applied our method to two different example applications, which include shape detection and discrimination, as well as localization of the desired object under occlusions. Experimental results are promising and show the good performance of our approach. It is robust to changes in the values of parameters used and requires no prior training with regard to shape variation and image characteristics. By utilizing a quality–of–fit function the model explicitly recognizes missing discrete parts of a complex shape, thus allowing for categorization between shape classes.

1   Introduction

Modeling global and local aspects of shape separately is useful for many image processing tasks including object recognition, pose estimation, medical image segmentation and motion tracking, as indicated by [1–4, 12, 14, 18, 25, 39, 41, 44, 50, 51, 54–56, 58, 61–64], among others. Studies on the human visual perception also provide evidence that a representation suitable for object recognition should include a structural decomposition of the object into parts and a description of parts and relations between them [2, 13]. Such representation should cover variations and irregularities in shape and structure due to image noise, object

deformation and possibly change of view point, and should allow representation of objects under occlusion.

The main contribution of this work is a novel representation of complex variable objects that contain multiple parts. It is inspired by Biederman's theory of Recognition by Components [2], but uses the structural decomposition into specific shapes in a top–down manner. Knowing the context of object topology provides crucial information for matching, which makes our approach suitable for model–based recognition, including object localization, segmentation and classification (categorization).

Automatically computing part–based descriptions from a real image is difficult, because this requires object detection, segmentation and decomposition. Our method follows a top–down strategy computing the most plausible explanation of image content given a (set of) prior model(s) in terms of segmentation with maximum quality. A major issue here is to abstract complex variable objects into a simplified representation which alleviates comparison between shapes. In our work, we choose a structural approach, which has the advantage of compositionality saving from an exponential increase in the number of single, linear prototypes required to represent valid variation in a specific shape class. In contrast to, e.g. exemplar–based representations, it can represent qualitative and discriminative features of objects. We propose a hierarchical model that also provides a natural framework for multi–scale decomposition of shape into parts, and parts into sub–parts. Using such hierarchy, interactions are more global, which has the advantage that information can propagate faster between distant shape parts.

This paper is structured as follows. Section 2 discusses related work in the context of structural models and techniques for deformable shape search for locating object instances in images. In section 3 we introduce our shape model for use in part–based recognition. Section 4 describes the algorithm for hierarchical shape matching using dynamic Finite Element Models (FEM). In section 5 we introduce our strategy for automating the shape search. Finally, we present two case studies for applying our model to recognition and classification tasks and discuss the results in section 6. At the present state, we do not attempt to estimate locations for object parts in arbitrary poses, instead restricting ourselves to cases in which the set of poses is limited (e.g., due to anatomical laws, standardized image acquisition, etc.). First, we provide results on the localization of a specific cortical folding pattern in 2D flat maps of the cortical surface that we recently presented in a short conference article [14]. It shows exemplarily how to specify different spatial relationships of shape parts. It also provides an example for representing shape parts with different level of detail, and further indicates how to use the structural shape matching as a focus–of–attention technique for shape detection (detecting the presence of a specific shape in an image). In the

second application example, the model is applied to the recognition of ant species from 2D color images from a database, which was also used by [4], and allows us to compare our method directly with their results. In this application we analyze insensitivity w.r.t. hidden object parts. Here, the model is also used for classification by comparing segmentation quality under different class–specific models. A conclusion is drawn in section 7 along with suggestions for further research.

2   Related Work

Variations in shape and in the structural relations between shape parts can be described in different ways. Many existing structural approaches are data–driven in that prior knowledge about the specific shape and the structural aspect of shape is not used for recovering the object parts. This often involves fitting parametric shape models (e.g., geons [2], generalized cylinders [3] and superquadrics [52]) to preprocessed data or images which are pre–segmented into regions corresponding to parts, and comparing the obtained description, e.g. with a database [2, 3, 39, 44, 50, 52, 58]. In many applications, however, specific (classes of) shapes have to be detected in images. Structural models that do not use prior knowledge about specific shapes, may then produce ambiguous results, because generic templates can fit a large range of shapes. (Local) variation may indicate both differences between class members and variation between shape classes. Another problem is that some objects may not have a clear decomposition into shape primitives, while the extraction of generic parts from images in a robust way remains elusive.

   A majority of structural approaches use pre–defined combinatorial constraints (e.g., by specifying attachment surfaces, articulations and joining angles) between simplified object parts to encode the compound shape, e.g. in a structural description graph [18, 31, 44, 53, 64], by a grammar [50], using coupled/split shapes [22, 51, 61], a blending function [12], or an expert model [60]. These approaches either cannot describe structurally variable shapes, or they capture only relatively weak structural properties of shape in their tree–structured models. This may not always be appropriate. If shape parts are constrained to have certain spatial relationships with each other, this information should be used in object detection, such that structural co–variations may induce deformations of the sub–shapes. Several approaches use a statistical/trained model of shape locations [6, 20, 41, 56], e.g., in terms of trained relations between prototypical sub–shapes for generating expectation maps in a sequential recognition process [4, 19, 54, 55]. Representing shape and structure by different models can be a drawback of these methods, as it does not allow structural deformations to directly influence morphological variation, and vice versa. This, however, may be required because often the structural aspect of shape is not independent from

local shape variations. The ASSM approach by Al-Zubi et al. [1] combines both statistical and structural a-priori knowledge about shape variation for 2D shape matching. Their structural representation was shown to profit mainly from the uniform description of co–variation and shape context based on a joint Gaussian model. Similarly, Cristinacce and Cootes [6] use a set of appearance models in a shape constrained local model search. These approaches outperform the original ASM [8] and AAM search [10], respectively. However, they require intensive training with labeled data of high reliability for separating valid from invalid variation. Shape variation may alternatively be constrained a-priori by means of physically based models, such as Finite Element Models (FEM) [40, 46], and training the natural variation modes may not be required.

In this paper, we describe a novel method to present and segment complex objects of specific classes from images. It resolves the addressed problems, in that we adopted the ASSM approach by Al-Zubi et al. [1], and propose a hierarchical Finite Element Model, which gives a uniform and multi–resolution description of shape without the need for prior training. By using deformable shape models, in contrast to simplified fixed templates for shape parts arranged according to a tree–structure (e.g. [18]), we are able to model variations in shape, contour and appearance. The shape hierarchy can capture long–range contextual information, as opposed to appearance and localized structural information only. Compound objects introduce many new issues related to the use of structural relationships to effect model–based segmentation and shape description. We address these issues through specific properties of the proposed model and matching algorithm. The shape model is applied in a hierarchical fashion where contextual shape information is used top–down to eliminate false interpretations of the data. It extends the ability of FEM in the sense that it enables structural variability in terms of elastic co–variations between specific deformable shapes. A significant result of using finite element vibration modes in the shape–structure hierarchy is that it combines noise robustness from deformable shape models and validation of structure from structural models. This reduces the complexity of the distribution function–which would be needed to model the spatial relations of parts of the compound shapes statistically–while modeling valid variation under the following assumptions. First, the desired object can be detected in an image and is (at least partially) visible. Second, variation due to change of view point is negligible, if the set of poses is limited. Finally, shape classes can be differentiated for comparison of shapes based on their structural configuration and/or morphology of the shape parts.

As any deformable model, the resulting shape model relies on the initial estimate. Our framework for shape search (i.e. localization of the shape within the image) is different from other object recognition approaches (e.g. [43, 63]) that guide the search for good object configurations in a bottom–up fashion.

These methods have the disadvantage that false negatives in the feature detection step can prevent parts from being properly localized. In order to deal with the complexity due to variations in pose, shape and appearance of compound objects, in [48] a brute force search is adopted. Felzenszwalb and Huttenlocher [18] developed an efficient method for matching tree–structured models that consider a small number of simple parts connected in a deformable configuration. Their pictorial structure matching provides a global search method [17–19] based on dynamic programming, which was also used in [11]. Their approach has the drawback that it uses pairwise constraints instead of a "full" shape model, where every sub–shape may be related to every other sub–shape. This usually requires a triangulated template of low resolution which is transformed to a small fixed number (e.g. 60×60) of possible locations in the image. Although an optimal non–rigid transformation of the template can be found in polynomial time (w.r.t. sampling density) [19]–which is better than an exhaustive search–such matching algorithm may practically still be ineffective unless being combined, e.g. with an oversegmentation of the image [38, 42]. Moreover, since their energy function is a sum of costs per triangle, relationships between distant parts of an object cannot be captured explicitly and the search is inherently a sequential process [19]. Again, false negatives in the part detection can prevent the shape (parts) from being properly localized. In contrast, our shape search uses an evolutionary strategy, which is inspired by the genetic algorithm Hill and Taylor employed for model–based segmentation of medical data [24]. It combines aspects from [4, 8, 23, 35] in that our local search procedure also uses a top–down, knowledge–driven model–fitting strategy. The main contribution of our approach is twofold. First, training regarding data input, shape (context) and appearance is not required, although this may improve the recognition rate. Second, in contrast to sequential search methods, our method uses a top–down method for hierarchical initialization and validation of the local searches with specific sub–shapes, which is combined with the bottom–up flow of information derived from the data. This integrates a focus–of–attention like search strategy and the segmentation of the desired objects into the object recognition process.

3   Hierarchical Shape Model

As in our case a structural description of the image cannot be derived automatically, the prototypical shape templates are generated from a manually segmented example image. A class-specific prototype is here represented as a hierarchy of Finite Elements Models to account for variability in relationships between the shape parts (which also refer to morphological components, sub–shapes or local shapes) and local shape deformation. This allows to specify the model from domain knowledge by first deciding on the decomposition into simpler components, e.g. convex shapes, which may represent (functional) shape units, e.g.

body parts. Secondly, it allows combination of components by a higher level FEM (the structural, or global model) whose decomposition describes the desired structural constraints between local shapes. Finally, the finite element decomposition supports an efficient simulation of deformation.

A hierarchical shape model $\mathcal{T}_{\mathcal{X}} = \bigcup_{\{l,j\}} \mathcal{T}_j^{(l)}$ represents the structural decomposition of a complex shape of class $\mathcal{X}$ into multiple discrete shapes $\{\mathcal{T}_j^{(l)}\}$, which contribute to different hierarchy levels l, and are represented as triangular finite element meshes. The shape parts at each level $l-1$ are coupled to form a higher level l of the hierarchical shape model, while any $\mathcal{T}_j^{(l-1)}$ may represent a compound shape on its own. Such a hierarchical model deforms into an object instance supported by image features. It should be able to localize instances of an object class based on the amount of structural and morphological deformation necessary to fit the features. We assume that it sufficiently separates information about variation within an object class and between different object classes in order to differentiate between shapes based on deformation parameters. The former is mainly determined by sub–shape deformation and their co–variation, while the latter is given by the decomposition into sub–shapes and structural variation parameters. This hypothesis allows for a certain amount of ambiguity, but we believe that this separation of within–class and between–class variation is sufficient for localizing and classifying objects within the same representational framework.

Under these hypotheses, morphological and structural aspects can be represented by a shape–structure hierarchy of FEM which refers to a 2–level FEM. In this case, the first level represents local shape variations, while the top level represents a-priori information about the co–variations, i.e. class–specific configuration and spatial relationships of the morphological FEM, $\mathcal{T}_j^{(1)}$, in terms of a structural FEM, $\mathcal{T}_{\mathcal{X}}^{(2)}$. Although we restrict our discussion to a shape–structure hierarchy of two levels, the decomposition may continue through additional levels yielding a hierarchy of shapes, structures, super–structures, etc.

The spatial configuration of such system is described by its degrees of freedom (DOF). The finite element nodes on each level represent object details, and are mapped to "sensors" which extract object–specific image features at the location of the node (sect. 4.2). The a-priori information is used for hierarchically combining object features to a more complex object (figure 1), as the shape model deforms according to sensor input (sect. 4.3). Thereby, characteristic features are hierarchically derived from the image. Their significance and semantics increase from the bottom to the top level.

The global structure of the compound object is embedded into the shape hierarchy such that the boundaries of the structural finite elements approximate elastic co–variations (i.e. relative positions) of "high–level image features" which represent a correlation with sub–shapes. Each node $\mathbf{x}_j^{(l)}$ of the structural

Hierarchical Shape Model $\mathcal{T}_{\mathcal{X}} = \bigcup_{\{l,j\}} \mathcal{T}_j^{(l)}$

Structural Fit
$l >= 2$
(sect. 4.3)

Shape Fit
$l = 1$
(sect. 4.1)

Data
(sect. 6)

across–level
spring forces
(eq. 6)

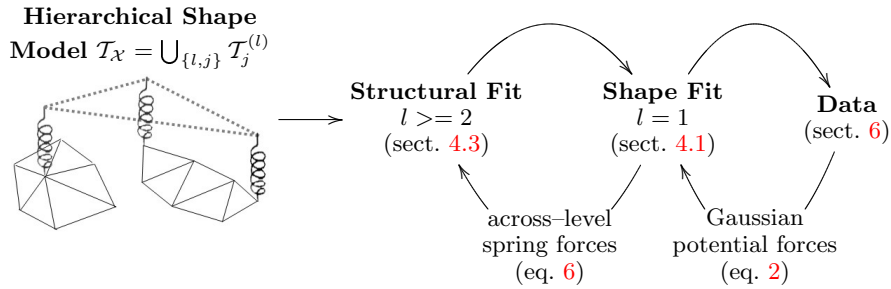Gaussian
potential forces
(eq. 2)

Figure 1: Algorithm overview. In the bottom–up flow of information specific features are extracted from the data. These are combined to more complex objects in a top–down manner using a hierarchy of shape models. An example of such model is depicted on the left. Here, the shapes $\mathcal{T}_j^{(l)}$ of level $l = 1$ (solid lines) and the top–level model, $\mathcal{T}_{\mathcal{X}}^{(l)}, l = 2$ (dotted lines), contribute to a shape–structure hierarchy, or 2–level FEM. The shapes are coupled across different levels l using virtual springs between specific link nodes. Note that the kind of coupling determines the form of co–deformation between sub–shapes.

FEM $\mathcal{T}_{\mathcal{X}}^{(l)}$ is therefore mapped to a high–level sensor representing a particular sub–shape $\mathcal{T}_j^{(l-1)}$, which is likewise implemented as dynamic FEM. Each $\mathbf{x}_j^{(l)}$ is related to the specific morphological FEM through a (set of) link node(s), $\mathbf{x}_j^{(l-1)}$, by a virtual spring of zero length (figure 1). The structural model $\mathcal{T}_{\mathcal{X}}^{(l)}$ defines the (initial) placement of the morphological FEM on the hierarchy level $l - 1$, whose deformation will cause the sub–shapes to fit local image features. In the bottom–up flow of information, the input for the finite element nodes of level $l > 1$ does not stem directly from the underlying image, but from the output of level $l - 1$ of the hierarchy of FEM, i.e. only sub–shapes of level $l = 1$ have direct access to the image. Here, as usual, external model forces subject to (s.t.) the nodes at level $l = 1$ are defined via Gaussian potentials based on the image data. The different levels influence each other during hierarchical shape fit, but in contrast to, e.g. [51], shapes on consecutive levels are not coupled directly. We introduce "high–level external model forces" in terms of across–level spring forces s.t. pairs of link nodes for applying constraints on the relative positions and dependent deformation of shape parts at the lower levels (sect. 4.2). Sub–shape displacements will thus cause deformation of the structural FEM, which will then again effect morphological FEM by propagating the displacement of top–level link nodes $\mathbf{x}_j^{(l)}$ to the lower level link nodes $\mathbf{x}_j^{(l-1)}$. Thus it is possible to separately analyze the deformation behavior of the sub–shapes and their structural relations. In contrast to [51], this yields a hierarchy of FEM, whose nodes are subject to external model forces which are derived from the image in a

bottom–up fashion, and by employing the deformable model paradigm. Thereby, our model is capable of representing local and structural variability of shapes within a uniform framework.

Depending on the number and location of the sub–shapes link nodes different structural attributes of the compound shape can be represented by the related global model, including high–order relations, such as parallelism (figure 2). We employ the following rules for building such hierarchical model based on an example segmentation: First, the sample is decomposed by outlining the parts which are subdivided into geometrically simpler finite elements, e.g. using Constrained Delauny triangulation and Blum's medial axis transform. Next, the shape parts are mapped to nodes $\mathbf{x}_j^{(2)}$ of the structural model, i.e. compact shapes are mapped to a single top–level node (using a single (internal) link node) (figs. 2b–2d), while more elongated shapes might be constrained by at least two top–level nodes (using at least two link nodes) (figs. 2a, 2c). Finally, the structural model is also triangulated.

4   Multi–Resolution Shape Fit using a Hierarchy of Vibrations

By (hierarchically) linking a set of sub–shapes with a structural model, shape matching can be performed on a global and local context iteratively and fully automatically. The dependent deformations between the morphological and structural FEM are realized through virtual links which provide for boundary conditions (BC) for the particular linear equations of motion.
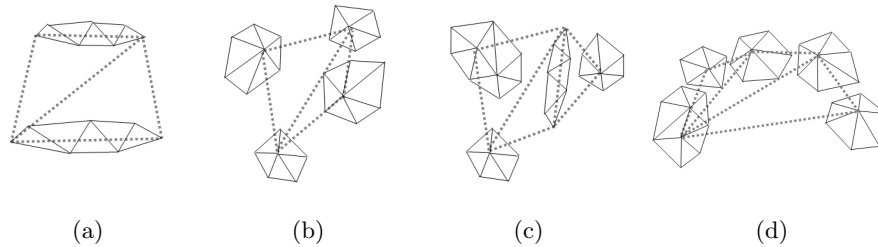

(a)   (b)   (c)   (d)

Figure 2: Examples of hierarchical FEM. Different structural attributes of the compound shape, e.g. the spatial configuration of the sub–shapes, are explicitly defined by the top–level model (dotted lines). In comparison to tree–structured models, such as [4,18], and interconnected/coupled shapes, e.g. [51], distant relations (c,d), parallelism (a) and curved shapes (d) can be represented. Additional desired constraints, e.g. regarding relative orientations, will require the definition of additional link nodes, yielding more complex structural models.

4.1 Dynamic Finite Element Model

Each local shape fit is achieved by deformations which are described by employing finite element vibration modes of the model [36, 40, 46][1],

$$\ddot{\mathbf{q}}(t) + \hat{\mathbf{C}}\dot{\mathbf{q}}(t) + \Lambda \mathbf{q}(t) = \Psi(t), \tag{1}$$

with modal damping matrix $\hat{\mathbf{C}} = \alpha\mathbf{I} + \beta\Lambda, \alpha, \beta \in \mathbb{R}$, identity $\mathbf{I}$ and modal loads $\Psi(t) = \Phi^T\mathbf{f}(t)$. Given a 2D FEM with finite element nodes at positions $\mathbf{x}_i(t) = (x_i, y_i), i = 1, ..., N$, the modal frequencies $0 \leq \omega_1^2 \leq \ldots \leq \omega_n^2 \in \Lambda_{n \times n}, n = 2N$, determine the amount of variation with regard to the modal vectors $\Phi = (\phi_1\ \phi_2\ldots\phi_n)$. Together with the values ω_k^2, the vibration modes $\phi_k, k = 1, \ldots, n$, are solutions to the generalized eigenproblem $\mathbf{K}\Phi = \Lambda\mathbf{M}\Phi$. The $2N \times 1$–composite vectors $\mathbf{q}(t)$ and $\mathbf{f}(t)$ denote modal displacements and nodal loads at time t, respectively.

A low dimensional displacement field can be obtained by neglecting the high–frequent vibration modes using $\mathbf{u}(t) \approx \bar{\Phi}\mathbf{q}(t), \bar{\Phi} = (\phi_1\ \phi_2\ldots\phi_m), m \ll n$, and interpolating displacements $\mathbf{u}^e$ over the $e = 1, ..., M$ finite elements, i.e. $u(x, y) = \sum_e H^e\mathbf{u}^e$. In our case, H^e contains linear finite element approximation functions. See [46] for details. To simulate the deformation of the template, the finite element equations are integrated over time until minimum change in nodal displacements. Note that the coordinates of a point originally located at $\mathbf{x}_i^0$ are $\mathbf{x}_i(t) = \mathbf{x}_i^0 + \mathbf{u}_i(t)$ in the deformed body.

4.2 External Model Forces

In our case, external model forces $\mathbf{f}(t)$ shall attract the nodes to object features in the image. Such dynamic loads are created by sensors at the finite element nodes, which sample normalized feature maps $I^{\mathcal{N}}$.

For the low–level nodes $\mathbf{x}_i^{(l)}, l = 1$, the features are derived from the image I by spatial filtering. The corresponding external forces, or Gaussian potential forces, are defined based on the gradient of feature intensity at the current nodal position $\mathbf{x}_i^{(1)}(t), i = 1, \ldots, N$, i.e.

$$\begin{aligned}
\mathbf{f}_i^{(1)}(t) &= \kappa\nabla I^{\mathcal{N}}(\mathbf{x}_i^{(1)}(t)), \\
\mathbf{f}_i^{(1)}(t) &= \kappa(\nabla I^{\mathcal{N}}(\mathbf{x}_i^{(1)}(t)) \cdot \mathbf{n}_i)\mathbf{n}_i,
\end{aligned} \tag{2}$$

where $I^{\mathcal{N}} \in [0, 1]$ and κ is a constant weight. The second formulation allows to restrict deformations according to the actual orientation of the shape boundary [32, 59], by using only the locally radial component of the forces, where $\mathbf{n}_i$

---

[1] For clarity of presentation, we will omit the upper index l, given the method applies to shape models of any hierarchy level l.

is the unit normal in $\mathbf{x}_i^{(1)}(t)$. Usually, two types of low–level sensors are of interest for segmentation of objects. Intensity sensors sample a Gaussian low pass filtered version $I_\sigma^\mathcal{N} = G_\sigma * I$ of the image[2], and are typically assigned to nodes representing the homogenous object interior. The corresponding intensity forces are then computed using $I^\mathcal{N} = I_\sigma^\mathcal{N}$, with $\kappa > 0$ in cases of high intensities in the interior and low intensities in the background. The contour sensors typically sample gradient magnitude maps, i.e. we let $I^\mathcal{N} = I_\nabla^\mathcal{N} = |\nabla I_\sigma^\mathcal{N}|^2$, and $\kappa > 0$ in equation 2.

The input for the high–level sensors, $\mathbf{x}_j^{(l)}, l > 1$, of the structural FEM depends on the behavior of the underlying morphological FEM $\mathcal{T}_j^{(l-1)}$. More specifically, their deformation and correspondence with low–level features is used to set up the high–level external forces that are described in sect. 4.3.

For using the hierarchical model in a search for the desired object, the quality of a model instance placed within the image and deformed according to external model forces, has to be evaluated. Our quality–of–fit (QOF) function for a model instance $\mathcal{T}^{(l)}$ combines a measure of deformation, $Q_{def}^{(l)}$, with the correlation with the data, $Q_s^{(l)}$, for estimating overall energy,

$$\mathcal{Q}^{(l)} = \zeta Q_{def}^{(l)} + (1 - \zeta)Q_s^{(l)}, \ \zeta \in [0,1]. \tag{3}$$

In our case, the strain energy which is associated with the nodes of model instance $\mathcal{T}^{(l)}$ is adapted from [46], such that

$$Q_{def}^{(l)} = (1 + \mu\left(\left\{\frac{1}{2}(\mathbf{q}_i^{(l)})^2(\omega_i^{(l)})^2\right\}\right))^{-1} \in [0,1], \tag{4}$$

estimates the energy needed to match model and data. The mean value of the sensor inputs,

$$Q_s^{(l)} = \mu(\left\{\gamma_i I^\mathcal{N}(\mathbf{x}_i^{(l)})\right\}) \in [0,1], \tag{5}$$

indicates the amount of correspondence of model and data. The weights $\gamma_i \in [0,1]$ allow emphasizing nodes which represent significant object details. Given appropriate feature maps $I^\mathcal{N}$, the QOF–function estimates segmentation quality for a shape model of arbitrary complexity in terms of deformation cost and correspondence with the chosen image features. In our case, high–level feature maps which are input to the matching of a shape model of level $l > 1$ represent the $\mathcal{Q}_{\mathcal{T}_j}^{(l-1)}$–values of the associated sub–shapes.

4.3   Hierarchical Shape Matching

Each shape fit is achieved by deformations which are determined by a set of constraints corresponding to finite element vibration modes, as described in sect.

---

[2] σ denotes the standard deviation of the Gauss kernel.

4.1. This concept for local optimization is extended such that our matching algorithm fits the structural model instances to the data in a hierarchical manner. Algorithm 1 summarizes the fit of a 2–level FEM. However, it can be easily extended to hierarchically match any shape hierarchy with $l > 2$ levels.

Let $\mathcal{T}_\mathcal{X} = \bigcup_{\{l=\{1,2\},j\}} T_j^{(l)}$ denote a given 2–level FEM. The global shape model $\mathcal{T}_\mathcal{X}^{(2)}$ restricts parametrization of the morphological FEM $\mathcal{T}_j^{(1)}$ according to the displacements of the $j = 1, ..., N_{\mathcal{T}_\mathcal{X}^{(2)}}$ structural nodes. More specifically, after initializing an instance of the global model $\mathcal{T}_\mathcal{X}^{(2)}$, the instances of the local models are aligned to it by propagating the displacements of the link nodes $\mathbf{x}_j^{(2)}$ in the global model $\mathcal{T}_\mathcal{X}^{(2)}$ to the low–level link nodes $\mathbf{x}_i^{(1)}$ of the sub–shapes $\mathcal{T}_j^{(1)}$ (figure 3a). In this case, the displacement $\mathbf{u}_j^{(2)}$ directly affects the degree of freedom (DOF) associated with link node $\mathbf{x}_i^{(1)}$, and is imposed as displacement boundary condition on the particular equations 1.

The first two steps of the iterative hierarchical shape matching algorithm account for the bottom–up flow of information between the two levels of the model. It is implemented using the hierarchy of forces derived from the particular feature maps described in sect. 4.2. Deformation of a FEM at level $l = 1$ uses external model forces computed by spatial filtering (step M1 in algorithm 1), and for $l \geq 2$ it uses across–level spring forces between pairs of link nodes of levels $l - 1$ and l (M2). The final step (M3) defines the top–down flow of information, which is realized through essential (displacement) boundary conditions (BC) for the particular linear equations 1.

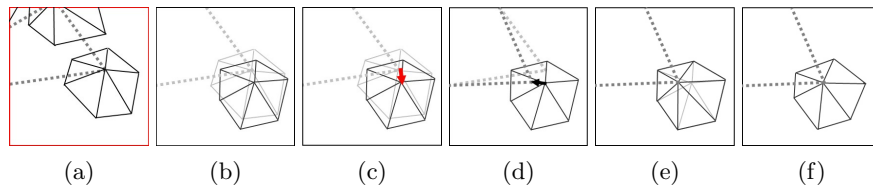(a)   (b)   (c)   (d)   (e)   (f)

Figure 3: Hierarchical Shape Matching. Shape interactions during the hierarchical shape fit are exemplarily depicted for a detail of a shape (solid lines)–structure (dotted lines) hierarchy (a). (b) Result of the morphological shape fit (step M1 in algorithm 1) for a sub–shape $\mathcal{T}_j^{(1)}$ (initial configurations are shown in light gray). (c) The resulting across–level spring force s.t. top–level node $\mathbf{x}_j^{(2)}$ (equation 6) is depicted by the red arrow. (d) The respective second–level nodal displacement determines a displacement boundary condition (black arrow) for the particular first–level link node $\mathbf{x}_i^{(1)}$ (M3), which yields a deformation of the sub–shape $\mathcal{T}_j^{(1)}$ (e). (f) Result after one iteration of the hierarchical matching including steps M1–M3 (sect. 4.3).

Algorithm 1 Hierarchical Shape Fit

1: Initialize the FEM for $\mathcal{T}_\mathcal{X} = \{\mathcal{T}_\mathcal{X}^{(2)}, \mathcal{T}_j^{(1)}\}, j = 1, ..., N_{\mathcal{T}_\mathcal{X}^{(2)}}$.

2: Specify a fixed time step of the simulation $\triangle t$,
   let $t = 0$, and initial values $\mathbf{q}(0) = \mathbf{q}^0, \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}^0$ for problem 1.

3: **repeat** $\{t = (t+1)\triangle t\}$

4:    Select $m(t), \sigma(t)$.

5:    **(M1) Morphological Shape Fit.**
      Each morphological FEM instance $\mathcal{T}_j^{(1)}$, $j = 1, ..., N_{\mathcal{T}_\mathcal{X}^{(2)}}$, of the 2–level FEM
      deforms independently under low–level image forces $\mathbf{f}_i^{(1)}$, $i = 1, \ldots, N_{\mathcal{T}_j^{(1)}}$, dy-
      namically derived from the image according to equation 2.
      The shape fit is iterated until minimum change in nodal displacements. It results
      in sparse high–level feature maps representing the QOF–values $\mathcal{Q}_{\mathcal{T}_j}^{(1)}(t) \in [0, 1]$
      of the morphological FEM instances after local optimization (Note that local
      optimization according to equation 1 should maximize $\mathcal{Q}_{\mathcal{T}_j}^{(1)}$.). This output is
      used to set up suitable forces for the structural search (figure 3b).

6:    **(M2) Structural Shape Fit.**
      The deformation of the instances of the $j = 1, ..., N_{\mathcal{T}_\mathcal{X}^{(2)}}$ morphological FEM is
      used to define the across–level spring forces in terms of high–level dynamic loads
      for the associated second–level nodes $\mathbf{x}_j^{(2)}$:

$$\mathbf{f}_j^{(2)}(t) = \kappa_s \Delta\mathcal{Q}_{\mathcal{T}_j}^{(1)}(\mathbf{x}_i^{(1)}(t) - \mathbf{x}_j^{(2)}(t-1)), \ \kappa_s > 0, \tag{6}$$

      where $\mathbf{x}_i^{(1)}(t)$, $i \in \{1, \ldots, N_{\mathcal{T}_j^{(1)}}\}$, denotes the displacement of the first–level
      link node of morphological FEM $\mathcal{T}_j^{(1)}$ after step (M1) and $\mathbf{x}_j^{(2)}(t-1)$ denotes
      the second–level nodal displacement computed in the previous iteration of the
      hierarchical matching algorithm. We let $\kappa_s = \kappa|\mathbf{x}_i^{(1)}(t) - \mathbf{x}_j^{(2)}(t-1)|^{-2}, \kappa > 0$,
      while $\Delta\mathcal{Q}_{\mathcal{T}_j}^{(1)} = \mathcal{Q}_{\mathcal{T}_j}^{(1)}(t) - \mathcal{Q}_{\mathcal{T}_j}^{(1)}(t-1)$ is the change in quality ("quality gradient")
      for sub–shape $\mathcal{T}_j^{(1)}$. The dynamic load is imposed on the vector $\mathbf{f}^{(2)}$ as a natural
      BC for the particular linear equations 1 (figure 3c).

7:    **(M3) Computation of the First–level Nodal Displacements due to
      Structural Deformation.**
      The resulting second–level nodal displacements $\mathbf{u}_j^{(2)}(t)$ (figure 3d) determine the
      final displacement of the first–level link nodes $\mathbf{x}_i^{(1)}$ of the related morphological
      FEM instances $\mathcal{T}_j^{(1)}, j = 1, ..., N_{\mathcal{T}_\mathcal{X}}^{(2)}$. More specifically, $\mathbf{u}_j^{(2)}(t)$ define essential BC
      that directly affect the DOF associated with the first–level link nodes $\mathbf{x}_i^{(1)}$ of
      sub–shapes $\mathcal{T}_j^{(1)}$ (figs. 3e, 3f). Essential BC are imposed directly on the nodal
      displacements in equation 1.

8: **until** a stopping criterion is fulfilled.

This hierarchical shape fit continues until minimum change in nodal displacements of the global model instance. With this scheme it is possible to separately analyze the deformation behavior of the sub–shapes and their structural relations, while the shapes at different levels can influence each other.

A computationally efficient coarse–to–fine implementation of our matching algorithm fits the structural model instances in a hierarchical manner to the data. In our case, it uses feature maps of different scales of resolution. Low–level feature maps are computed from a Gaussian low pass filtered version of the image using dynamically decreasing values for σ. The high–level features are computed using a dynamically increasing number of $m \leq n$ vibration modes of the elastic shape templates at any level l, e.g. starting with the rigid–body modes. Values for σ and m are updated after each step of the matching.

5   Evolutionary Shape Search

Our algorithm computes the most plausible explanation of image content given a (set of) prior model(s) in terms of segmentation with maximum quality. The best fitting shape instance extracts the desired object from the image and classifies it. For automating the search for the best fitting shape instance(s) within an image, the hierarchical shape matching is directly integrated into an evolutionary search which is inspired by [4, 8, 23, 24]. In contrast to [23], where a highly application-specific segmentation plan is carried out by a single "self–aware" deformable agent, it initializes and optimizes multiple model instances in parallel by using the quality–of–fit (QOF) function introduced in sect. 4.2.

Model instances are initialized by transformation of the prototype $\mathcal{T}_{\mathcal{X}} = \bigcup_{\{l,j\}} T_j^{(l)}$ from the model coordinate frame to the image coordinate frame. Here, we only consider the transformations translation, rotation and scaling, although it is straightforward to use other sets of transformations. The Euclidean transformations are characterized by the set of parameters position $\mathbf{c} = (c_x, c_y)$, orientation ψ w.r.t. a predefined axis, and (isotropic) scaling $\mathbf{s} = (s_x, s_y)$, of the model instance in the image[3]. The parameters $\xi \in X = \{c_x, c_y, \mathbf{s}, \psi\}$ might be considered as variates with a presumed Gaussian[4] distribution $\xi \sim N(\mu, \varsigma)$, and random samples are computed according to

$$x = \hat{\mu}(\xi) + z\sqrt{\hat{\varsigma}(\xi)}, \quad z \sim N(0, 1). \tag{7}$$

This random initialization replaces the initial global (rigid) matching used in [8] to determine a single best first guess which is then deformed to locally

---

[3] We use the angle with the principle axis of a shape to define ψ and the standard deviation in nodal positions to estimate $\mathbf{s}$.

[4] Gaussian sampling might be illposed. The desired function must conform with the data, e.g. acquisition parameters.

adapt to the image features. Because the correlation with a rigid template is likely to provide less significant estimates for promising object locations, we use a quality–of–fit function for deformable templates which includes local optimization. Compared with the sequential search in [4, 19, 55], each model instance is randomized w.r.t. all its levels, and then fitted to the data in a hierarchical manner. This will reduce the risk that false negatives in the feature detection step prevent parts from being properly localized.

Our approach cannot guarantee to find the globally optimal parameter set. This would require a recursive subdivision of the parameter space and analysis of all possible matchings for transformations with parameters contained in the respective region. Also, proper parametrization would require estimation of the parameters $\hat{\mu}$ and $\hat{\varsigma}$ of the probability density functions (PDF). However, in our case, samples in terms of parameter values of accepted solutions are not available. Hence, we specify an initial region of parameter values we are interested in and use pre–set tolerances $\hat{\varsigma}$ from the parameter values x' of a single model instance $\mathcal{T}_{\mathcal{X},\mathcal{M}}$ generated from the representative manual segmentation, which serve as estimates for $\hat{\mu}$ (see sect. 6 for settings we used in our experiments). Note that the parameters for the particular PDF have to be estimated for any $\xi(\mathcal{T}), \mathcal{T} \subseteq \mathcal{T}_{\mathcal{X}}$, which is not constrained by the top–level model parametrization.

Each of the multiple model instances initiates an optimization process in order to adapt to the local conditions in the data (sect. 4). When the hierarchical matching converges, the structural QOF $\mathcal{Q}_{\mathcal{T}_{\mathcal{X}}}^{(l)}$ is computed according to equation 3. Since for $l \geq 2$ the top–level model groups image features in terms of (a hierarchy of) sub–shape candidates (of level $l-1, l-2, ..., 1$), we let $I^{\mathcal{N}}(\mathbf{x}_j^{(l)}) = \mathcal{Q}_{\mathcal{T}_j}^{(l-1)}$ in equation 5 for extending the QOF–function to evaluate combined shapes. For example, the QOF–function defined at the level $l = 2$ estimates the energy of deformation and the degree of correlation between proposed part location and image features.

We can now organize the search by employing a priority queue of regions within the search space, where we use the quality–of–fit of the current model instances as the priority. Solutions with high energy (i.e. quality–of–fit, $\mathcal{Q}_{\mathcal{T}_{\mathcal{X}}}^{(l)}$) are selected by applying a threshold τ_Q, and further evolved until the overall quality of the current model instances, $q(\mathcal{T}_{\mathcal{X}}^{(l)})$, converges. New shape generations are generated based on parametrization of the regionally best fitting shapes. In our case, each selected model instance is randomized w.r.t. all its levels. More specifically, for each shape $\mathcal{T} \subseteq \mathcal{T}_{\mathcal{X}}$ that contributes to an instance of the compound shape with high energy, we use in equation 7, $\hat{\mu}(\xi) = x'$ and $\hat{\varsigma}(\xi) \in [0, 1]$, where x' is the actual value for $\xi(\mathcal{T})$. Shape model instances $\mathcal{T}_{\mathcal{X}}$ with low energy are replaced by new instances based on the initial settings accordingly. Misleading shape searches due to an insufficient parametrization as well as an

exponential increase in the number of solutions can thus be avoided, while the additional "new" trials keep the search independent of known solutions.[5]

For determining $q(\mathcal{T}_{\mathcal{X}}^{(l)})$ and τ_Q, clusters of model instances with high energy are built using a regular grid of bins over the image. Since the exact relation of the desired object and background is not known in advance, we choose the grid size g according to the parametrization of current model instances of the lowest level, $l = 1$, i.e. $g < \min_{\{j\}}(s_x(\mathcal{T}_j^{(1)}), s_y(\mathcal{T}_j^{(1)}))$. Each instance is assigned to the according bin $b(g) = 1, \ldots, g^2$, by means of the geometric mean of its node positions. The overall QOF is then defined by the mean QOF–value of the best rated current model instances over all clusters, i.e. $q(\mathcal{T}_{\mathcal{X}}^{(l)}) = \mu_{b(g)}\big(\max(Q_{\mathcal{T}_{\mathcal{X}}}^{(l)} : \mathcal{T}_{\mathcal{X}}^{(l)} \in b(g))\big)$, and we let $\tau_Q = q(\mathcal{T}_{\mathcal{X}}^{(l)}) - \frac{1}{10}$. The multi–resolution shape search continues until the overall QOF converges, such that the desired shape is finally represented as the best rated structural configuration of sub–shapes in the current image. (Depending on the desired precision, the search may also be terminated if a solution can be reported with a QOF–value above a pre–defined threshold, τ'.) Note that our algorithm can determine the $k > 1$ best matches, which is useful, if the number of instances of the desired object within the image is not known. If the QOF–value of the best match is below a pre–defined threshold, the specific object could not be detected in the given image.

Model–based approaches that use prior knowledge about specific shapes offer a complete characterization of the fitted shapes and imply classification. Each object is identified under a given model $\mathcal{T}_{\mathcal{X}}$ with a probability depending on the structural quality–of–fit $\mathcal{Q}_{\mathcal{T}_{\mathcal{X}}}^{(l)}$ of the best fitting instance of this prototype. Hence, the competitive use of different class–specific shape models allows for classification of objects within the image by comparing $\mathcal{Q}_{\mathcal{T}_{\mathcal{X}}}^{(l)}$–values for the different prototypes $\mathcal{T}_{\mathcal{X}}, \mathcal{X} = \mathcal{A}, \ldots, \mathcal{Z}$.

---

[5] In order to avoid repetitive computations one may use a match list of known solutions, and modify/discard parametrizations in case of close proximity.

6   Localization and Discrimination of Complex Objects within 2D Images

We selected two specific examples, where it is relatively easy to quantify success or failure, in order to explore a number of different trade–offs, such as the role of low–level versus high–level features, or cues, the role of (valid) decomposition of shapes, and the appropriateness of a-priori constrained prototypical "zero–order" models of shape variation. In both examples, the desired object can be identified based on the class–specific configuration of specific shape parts. In contrast to the classical geon approach [2], identification and combination of the shape parts is in our case governed by general assumptions on structural interpretation of the object anatomy. Hence, a sub–shape may not necessarily represent a geometrical unit, but a functional part of the compound shape. This kind of assumptions on the structural decomposition may be introduced by a human expert, but can often also be derived from inspection of the images and example segmentations.

In the first example application, a specific gyrus which contributes to the auditory cortical folding pattern is localized in 2D flat maps of the cortical surface (first results have been presented in [14]). It shows exemplarily how to specify different spatial relationships on the top level of the shape hierarchy. It also provides an example for demonstrating the use of hierarchical deformable template matching as a means to detect and discriminate the desired object from similar shapes within the image. The automatic delineation of regions of interest is a basis for the accurate functional parcellation of the human cortex using neuroimaging [5]. Here, our approach can provide a means to overcome problems inherent to the gross inter–subject variability in brain anatomy by the top–down utilization of structural a–priori knowledge [15, 16].

In the second case study, our model is applied to the recognition of ant species from 2D color images from a database. This specific application is well–suited for a number of different reasons. First, such image databases are becoming an increasingly powerful tool for a wide range of applications, including entomology, biology and longitudinal studies using medical imaging. Since these databases are constantly extended, it is desirable to have tools which automate and support classification. Second, in this application we can analyze insensitivity w.r.t. hidden object parts. Third, it allows us to compare our results with the results presented in [4], who used statistical color–classification for sensor input, and a statistical structural model in a sequential search to recognize ants from the same database. In contrast, no prior training is required in our approach.

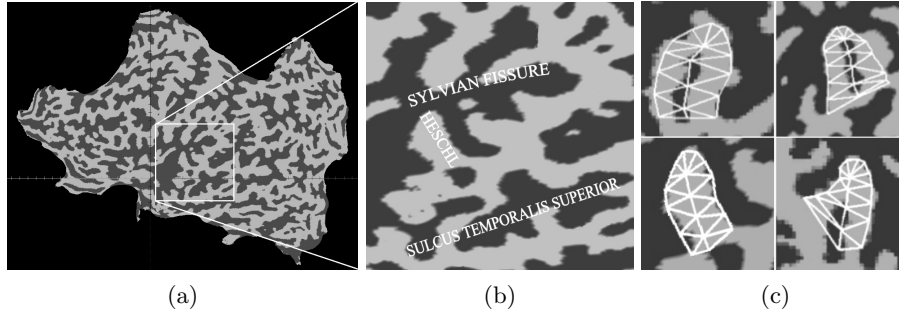<div align="center">(a)   (b)   (c)</div>

Figure 4: Auditory folds on a cortical flat map. (b) Trimmed patch of the flat map (a) containing the Sylvian fissure, Sulcus temporalis superior and transverse temporal gyri (Heschl's gyri). (c) The morphology of (the first) Heschl's gyrus, its size, orientation and absolute position within a reference space vary dramatically between different cortices.

6.1   Case Study 1: Automated Labeling of Cortical Folds

For the identification of Heschl's gyrus[6], which contributes to the auditory cortical folding pattern, our algorithm utilizes the properties of the cortical surface, e.g. 2D flat maps. Such flat maps are in our case derived [21] from the reconstructed cortical white matter [57] using BrainVoyager[7]. It represents the mean curvature pattern of the cortical surface (figure 4a, gyri are indicated by high intensities, sulci by low, non–zero intensities). Locating Heschl's gyrus includes discriminating the desired object from several similar shapes within the flat map. This requires the use of a–priori information regarding its relations with anatomical landmarks for restricting the search space.

A Shape–Structure Hierarchy for the Representation of the Auditory Cortical Folds. As depicted in figure 4b, the auditory cortical folding pattern can be described as a variable configuration of single deformable folds. Heschl's gyrus (HG) is located in Sylvian fissure (SF) and is restricted laterally by Sulcus temporalis superior (STS). We use a 2–level FEM, whose second level represents a–priori information about the co–variations, i.e. structural relations, of the auditory folds. The morphological FEM $\mathcal{T}_{hg}^{(1)}, \mathcal{T}_{sf}^{(1)}$ and $\mathcal{T}_{sts}^{(1)}$, which represent HG, SF and STS, contribute to the first level of the shape model. Since Heschl's gyrus

---

[6] The transverse temporal gyri (Heschl's gyri) contribute to the superior temporal gyrus of the human brain and extend mediolaterally into the Sylvian fissure. As their number varies between brains and hemispheres, we will refer to the *first* transverse temporal gyrus as Heschl's gyrus.

[7] http://brainvoyager.com

(a) $\mathcal{T}_{\mathcal{H}^+}$
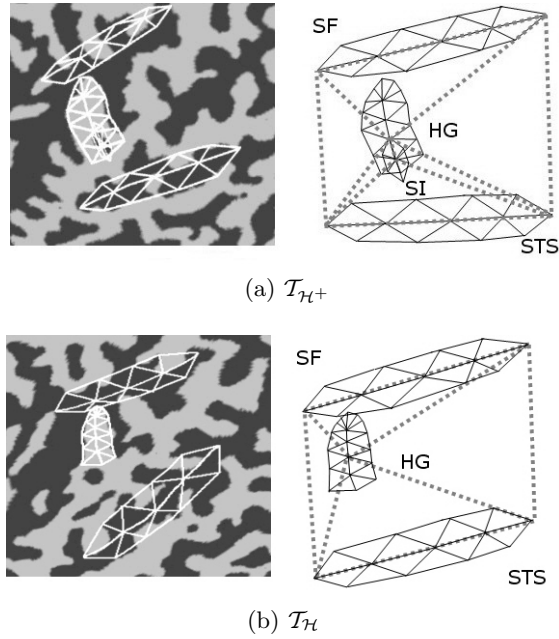

(b) $\mathcal{T}_{\mathcal{H}}$

Figure 5: 2–level FEM of the auditory cortical folding pattern. The morphological components (solid lines) are related to each other as determined by the structural FEM (dotted lines) of the two class–specific models $\mathcal{T}_{\mathcal{H}^+}$ and $\mathcal{T}_{\mathcal{H}}$. In both cases, a central top–level node is mapped to the morphological FEM of HG. In the $\mathcal{T}_{\mathcal{H}}$–model it is related to all other morphological components through a single (internal) link node (b). The $\mathcal{T}_{\mathcal{H}^+}$–model consists of one additional internal second–level node, which is mapped to the morphological FEM of SI (a). SF and STS are represented by pairs of top–level boundary nodes, whose interconnection defines the parallel arrangement of SF and STS.

may or may not show a Sulcus intermedius (SI), two class–specific models ($\mathcal{T}_{\mathcal{H}^+}$ and $\mathcal{T}_{\mathcal{H}}$) are defined accordingly (figs. 4c, 5b, 5a).

For reasons of simplicity the flat maps $\mathbf{V} = \{\varphi_i\}$ are mapped to image matrices, $I^{\mathcal{N}}$, by averaging the curvature values over vertices φ_i which are assigned to pixels of resolution $1mm^2$ based on their position in the flat map coordinate system [21], and mapping the resulting values to $[0, 1]$ (figure 6). The low–level feature maps can then be computed by spatially filtering the images (sect. 4.2).

The internal nodes of the morphological FEM $\mathcal{T}_{hg}^{(1)}$ represent the positive curvature pattern of Heschl's gyrus. The corresponding sensors are mapped to intensity images, $I_\sigma^{\mathcal{N}}$, and in equation 2 we let $\kappa > 0$. Since all nodes of the remaining sub–shapes $\mathcal{T}_{sf}^{(1)}, \mathcal{T}_{sts}^{(1)}$ and the internal node of $\mathcal{T}_{si}^{(1)}$ represent negative mean curvature indicating concave folds, i.e. sulci, we let in these cases $\kappa < 0$.

The boundary nodes of $\mathcal{T}_{hg}^{(1)}$ and $\mathcal{T}_{si}^{(1)}$ represent borders between areas of positive and negative curvature. The relating sensors are thus mapped to gradient magnitude maps, $I_{\nabla}^{\mathcal{N}}$.

The sub–shapes $\mathcal{T}_{sf}^{(1)}, \mathcal{T}_{sts}^{(1)}$ and $\mathcal{T}_{si}^{(1)}$ are modeled with low detail because the exact representation of the highly variable shape of these sulci is not essential w.r.t. the recognition task. We used medial–axis shape representations for all single folds in order to account for the expected high variability in morphology (figs. 4c, 5a, 5b). The class–specific configuration of the morphological components is determined by the second level of the 2–level FEM. (Note that the two models $\mathcal{T}_{\mathcal{H}}$ and $\mathcal{T}_{\mathcal{H}^+}$ share the morphological prototypes.) The elongated shapes $\mathcal{T}_{sf}^{(1)}$ and $\mathcal{T}_{sts}^{(1)}$ have two link nodes each. The corresponding boundary nodes $\mathbf{x}_j^{(2)}, j = 2, 3$ and $j = 4, 5$, of the structural model $\mathcal{T}_{\mathcal{X}}^{(2)}, \mathcal{X} = \mathcal{H}, \mathcal{H}^+$, are connected in order to enforce the near parallelism between SF and STS. The morphological FEM $\mathcal{T}_{hg}^{(1)}$ has a single link node whose corresponding second level node $\mathbf{x}_1^{(2)}$ is connected to the boundary nodes of $\mathcal{T}_{\mathcal{X}}^{(2)}$ in order to make HG be positioned between SF and STS (figure 5b). Thereby, only the relative position of HG is restricted by the top–level model, since it would otherwise erroneously imply significant co–variations in size and orientation, e.g. of HG and SF. The model $\mathcal{T}_{\mathcal{H}^+}^{(2)}$ has an additional internal node $\mathbf{x}_6^{(2)}$ for relating $\mathcal{T}_{si}^{(1)}$ to all other shapes (figure 5a).

Localization of Heschl's gyrus in 2D Cortical Flat Maps. We used 80 flat maps of left and right hemispheres created from high resolution anatomical MRI data in order to test the ability of the 2–level FEM for automatic identification of Heschl's gyrus. A successful shape search required the correct solution–which was confirmed by a neuroscientist expert–to be the best–rated in the ordered list of solutions (priority queue) according sect. 5. On a pixel to pixel comparison the segmentations overlap by $> 90\%$ with the manual segmentations.
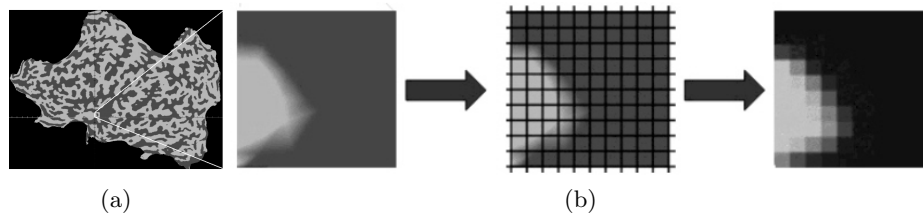

(a)   (b)

Figure 6: For segmentation, each flat map (a) is mapped to an image matrix as illustrated in (b) for a flat map patch of $10 \times 10 \, mm^2$ size. Depending on the amount of distortion due to flattening of the cortical surface, the image pixels represent the curvature pattern integrated over about $1 - 4$ flat map vertices.

All experiments have been done with the same set of empirically chosen material parameters[8], $\mathcal{E} = 0.85 \cdot 10^9 [Pa], \nu = 0.19, \rho = 1[\frac{g}{cm^3}]$, $\zeta = \frac{1}{2}$, $\gamma_i = 1$, $\forall \mathbf{x}_i^{(l)}, l = 1, 2$, as parameter values for the QOF–function (equations 3, 5), and $m(1) = 4, \sigma(1) = 10$ for shape matching (sect. 4.3). For the search we used the following initial set of parameter values, $\hat{\mu}(\xi) = x'$, where x' is known for $\xi \in X$ of all components $\mathcal{T}$ of the model instances $\mathcal{T}_{\mathcal{X},\mathcal{M}}$ generated from the manual segmentations (sect. 5). We further let $\hat{\varsigma}(\psi) = 10°$ for $\{\mathcal{T}_{\mathcal{X}}^{(2)}, \mathcal{T}_{hg}^{(1)}, \mathcal{T}_{si}^{(1)}\}$, $\hat{\varsigma}(c_x) = \hat{\varsigma}(c_y) = 10mm$ for $\mathcal{T}_{\mathcal{X}}^{(2)}, \mathcal{X} = \mathcal{H}, \mathcal{H}^+$, and $\hat{\varsigma}(\mathbf{s}) = \frac{1}{10}\mathbf{s}'(\mathcal{T})$ for $\mathcal{T} = \mathcal{T}_{\mathcal{X}}^{(2)}, \mathcal{T}_j^{(1)}, j = hg, si$.[9]

We then analyzed the performance and robustness in parametrization of the 2–level FEM in comparison with the morphological FEM of HG used in isolation. Using the morphological FEM, $\mathcal{T}_{hg}^{(1)}$, for localizing Heschl's gyrus was successful in 54% of all cases, indicating that matched model instances not necessarily represent correct solutions independent of their initial placement. Adding structural knowledge using the 2–level FEM significantly increased the recognition rate to 68%. The representation of structural attributes of the auditory folding pattern provided additional information which restricts the search space for HG to a region well–defined by SF and STS (figure 7). In all other cases the user had to select the correct solution from the priority queue. Here, with a probability of 76% the correct solution was included within the top 2% system solutions (i.e. at average at first to third position in the queue).

The structural attributes are not independent of the morphological variation of the auditory cortical folds, while the structure itself is variable. This results in a large number of variation modes mainly caused by non–linear displacements. Consequently, a 1–level FEM which includes HG (SI), SF and STS was likely to be less adequate for representing the configuration of the auditory cortical folds. We constructed such model for both classes $(\mathcal{T}_{\mathcal{X}}^{(1*)}, \mathcal{X} = \mathcal{H}, \mathcal{H}^+)$[10]. A recognition rate of 58% showed that these linear models did not provide extra knowledge compared with $\mathcal{T}_{hg}^{(1)}$. An ideal shape fit, i.e. a correct match, yielded a low overall energy $Q_{\mathcal{X}}^{(1*)}$, and vice versa. The parameters ζ and γ_i of the QOF–function had to be explicitly fine–tuned in order to increase the significance[11] of the 1–level

---

[8] To build the finite element model (sect. 4.1), stiffness ($\mathbf{K}$) and mass ($\mathbf{M}$) matrices have to be computed. In our case, $\mathbf{M}$ is a constant function of material density, ρ. $\mathbf{K}$ is a function of the material constitutive law, and is related to the material–specific Young's modulus $\mathcal{E}$ and Poisson's ratio ν.

[9] Note that scaling of $\mathcal{T}_{sf}^{(1)}$ and $\mathcal{T}_{sts}^{(1)}$ is restricted by the scaling of $\mathcal{T}_{\mathcal{X}}^{(2)}$.

[10] Here, based on the manual example segmentations, the auditory cortical folding pattern was subdivided into finite elements in a single decomposition step, yielding a set of nodes of level $l = 1$, which contribute to SF, STS, HG (and SI).

[11] We formally defined the probability of the correct solution being provided within the top 2% of the priority queue as the *significance* of a model and associated QOF.

FEM. In contrast we found that for the 2–level FEM any $\zeta < \frac{3}{10}$ decreased its significance, while it did not change for a large range of values $\zeta \in [\frac{3}{10}, \frac{7}{10}]$. This leads us to the assumption that the differentiation between the influence of model and data is less clear for the 1–level FEM compared with the hierarchy of FEM. Another example is the more adequate representation of the parallelism between SF and STS by the 2–level FEM (figs. 5a, 5b). Here, deformation of the coarsely structured global FEM due to a rotation, for example of the line–shaped model $\mathcal{T}_{sf}^{(1)}$, will enforce an equivalent rotation of the STS–model, while not affecting the rotation of the HG–model.

Another result of the high anatomical variability is that a prior model which is constructed based on a single data set might be insufficient to cover all possible variations. Similar to the construction of an ASM [8], we used known solutions
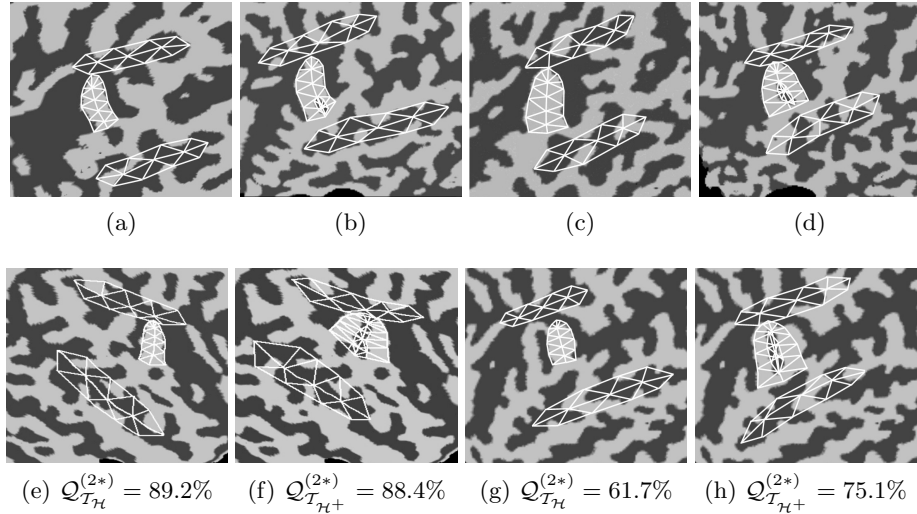

(a)   (b)   (c)   (d)

(e) $\mathcal{Q}_{\mathcal{T}_{\mathcal{H}}}^{(2*)} = 89.2\%$ (f) $\mathcal{Q}_{\mathcal{T}_{\mathcal{H}^+}}^{(2*)} = 88.4\%$ (g) $\mathcal{Q}_{\mathcal{T}_{\mathcal{H}}}^{(2*)} = 61.7\%$ (h) $\mathcal{Q}_{\mathcal{T}_{\mathcal{H}^+}}^{(2*)} = 75.1\%$

Figure 7: Results on the localization and classification of Heschl's gyrus in flat maps. Typical segmentation results using the appropriate 2–level FEM, $\mathcal{T}_{\mathcal{X}}, \mathcal{X} = \mathcal{H}, \mathcal{H}^+$, are shown in (a)–(d). The bottom row shows the best–rated instances of the two different improved class–specific prototypes $\mathcal{T}_{\mathcal{X}}^{(2*)}, \mathcal{X} = \mathcal{H}, \mathcal{H}^+$ in two different flat maps (sect. 6.1). In (e) Heschl's gyrus is correctly localized using the $\mathcal{H}$–model, while $\mathcal{T}_{\mathcal{H}^+}$ erroneously classifies two transverse temporal gyri in the same data set (f). In this example, the QOF does not allow for a positive classification. In the second example, Heschl's gyrus shows a Sulcus intermedius, and the data set was classified correctly with probability $\mathcal{Q}_{\mathcal{T}_{\mathcal{H}^+}}^{(2*)} > \mathcal{Q}_{\mathcal{T}_{\mathcal{H}}}^{(2*)}$ (h). Here, a low deformation energy (equation 4) due to the large displacement of HG and SF indicated a mismatch of the improved $\mathcal{H}$–model, $\mathcal{T}_{\mathcal{H}}^{(2*)}$ (g).

on a subset of 40 data sets to compute new global ("zero–order") prototypes $\mathcal{T}_{\mathcal{X}}^{(2*)}$ based on the samples mean nodal displacements. Both, the default and the adapted templates were then applied simultaneously for localizing Heschl's gyrus. In addition to that, given K samples x' for each transformation parameter $\xi(\mathcal{T}) \in X$ (sect. 5) and $\mathcal{T} \subseteq \mathcal{T}_{\mathcal{X}}$, we estimated the parameters of the PDF by $\widehat{\mu} = k^{-1} \sum_{s \leq K} h_s x'_s$, $\widehat{\varsigma} = (K-1)^{-1} \sum_{s \leq K} h_s (x'_s - \bar{x}')^2$, where h_s denotes frequency of occurrence of realization x'_s.

Inclusion of information from statistical analysis led to an increase of the significance of the 2–level FEM (from 76% to 98%), and in recognition rate (from 68% to 80%). As the top level of the 2–level FEM specifies linearly correlated shapes, after adapting the structural variation parameters false positive solutions could be rejected because of a lower structural QOF. For instance, with the initial global prototypes, $\mathcal{T}_{\mathcal{X}}^{(2)}$, a transverse temporal gyrus–but not the *first* Heschl's gyrus–was labeled in most cases of a failure. The displacement of the 2nd–level nodes $\mathbf{x}_j^{(2)}, j = 1$ (and $j = 1, 6$ for $\mathcal{X} = \mathcal{H}^+$, resp.) and $\mathbf{x}_2^{(2)}$ represents the spatial relation of HG (+SI) and the anterior[12] part of SF. It turned out to be much smaller and of smaller variance compared with the displacement of 2nd–level nodes $\mathbf{x}_j^{(2)}, j = 1$ (and $j = 1, 6$, resp.) and $\mathbf{x}_i^{(2)}, i = 3, 4, 5$. Consequently, large deformations of the improved prototypes $\mathcal{T}_{\mathcal{X}}^{(2*)}$ would yield a low overall energy (figs. 7g, 7h).

Our results show convergence after 8.3 ± 1.9 iterations[13] of the shape search. The total number of 2-level FEM instances initialized varied from 550 to 1280 due to the evolutionary strategy and combinatorial problems inherent to structural models[14]. However, even in the experiments with the isomorph structural prototypes, $\mathcal{T}_{\mathcal{X}}^{(2)}$ and $\mathcal{T}_{\mathcal{X}}^{(2*)}$, the search was more effective compared with the morphological FEM, $\mathcal{T}_{hg}^{(1)}$, and 1–level FEM, $\mathcal{T}_{\mathcal{X}}^{(1*)}$. (Here, the number of parallel shape searches was about two times higher.)

Classification Based on the Anatomical Configuration. The 2–level FEM was used for classification by competitively applying the class–specific models to the data. We therefore merged the priority queues after the shape searches using the two different models $\mathcal{T}_{\mathcal{X}}, \mathcal{X} = \mathcal{H}, \mathcal{H}^+$, and computed the index of the correct solution, which was known from the expert's classification.

In 90% of all cases the appropriate class–specific model $\mathcal{T}_{\mathcal{X}}$ exhibited a higher structural QOF, $Q_{\mathcal{T}_{\mathcal{X}}}^{(2)}$ (figs. 7e–7h). The difference in the QOF–values w.r.t. the correct class was significant ($p < 0.01$, one–sided t–test). In contrast, the linear

---

[12] The anterior–posterior axis corresponds to the x–axis of the flat map coordinate system, which is flipped in the right hemisphere.

[13] Expressions of this form denote mean value $\pm$ standard deviation from the mean.

[14] A complete search took 224 ± 67 seconds on a 2.4GHz P4, unoptimized Matlab/C.

model did not explicitly recognize missing discrete parts (here, the SI) of the folding pattern. In only 30% of all cases the appropriate 1–level FEM $\mathcal{T}_{\mathcal{X}}^{(1*)}$ was indicated by the respective $Q_{\mathcal{T}_{\mathcal{X}}}^{(1*)}$–values. The difference was not significant ($p > 0.4$), which shows that it did not allow for distinguishing between placement and structure error.

6.2   Case Study 2: Improved Recognition of Ants

In this section, we apply our hierarchical FEM to the automated recognition of ants in 2D color images from an ant database. As due to standardized acquisition of the images localization is less complicated compared with the first case study, here we test our hypothesis that a hierarchical representation of shape by means of vibration modes sufficiently captures variation within and between object classes, such that all ant body parts can be located even in cases of occlusion. Our particular database has been obtained from MCZ database of the Museum of Comparative Zoology at Harvard University[15] and AntWeb by the Californian Academy of Sciences[16].

Design of an "Ant Sensor". The given database images show examples of different ant species in standardized perspectives. We used 260 images from our database including the genera $\mathcal{A}$nochetus, $\mathcal{C}$erapachys and $\mathcal{P}$heidole[17]. We use the lateral views which allow for shape–based recognition in 2D, and assume that exactly one ant is displayed in each image, although parts of it may be missing. For each of the different classes ($\mathcal{A}$nochetus, $\mathcal{C}$erapachys and $\mathcal{P}$heidole) a prototypical geometric template $\mathcal{T}_{\mathcal{X}}, \mathcal{X} = \mathcal{A}, \mathcal{C}, \mathcal{P}$, is generated based on a manually segmented example image of the database. Morphological and structural aspects are again represented by class–specific 2–level FEM. Each ant shape is therefore subdivided into multiple sub–shapes $\mathcal{T}_{j}^{(1)}$, such as head, thorax, back, et cetera. The top–level models $\mathcal{T}_{\mathcal{X}}^{(2)}$ constrain structural variation for anatomical reasons. Besides from the class–specific kind and number of ant body parts, the standardized positioning of the ants, e.g. on wooden sticks, causes a curved organization of the parts, which determines the subdivision of the shape domain into the specific structural FEM. These assumptions yield coarsely structured global prototypes $\mathcal{T}_{\mathcal{X}}^{(2)}$, which account for certain anatomical variability within the ant classes (figure 8).

To set up suitable external model forces according to sect. 4.2, we use internal and contour sensors for the sub–shapes. The internal first–level nodes ideally

---

[15] `http://mcz-28168.oeb.harvard.edu/mcztypedb.htm`

[16] `http://www.antweb.org/`

[17] It was necessary to generalize among distinct genera because most databases provide only very few examples for each species, which would not allow for a validation of the proposed algorithm. Note that we used the same genera as Bergner et al. [4].
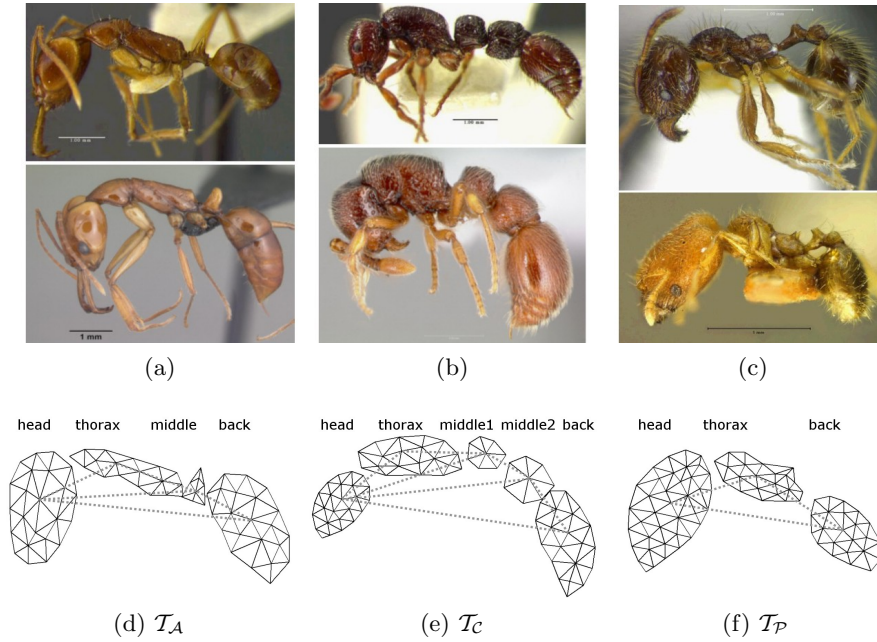
Figure 8: 2–level FEM of ant genera. Class–specific shape models for the ant genera $\mathcal{A}$nochetus (d), $\mathcal{C}$erapachys (e) and $\mathcal{P}$heidole (f). Example species are displayed in (a)–(c). (Note the high variability in shape and color of different specimen in our test images.) The number and kind of morphological components (solid lines) differ between the three prototypical structural FEM (dotted lines).

represent the typical color of the ant's body parts. However, since we generalize among distinct ant genera, no discriminant variation in body color among classes, but considerable within–class variation in color can be expected. We assume that each ant can be extracted from the background by exploiting the fact that due to the standardized acquisition most background in the images will be homogeneously gray. The internal low–level sensors are therefore mapped to "ant color" intensity images $I^{\mathcal{N}}_{\Sigma}$, which are computed by suppressing such image regions:

$$I^{\mathcal{N}}_{\Sigma} = (|I^{\mathcal{N}} - \mathbf{I}|^2) * G_{\sigma}. \tag{8}$$

Here $I^{\mathcal{N}}$ is the (triplet–wise normalized) input image I, which is a 3D matrix of values in rgb–color space, $\mathbf{I}$ is the 3D identity matrix and G_{σ} is a Gaussian low pass filter. The contour nodes of the morphological FEM represent borders between ant and background. The relating sensors are mapped to gradient magnitude maps $I^{\mathcal{N}}_{\nabla}$ of the form $|\nabla I^{\mathcal{N}}_{\sigma}|^2$ (sect. 4.2, figure 9).

(a) Input $I^{\mathcal{N}}$ (b) "Ant color" $I_{\Sigma}^{\mathcal{N}}$ (c) $I_{\nabla}^{\mathcal{N}}$ (d)
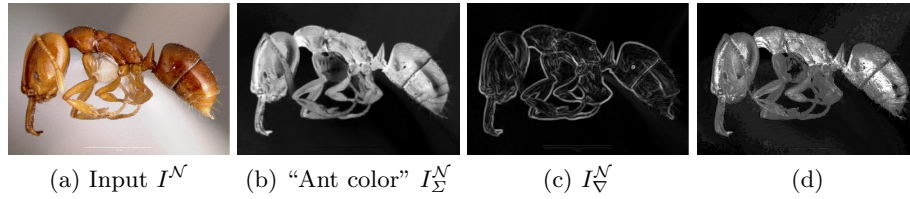
Figure 9: Low–level feature maps. The "ant color" image (b) estimates the difference in color from homogeneously gray background (eq. 8). It is used here because a statistical model of the ant color distribution is *not* available for color–based classification (d), which was used in [4].

Automatic Recognition of Ants. Because of the properties of the modeled objects, we could successfully apply the proposed multi–resolution shape search. This means that under the appropriate model $\mathcal{T}_{\mathcal{X}}, \mathcal{X} = \mathcal{A}, \mathcal{C}, \mathcal{P}$, the ant was successfully segmented in all cases of a subset of 120 test images. All experiments have been done with the same set of empirically chosen material parameters, $\mathcal{E} = 2 \cdot 10^9 [Pa], \nu = 0.4, \rho = 1.1 [\frac{g}{cm^3}]$. We used as initial settings $m(1) = 4, \sigma(1) = 10, \hat{\varsigma}(\psi) = 10°$ and $\hat{\varsigma}(\mathbf{s}') = \frac{2}{10} \mathbf{s}(\mathcal{T})$ for $\mathcal{T} = \mathcal{T}_j^{(1)}, \mathcal{T}_{\mathcal{X}}^{(2)}, \forall j, \mathcal{X} = \mathcal{A}, \mathcal{C}, \mathcal{P}$, based on the parameter values $\xi(\mathcal{T})$, e.g. known scaling $\mathbf{s}'(\mathcal{T})$, for $\mathcal{T} \subseteq \mathcal{T}_{\mathcal{X}, \mathcal{M}}$ (sect. 5). According to our working assumptions, we use the image center to set up $(\hat{\mu}(c_x), \hat{\mu}(c_y))$ for $\mathcal{T}_{\mathcal{X}}^{(2)}, \mathcal{X} = \mathcal{A}, \mathcal{C}, \mathcal{P}$, while $\hat{\varsigma}(c_x) = \hat{\varsigma}(c_y)$ is set to $\frac{1}{5}$ of the image size. We further let $\zeta = \frac{1}{2}, \gamma_i = 1, \forall \mathbf{x}_i^{(l)}, l = 1, 2$, in equations 3 and 5.

Figures 10a–10d show typical segmentations. The search converged after 4 ± 1.4 iterations. Here, the total number of model instances initialized (i.e. the number of parallel shape searches) varied from 290 to 620[18].

The proposed QOF–function according to equation 3 was sufficient for evolving and selecting the correct solution from the priority queue (sect. 5). As in the first case study (sect. 6.1), we could reproduce recognition rates of $95 - 100\%$ with the same experiment and $\zeta \in [\frac{3}{10}, \frac{7}{10}]$.

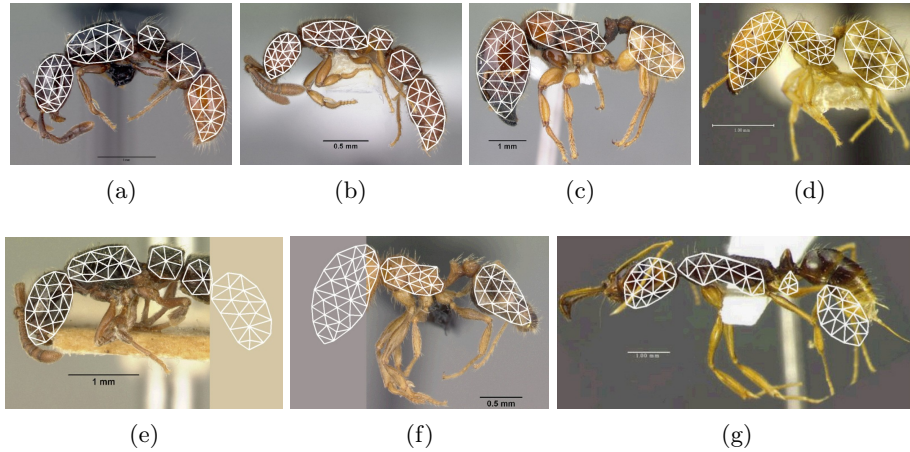[18] A complete search took 129 ± 48 seconds.

Figure 10: Results on the recognition of ants. Typical segmentation results are shown in (a)–(d). Our algorithm utilizes shape information in a hierarchical manner. This increases its robustness in cases of occlusion, such as depicted in (e) and (f). Mismatches happen very rarely. In most cases they are due to variability that cannot be captured by variation of the structural model parameters. (g) shows an example which would require all structural nodes of the $\mathcal{A}$nochetus–model (figure 8d) to be aligned horizontally.

We further analyzed the robustness of our shape search in cases of manually introduced occlusions (as depicted in figs. 10e, 10f) in a second subset of 30 images[19]. Recognition was still successful under occlusions of either part of the ant and up to 30% of the ant shape (i.e. in some cases more than one part was missing). This makes our method superior to the approach of Bergner et al. [4], as their sequential shape search requires the "head" sub–shapes to be found, and failed otherwise.

Here, our experiments were performed without any training. Results show that, given an appropriate finite element decomposition, certain variability could be covered by elastic deformation. Only in some cases, the shape constraints were not adequate to guarantee accurate segmentation. For example, the misregistration shown in figure 10g was mainly due to our assumption of curved organization of the ant body parts in the images, which is not always valid. It could

---

[19] Note that all subsets of images were chosen at random, i.e. we did not check for a clear appearance, and did not exclude examples in case our working assumptions were not met (e.g., regarding the supposed homogeneous background and curved arrangement of the body parts).

not be avoided by improving the prototypical templates $\mathcal{T}_{\mathcal{X}}^{(2)}$ based on training data, but would require inclusion of additional statistical modes of shape variation (e.g. [9], [1]), which can describe natural variations. Our experiments with improved prototypical templates and parametrization according to 6.1 showed that the number of false positive solutions due to insufficient initialization can be reduced and computation of a shape search can speed up by a factor of 2. However, the recognition rate did not improve in this application.

Model–based Classification of Ant Species. For classification we performed the multi–resolution shape search with all three class-specific models on a subset of 60 input images (we selected 20 images per class). Again, the ordered lists of solutions of our algorithm were merged for shape searches using all three models $\mathcal{T}_{\mathcal{X}}, \mathcal{X} = \mathcal{A}, \mathcal{C}, \mathcal{P}$, for the different classes. We computed the index of the correct solution in the priority queue, which was known from the database classification. In 93.3% of all cases the correct model $\mathcal{T}_{\mathcal{X}}$ exhibited a higher structural QOF–value $\mathcal{Q}_{\mathcal{T}_{\mathcal{X}}}^{(2)}$ (figure 11). The difference in the QOF–values w.r.t. the correct class was significant ($p < 0.01$, one–sided t–test). This indicates that ants can be classified using our model–based segmentation given that kind, number and

(a) $\mathcal{P}$. zelata, $\mathcal{Q}_{\mathcal{T}_{\mathcal{P}}}^{(2)} = 85.6\%$ (b) $\mathcal{A}$. madagascarensis, $\mathcal{Q}_{\mathcal{T}_{\mathcal{A}}}^{(2)} = 87.2\%$ (c) $\mathcal{C}$. gilesi, $\mathcal{Q}_{\mathcal{T}_{\mathcal{C}}}^{(2)} = 89.1\%$ (d) $\mathcal{A}$. graeffei, $\mathcal{Q}_{\mathcal{T}_{\mathcal{P}}}^{(2)} = 84.2\%$

(e) $\mathcal{Q}_{\mathcal{T}_{\mathcal{A}}}^{(2)} = 71.9\%$ (f) $\mathcal{Q}_{\mathcal{T}_{\mathcal{C}}}^{(2)} = 76.0\%$ (g) $\mathcal{Q}_{\mathcal{T}_{\mathcal{P}}}^{(2)} = 58.4\%$ (h) $\mathcal{Q}_{\mathcal{T}_{\mathcal{A}}}^{(2)} = 81.6\%$
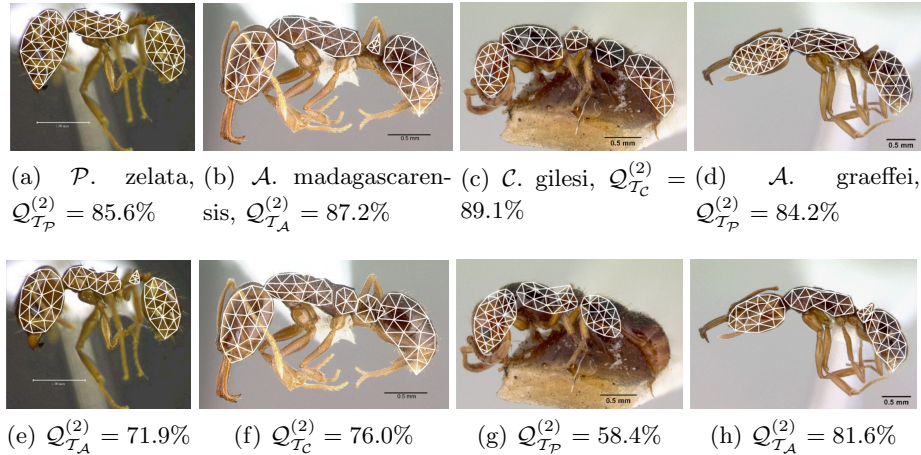
Figure 11: Classification results. The top row shows the best–rated model instances (which is for (a)–(c) an instance of the correct class–specific model $\mathcal{T}_{\mathcal{X}}$) together with the structural QOF–value $\mathcal{Q}_{\mathcal{T}_{\mathcal{X}}}^{(2)}$. The second row shows in each case the best solution with a shape model of another class. For example, $\mathcal{P}$heidole zelata was classified correctly with probability $\mathcal{Q}_{\mathcal{T}_{\mathcal{P}}}^{(2)}$ (a), which is higher than $\mathcal{Q}_{\mathcal{T}_{\mathcal{A}}}^{(2)}$ for the best match with the $\mathcal{A}$nochetus–model (e). (d) and (h) show $\mathcal{A}$nochetus graeffei, which was erroneously classified as $\mathcal{P}$heidole due to a higher $\mathcal{Q}_{\mathcal{T}_{\mathcal{P}}}^{(2)}$.

spatial configuration of the morphological components differ between classes[20]. In our case, the shape models are sufficiently different w.r.t. morphology and structure, i.e. $\mathcal{T_A}, \mathcal{T_C}$ and $\mathcal{T_P}$ consist of $3 - 5$ class–specific sub–shapes (figure 8). In some cases the difference in the QOF–values did not allow for a positive classification (figures 11d and 11h), although the correct solution was always contained within the first 12 ± 7 positions in the priority queue. Classification might improve by training a color–classifier for separating the desired object from the background. However, our assumption of gray background was already sufficient for reducing irrelevant input for the internal first–level sensors, while avoiding misclassification due to poor response to low–level feature detectors, such as reported in [4].

In a second classification experiment we used all test images and applied only the $\mathcal{P}$heidole–model $\mathcal{T_P}$, in order to select all images of ants from this genus. We computed 16.7% false positive solutions and 92.2% true positives. Bergner et al. [4] reported 84% correct classifications using a much smaller set of images which were preselected for clear appearance of exemplars from one specific genus. This shows that in direct comparison restricting shape variation to local vibration modes was adequate for this specific application, and superior to their statistical approach.

7   Discussion and Outlook

We have presented an effective approach to extract the desired complex object from an image by using a deformable shape hierarchy. For the first time it extends the ability of FEM to represent the structural decomposition of complex shapes such that the different levels of the model hierarchy describe different aspects of shape. The resulting shape description allows to separately analyze the deformation behavior of shapes and their structural relations, while the different levels of the shape hierarchy can influence each other based on their uniform representation.

We have shown that such hierarchical prototype provides useful context information for localizing complex shapes in images. Its application to the two different tasks required minimal tuning of the parameters. In both applications we used a similar set of parameters (e.g. similar FEM material properties and parameters for the shape search), which were also found to be robust. In the first case study, Heschl's gyrus was localized in flat maps of the human cortex, through the combination of global shape context (in terms of a structural model of the auditory cortical folding pattern) with local (per fold) deformable sub–shapes. Here, the top–level model estimates the fold locations, whose deformation is used for evaluating their characteristic structural configuration. Although

---

[20] Some ant types are more similar to other ants the closer they are in the family tree.

an automatic solution remains elusive because this task includes discriminating the desired object from several similar shapes within the flat maps, we demonstrated good results using the proposed focus–of–attention technique. A second application to the automatic recognition of ant species in standardized 2D color images was presented. Using a shape–structure hierarchy of two levels was shown to produce localization and classification results superior in direct comparison with a statistical structural model [4]. Detecting the head and back of an ant in isolation is a difficult problem. A back is not a back unless it is connected to a torso which is connected to a head. Unlike tree structures for use in a sequential matching, our model can capture this type of long–range contextual information, as shown in experiments with occlusions of the ant shapes.

Since our approach involves image and model features, it was possible to integrate the segmentation step into a shape search. Our experiments indicate that the hierarchical decomposition allows for different sources of shape variation, such that mislead shape searches can be analyzed in terms of placement and structure error. Despite considering class–specific shape variation, the proposed representation allows to classify shapes based on deformation parameters. Variation between classes is characterized by structural differences, while morphological variation mainly indicates variation between class members. Hence, we believe that the proposed combination of structural knowledge and deformable shapes is a very promising approach to shape search and will be of use in many applications, for example content–based image retrieval, pose estimation and tracking, medical image segmentation and model–based registration.

Assuming natural variation in shape is related to the geometry of the prototypical estimates, complex anatomical variation can be adopted by decomposition, and prior training can be avoided. This is essential, since from our experience often a rough estimate on the variation parameters will be sufficient due to the redundancy contained in model and data, while on the other hand in many cases a sufficient set of training data is not available. At present stage it is unclear how the method would deal in view of significant part articulation and deformation. For example, a sequential arrangement for aligning all shape parts in figure 2d horizontally cannot be represented with the triangular structural finite elements we used in our experiments. However, using spring/bar elements at the top levels would allow introducing such (weaker) shape constraints. On the other hand, significant "natural" deformations can be represented by statistical variation modes. Here, our shape model has the advantage that morphological and structural attributes can be optimized separately. As indicated by our results in sect. 6.1, the structural decomposition allows for a substantial improvement of the (co–)deformation parameters on the basis of small training sets.

The running time of the algorithm is largely determined by two factors. First, the time necessary to determine the optimum parametrization is bounded by the

dimensionality of the search space (which was fixed at four in our applications) and the imposed accuracy constraints. If a more precise solution is needed, the initial set of parameter values as well as the lower bound for the QOF–value τ' (sect. 5) would increase accordingly (and possibly yield an exhaustive search if a global optimum is desired). Second, the number of times the initial region of parameter values is hierarchically decomposed for the local matches before a solution is reported, is influenced by the number of hierarchy levels and number of shapes at each of the levels. Note that by using lists of known solutions repeated trials with the same set of parameters can be avoided, and the average number of trials can be reduced. All other computations are dominated by the deformable shape fit, which can be efficiently implemented, e.g. using [29], pre–computed low–level feature maps, as well as the proposed coarse–to–fine strategy.

Although we restricted our discussion to 2D shape matching by using specific aspects of the shape, we think our approach can be easily extended to 3D, and shape matching of a compound shape from different view points, respectively. The latter would require different prototypes which represent multiple 2D views [37]. Results indicate that our approach provides for hierarchically imposing top–down expectations to disambiguate interpretations at lower levels. Furthermore, the representation of different aspects of shape within a uniform framework supports the separate improvement of (almost) independent deformation parameters. We therefore believe that a "good" hierarchical prototype would facilitate the search for the desired view by providing the context needed to reliably detect the object parts in applications such as human pose estimation and tracking. Here, it will be necessary to further evaluate the adequacy of the proposed quality–of–fit measure to distinguish between false positive solutions and largely occluded object instances in cluttered scenes.

In general the number of object details, level of detail and kind of relations is not known exactly. This may restrict the applicability of (top–down) model–based methods. Since our model provides the direct comparison of different structural interpretations, i.e. categorization, we will focus on the automatic adaptation of models. One possible way around this important problem is bootstrapping, including the evaluation of different decompositions of a particular shape which vary within and across hierarchy levels. We believe that the instability problem inherent to compositional approaches might be solved by employing spectral methods [26, 47, 49] in combination with scale space approaches [30] and pairwise segmentation techniques [7, 27, 33, 34, 45] (as preprocessing step, or to improve the efficiency and accuracy of model search in an image [28, 38]).

References

1. S. Al-Zubi and K. Toennies, Extending active shape models to incorporate a–priori knowledge about structural variability, Proc. DAGM Symp (2002) 338–344
2. I. Biedermann, Human image understanding: Recent research and a theory, Comp Vis Graph Imag Proc **32** (1985) 29-73
3. T. Binford, Visual perception by computer, IEEE Conf Sys Control (1971)
4. S. Bergner and S. Al-Zubi and K. Toennies, Deformable structural models, Proc. IEEE ICIP (2004) 1875–1878
5. M. Brett, I. Johnsrude and A. Owen, The problem of functional localization in the human brain, Nature Rew Neuroscience **3** (2002) 243-255
6. D. Cristinacce and T. Cootes, Feature detection and tracking with constrained local models, Proc. BMVC (2006) 929-938
7. D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Trans Patt Anal Mach Intell **24**(5) (2002) 603–619
8. T. Cootes, et al., Active shape models their training and application, Comp Vis Imag Understand **61**(1) (1995) 38-59
9. T. Cootes and C. Taylor, Combining point distribution models with shape models based on finite–element analysis, Imag Vis Comp **13**(5) (1995) 403-409
10. T. Cootes, J. Edwards and C. Taylor, Active appearance models, Proc. ECCV (1998) 448-498
11. J. Coughlan, et al., Efficient deformable template detection and localization without user initialization, Comp Vis Imag Underst **78**(3) (2000) 303-319
12. D. DeCarlo and D. Metaxas, Shape evolution with structural and topological changes using blending, IEEE Trans Patt Anal Mach Intell **20**(11) (1998) 1186-1205
13. S. Edelman, Computational theories of object recognition, Trends in Cognitive Sciences **1** (1997) 296-304
14. K. Engel, A. Brechmann and K. Toennies, A two–level dynamic model for the representation and recognition of cortical folding patterns, Proc. IEEE ICIP (2005) 297-300
15. K. Engel, A. Brechmann and K. Toennies, Model-based segmentation of regions of interest for multi-subject analysis of fMRI data, Proc. BVM (2008) 442-447
16. K. Engel, K. Toennies and A. Brechmann, Model-based labelling of regional fMRI activations from multiple subjects, Proc. MICCAI Workshop on Analysis of Functional Medical Images (2008) 9–16
17. P. Felzenszwalb and D. Huttenlocher, Efficient matching of pictorial structures, Proc. IEEE CVPR (2000) 66-75
18. P. Felzenszwalb and D. Huttenlocher, Pictorial structures for object recognition, Int J Comp Vis **61**(1) (2003) 55-79
19. P. Felzenszwalb, Representation and detection of deformable shapes, IEEE Trans Patt Anal Mach Intell **27**(2) (2005) 208-220
20. P. Felzenszwalb, D. McAllester and D. Ramanan, A discriminatively trained, multiscale, deformable part model, Proc. IEEE CVPR (2008) 1-8
21. B. Fischl, M. Sereno and A. Dale, Cortical surface-based analysis 2: Inflation, flattening, and a surface–based coordinate system, Neuroimage **9** (1999) 195-207

22. A. Frangi, et al., Automatic construction of multiple-object three-dimensional statistical shape models, Application to cardiac modeling, IEEE Trans Med Imag **21** (2002) 1151-1166
23. G. Hamarneh, T. McInerney and D. Terzopoulos, Deformable organisms for automatic medical image analysis, Proc. MICCAI (2001) 66-76
24. A. Hill and C. Taylor, Model–based image interpretation using genetic algorithms, Imag Vis Comp **10**(5) (1992) 295-300
25. J. Hummel and I. Biedermann, Dynamic binding in a neural network for shape recognition, Psych Review **99** (1992) 480-517
26. V. Jain and H. Zhang, Robust 3d shape correspondence in the spectral domain, Proc. IEEE Int Conf Shape Mod Appl (2006) 118–129
27. J. Keuchel, M. Heiler and C. Schnoerr, Hierarchical image segmentation based on semidefinite programming, Proc. DAGM Symp (2004) 120–128
28. J. Kim and K. Hong, A new graph cut–based multiple active contour algorithm without initial contours and seed points, Mach Vis App **19** (2008) 181-193
29. S. Krinidis and I. Pitas, Fast free–vibration modal analysis of 2d physics–based deformable objects, IEEE Trans Imag Proc **14**(3) (2005) 281-293
30. T. Lindeberg, P. Lidberg and P. Roland, Analysis of brain activation patterns using a 3–d scale–space primal sketch, Human Brain Mapping (1999) 166-194
31. T. Liu and D. Geiger, Approximate tree matching and shape similarity, Proc. IEEE ICCV (1999) 456-462
32. S. Lobregt and M. Viergever, A discrete dynamic contour model, IEEE Trans Med Imag **14** (1995) 12-24
33. B. Luo, R. Wilson and E. Hancock, A spectral approach to learning structural variations in graphs, Pattern Recognition **19** (2006) 1188-1198
34. J. Malik, et al., Contour and texture analysis for image segmentation, Int J Comp Vis **43**(1) (2001) 7–27
35. C. McIntosh, G. Hamarneh and G. Mori, Human limb delineation and joint position recovery using localized boundary models, IEEE Workshop Motion Video Comp (2007) 31
36. T. McInerney and D. Terzopoulos, A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4d image analysis, Comp Med Imag Graph **19**(1) (1995) 69-83
37. G. Mori and J. Malik, Estimating human body configurations using shape context matching, Proc. ECCV (2002) 666–680
38. G. Mori, Guiding model search using segmentation, Proc. IEEE ICCV (2005) 1417-1423
39. A. Pentland, Recognition by parts, Proc. ICCV (1987) 612-620
40. A. Pentland and S. Sclaroff, Closed-form solutions to physically based shape modeling and recognition, IEEE Trans Patt Anal Mach Intell **13**(7) (1991) 715-729
41. D. Ramanan and C. Sminchisescu, Training deformable models for localization, Proc. IEEE CVPR (2006) 206-213
42. X. Ren and J. Malik, Learning a classification model for segmentation, Proc. IEEE ICCV (2003) 10-18
43. X. Ren, A. Berg and J. Malik, Recovering human body configurations using pairwise constraints between parts, Proc. IEEE ICCV (2005) 824-831

44. E. Rivlin, S. Dickinson and A. Rosenfeld, Recognition by functional parts, Comp Vis Imag Understand **62**(2) (1995) 164-176

45. C. Rohkohl and K. Engel, Efficient image segmentation using pairwise pixel similarities, Proc. DAGM Symp (2007) 254–263

46. S. Sclaroff and A. Pentland, Modal matching for correspondence and recognition, IEEE Trans Patt Anal Mach Intell **17**(6) (1995) 545-561

47. G. Scott and H. Longuet–Higgins, An algorithm for associating the features of two images, Proc. R Soc Lond **244** (1991) 21-26

48. G. Shakhnarovich, P. Viola and T. Darrell, Fast pose estimation with parameter sensitive hashing, Proc. IEEE ICCV (2003) 750-758

49. L. Shapiro and J. Brady, Feature–based correspondence: an eigenvector approach, Imag Vis Comp **10**(5) (1992) 283-288

50. K. Siddiqi and B. Kimia, Toward a shock grammar for recognition. Proc. IEEE CVPR (1996) 507-513

51. H. Tao and T. Huang, Connected vibrations: a modal analysis approach to non–rigid motion tracking, Proc. IEEE CVPR (1998) 1424-1426

52. D. Terzopoulos and D. Metaxas, Dynamic 3d models with local and global deformations, Deformable superquadrics, IEEE Trans Patt Anal Mach Intell **13**(7) (1991) 703-714

53. C. Thies, et al., Formal extraction of biomedical objects by subgraph matching in attributed hierarchical region adjacency graphs, Proc. SPIE (2004) 172-183

54. A. Tsai, et al., Mutual information for medical image segmentation in oupled multi–shape model for medical image segmentation, Med Imag Anal **8**(4) (2004) 429-445

55. M. Vaillant and C. Davatzikos, Hierarchical modeling of cortical features for deformable brain image registration, Proc. IPMI (1999) 182-195

56. Y. Wang and G. Mori, Multiple tree models for occlusion and spatial constraints in human pose estimation, Proc. ECCV (2008) 710–724

57. C.Wasserthal, et al., Automatic segmentation of the cortical grey and white matter in MRI using a region-growing approach based on anatomical knowledge, Proc. BVM (2008) 437-441

58. K. Wu and M. Levine, Segmenting 3d objects into geons, Proc. ICIAP (1995) 321-334

59. C. Xu and J. Prince, Snakes, shapes, and gradient vectorflow, IEEE Trans Imag Proc **7**(3) (1998) 359-369

60. A. Yuille, D. Cohen and P. Hallinan, Feature extraction from faces using deformable templates, Int J Comp Vis **8**(2) (1992) 99-111

61. S. Zambal, J. Hladuvka and K. Buehler, Improving segmentation of the left ventricle using a two–component statistical model, Proc. MICCAI (2006) 151-158

62. X. Zeng, L. Staib, R. Schultz and J. Duncan, Segmentation and measurement of the cortex from 3d MR images using coupled surfaces propagation, IEEE Trans Med Imag **18**(10) (1999) 100-111

63. J. Zhang, et al., Body localization in still images using hierarchical models and hybrid search, Proc. IEEE CVPR (2006) 1536–1543

64. S. Zhu and A. Yuille, Forms: A flexible object recognition and modeling system, Int J Comp Vis **20**(3) (1996) 187-212