Vision, Modeling, and Visualization (2013)
Michael Bronstein, Jean Favre, and Kai Hormann (Eds.)

Closed-Form Hierarchical Finite Element Models for Part-Based Object Detection

M. Rak[1], K. Engel[2] and K. D. Tönnies[1]

[1]University of Magdeburg, Germany
[2]Leibniz Institute for Neurobiology, Germany

Abstract

In this work we address part-based object detection under variability of part shapes and spatial relations. Our approach bases on the hierarchical finite element modeling concept of Engel and Tönnies [ET09a, ET09b]. They model object parts by elastic materials, which adapt to image structures via image-derived forces. Spatial part relations are realized through additional layers of elastic material forming an elastic hierarchy. We present a closed-form solution to this concept, reformulating the hierarchical optimization problem into the optimization of a non-hierarchical finite element model. This allows us to apply standard finite element techniques to hierarchical problems and to provide an efficient framework for part-based object detection. We demonstrate our approach at the example of lumbar column detection in magnetic resonance imaging on a data set of 49 subjects. Given a rough model initialization, our approach solved the detection problem reliably in 45 out of 49 cases, showing computation times of only a few seconds per subject.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Computer Graphics]: Scene Analysis—Object recognition; I.4.8 [Computer Graphics]: Scene Analysis—Shape

1. Introduction

Many natural and artificial objects can be decomposed into semantically meaningful parts with known spatial relations among them. Therefore, part-based object detection is an important research topic. The combination of shapes of object parts and their relations typically carries a larger amount of information compared to objects without part relations. This makes part-based object detection interesting, especially when other sources of information - as for instance intensity, edges, texture, position and orientation - are unreliable due to poor image data quality or even unavailable due to the lack of prior knowledge.

The modeling of part-based objects is widely applicable. Besides being extensively used for object detection [BAZT04, FGMR10, FH05, KTZ04, KS11, KZT09], typical fields are object segmentation [RCP*09, ZHU*11] and object tracking [LH04, RF03]. Part-based approaches were used in natural scenes [FGMR10, KTZ04], urban environments [FH05, KZT09, LH04, RF03] and medical images [RCP*09, ZHU*11], dealing with two-dimensional and three-dimensional domains. They found application in face recognition [ET09a, ET09b, FH05, KS11], person detection [FGMR10, KZT09] and tracking [LH04, RF03], as well as in animal and car detection [BAZT04, FGMR10, KTZ04].

Although part-based approaches carry more information through part relations, these relations also render the modeling task more challenging. The first issue is to combine the morphological shape information of each object part with the structural information of their relationship in a computational framework. The second even more challenging issue is the incorporation of inter-object variability of both shape and structure in such a framework.

Most of the recent work deals only with variability of part relations, completely ignoring the variability of part shapes, e.g., [FGMR10, FH05, KZT09, LH04, RCP*09, RF03]. To be clear, each of these has its justification. Still, incorporating both aspects is essential for a general purpose framework. Today, only a couple of such frameworks exist. Rendering variability, some of them draw on pure statistical models, learned on a training data set, e.g., [KS11, KTZ04]. Others, such as [BAZT04, ZHU*11], rely on mixed models, having a statistical and a parameterizable component. For mixed models, the variability of either shape or relations is represented by a parameterizable prototype, while the other is

DOI: 10.2312/PE.VMV.VMV13.137-144

still learned from training data. Yet others [ET09a, ET09b] rely on pure prototypical descriptions, where both aspects of variability are dealt with via parameterization.

A considerable deficiency of pure statistical and mixed models is the substantial effort of creating training data sets. This may be a reason why these models are almost always trained on very small data sets, containing only a few samples. Although being insufficiently trained to capture the true variability, these models still perform well in practice. This interesting fact holds true for simple (not part-decomposable) objects as well. This strong evidence provides that an exact statistical description of variability is not crucial to most problems in object detection.

Consequently, we focus on a pure prototypical description. We base on the concept of Engel and Tönnies [ET09a, ET09b], which - to our knowledge - is the only work in the field. They proposed a hierarchical layer-wise elastic modeling scheme. Therein, object parts make up the bottom-most morphological layer (ML) and part relations are superimposed by one or more structural layers (SLs) on top of the ML. Object parts in the ML as well as relations in the SLs are modeled by elastic material using finite elements. These finite element models (FEMs) are connected among layers via shared nodes, elastically arranging the likewise elastic object parts. Image information like intensity, edge and texture as well as prior knowledge like position and orientation can be integrated into the framework via forces that act on the interior or the boundary of ML and SL submodels.

This hierarchical finite element modeling (HFEM) concept provides several beneficial properties:

- separate modeling of part shape and relations
- allows for variability of part shapes and relations
- supports complex part relations via hierarchy
- intuitively parameterizable by elasticity parameters
- integrates arbitrary prior information via forces
- immediately applies to n-dimensional domains

However, the concept has drawbacks: a heuristical optimization framework, which is prone to unpredictable model movements and oscillation, and an objective function (quality-of-fit measure) that does not fit the function that is actually optimized by the framework.

We treat these issues by a closed-form solution to the HFEM concept. We show that the solution of a HFEM can be traced back to that of a simple FEM. Therefore, we are able to apply FEM optimization techniques to HFEM problems, circumventing the aforementioned issues. Our closed-form solution renders an efficient optimization framework to the powerful HFEM concept, substantially simplifying prototypical part-based modeling. Our contributions are:

- a well-defined minimization problem
- an efficient optimization framework
- guaranteed convergence/oscillation prevention
- smooth, predictable model motion.

We demonstrate our HFEM solution at the example of lumbar column detection in T_1 and T_2 weighted magnetic resonance imaging (MRI) on a data set of 49 subjects. We use a two-layer hierarchy with ML submodels of the spinal canal and lumbar vertebrae L_1–L_5, which are bridged by a bar-shaped SL submodel. Detection quality is assessed manually by an expert, answering the following questions:

- Did the model assign L_1–L_5 correctly?
- Did the canal model align with the spinal canal?
- Did the vertebra models orient correctly?

Our work is structured as follows. In Section 2, we start with a background on FEMs and their use in object detection. Afterwards, we switch to HFEMs and derive their closed-form solution in Section 3. Section 4 comprises our experiments and discussion on lumbar column detection. In Section 5, we close with a brief summary of our work, our most important results and an outline of future work.

2. Background

In FEM-based detection the object is modeled by elastic material which, over time, adapts to forces derived from external information such as image data or other prior knowledge. In this section, we give some necessary background to the linearized adaptation process and the material parameterization. Afterwards, we describe how this can be used to solve the actually non-linear adaptation problem efficiently. We further give some details on the class of conservative forces, which render FEM-based detection to be an intuitively interpretable minimization problem.

2.1. Finite Element Models

In FEM, the complex shape of an object is decomposed into elements of simpler geometry. The simplest of these finite elements are simplices, e.g., line segments in one dimension, triangles in two dimensions, tetrahedra in three dimensions, and so forth. In what follows we concentrate on this element type for the sake of simplicity. The application of our work to more complex elements is straightforward.

Given such a decomposition of the object's shape in d-dimensional space, the material at rest is given by a vector $\tilde{\mathbf{x}} \in \mathbb{R}^{d|\mathcal{N}|}$, which node-by-node holds the position of all nodes of the finite element node set $\mathcal{N}$. Representing object movement and deformation, we can draw on displacement vector $\mathbf{u}$, which relates current node positions $\mathbf{x}$ with those of the rest state $\tilde{\mathbf{x}}$ by

$$\mathbf{u} = \mathbf{x} - \tilde{\mathbf{x}}. \tag{1}$$

For small displacements $\mathbf{u}$, the body motion under influence of external forces $\mathbf{f} \in \mathbb{R}^{d|\mathcal{N}|}$ is governed by the linearized motion equation

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{D}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f}. \tag{2}$$

Matrices $\mathbf{M}$, $\mathbf{D}$ and $\mathbf{K}$ summarize the material properties inertia (mass) and damping and stiffness (elasticity), respectively. They, as well as the force vector $\mathbf{f}$, are formed by aggregation of corresponding element quantities via

$$\mathbf{M} = \sum_{e\in\mathcal{E}} \mathbf{P}_e^{\mathrm{T}} \mathbf{M}_e \mathbf{P}_e \tag{3}$$

$$\mathbf{K} = \sum_{e\in\mathcal{E}} \mathbf{P}_e^{\mathrm{T}} \mathbf{K}_e \mathbf{P}_e \tag{4}$$

$$\mathbf{f} = \sum_{e\in\mathcal{E}} \mathbf{P}_e^{\mathrm{T}} \mathbf{f}_e \tag{5}$$

over the set of finite elements $\mathcal{E}$. $\mathbf{P}_e \in \mathbb{R}^{(d+1)d\times d|\mathcal{N}|}$ is a convenient projection matrix, which transforms from global degrees of freedom (DOF) into element DOFs. For simplices this means that $\mathbf{P}_e$ transforms from $|\mathcal{N}|$ into $(d+1)$ many nodes of dimensionality d. Although the above aggregation holds true for damping matrix $\mathbf{D}$ as well, we rather post-superimpose Rayleigh's damping model

$$\mathbf{D} = \alpha\mathbf{M} + \beta\mathbf{K} \tag{6}$$

for non-negative weights α and β. Element quantities $\mathbf{M}_e$, $\mathbf{K}_e$ and $\mathbf{f}_e$ in Eq. 3-5 stem from numerical integration of the material density $\rho \in \mathbb{R}^+$, the material law and the applied loads (forces) $\mathbf{l} \in \mathbb{R}^d$ via

$$\mathbf{M}_e = \int_{\Omega_e} \mathbf{H}_e^{\mathrm{T}} \rho \mathbf{H}_e \, d\Omega_e \tag{7}$$

$$\mathbf{K}_e = \int_{\Omega_e} \mathbf{B}_e^{\mathrm{T}} \mathbf{C} \mathbf{B}_e \, d\Omega_e \tag{8}$$

$$\mathbf{f}_e = \int_{\Omega_e} \mathbf{H}_e^{\mathrm{T}} \mathbf{l} \, d\Omega_e \tag{9}$$

over the element (sub)domain Ω_e. Here, matrix $\mathbf{H}_e$ interpolates material displacement over the element via barycentric interpolators. Matrix $\mathbf{B}_e$ relates displacement to strain via (partial) spatial derivatives of these interpolators. Strain and stress are related by matrix $\mathbf{C}$, which reflects the material law (see the work of Sclaroff and Pentland [SP93] for further details on these quantities).

The influence of density and load on the model is intuitive. However, for the material law this is more complex. If we, for instance, know that shape variation differs among dimensions, we could use an orthotropic or even a fully anisotropic material law, both of which allow for dimension-dependent elasticity parameterization. For the sake of simplicity we focus on isotropic materials without strain cross-relations between dimensions. In this case, the material law in $\mathbf{C}$ is governed by a single elasticity parameter $E \in \mathbb{R}^+$, which akin to material density ρ may vary over the material.

2.2. Co-rotational Formulations

Motion equation 2 becomes invalid for displacements containing large global or local material rotation. This is due to the fact that $\mathbf{K}$ represents linearized stiffness, leading to a substantial model growth when rotation occurs. Describing the non-linear behavior, we have to update $\mathbf{K}$ as the motion takes place. Solving this issue without reintegration is vital to the efficiency of FEMs. Several co-rotational formulations were proposed to treat this issue, all of which transform the problem into an unrotated reference frame, solve the problem therein and then rotate its solution forward again.

If the rotation is mostly global, i.e., the expected local material deformation is small, the linearization in Eq. 2 is sufficient as long as the global rotation is corrected. Based on the idea of Terzopoulos and Witkin [TW88] we substitute

$$\mathbf{K}\mathbf{u} = \mathbf{R}^{\mathrm{T}} \widetilde{\mathbf{K}} (\mathbf{R}\mathbf{x} - \widetilde{\mathbf{x}}) \tag{10}$$

$$= \mathbf{K}_{\mathbf{x}}\mathbf{x} - \mathbf{K}_{\widetilde{\mathbf{x}}}\widetilde{\mathbf{x}} \tag{11}$$

$$\text{with } \mathbf{K}_{\mathbf{x}} = \mathbf{R}^{\mathrm{T}} \widetilde{\mathbf{K}} \mathbf{R} \tag{12}$$

$$\mathbf{K}_{\widetilde{\mathbf{x}}} = \mathbf{R}^{\mathrm{T}} \widetilde{\mathbf{K}}, \tag{13}$$

where Matrix $\widetilde{\mathbf{K}}$ represents linearized stiffness w.r.t. the reference frame $\widetilde{\mathbf{x}}$. Matrix $\mathbf{R} = \mathrm{diag}(\mathbf{R}', \ldots, \mathbf{R}')$ contains $|\mathcal{N}|$ copies of model rotation matrix $\mathbf{R}' \in \mathbb{R}^{d\times d}$ along its diagonal. This approach conceptually backrotates $\mathbf{x}$ via $\mathbf{R}$ and rotates its solution with $\mathbf{R}^{\mathrm{T}}$ forward again.

If the global co-rotation of Eq. 10-13 is insufficient, we can draw on local formulations based on element co-rotation as proposed by Etzmuß et al. [EKS03] and Müller and Gross [MG04]. We may also draw on node co-rotation as proposed by Müller et al. [MDM*02] and Choi and Ko [CK05]. The latter avoids re-aggregation of $\mathbf{K}$, but introduces unwanted ghost forces, cf. Etzmuß et al. [EKS03]. The former is more exact, but requires re-aggregation of $\mathbf{K}$. We focus on local element co-rotation. Following Müller and Gross [MG04], we substitute

$$\mathbf{K}\mathbf{u} = \sum_{e\in\mathcal{E}} \mathbf{P}_e^{\mathrm{T}} \mathbf{K}_e \mathbf{P}_e \mathbf{u} \tag{14}$$

$$= \sum_{e\in\mathcal{E}} \mathbf{P}_e^{\mathrm{T}} \mathbf{R}_e^{\mathrm{T}} \widetilde{\mathbf{K}}_e (\mathbf{R}_e \mathbf{P}_e \mathbf{x} - \mathbf{P}_e \widetilde{\mathbf{x}}) \tag{15}$$

$$= \mathbf{K}_{\mathbf{x}}\mathbf{x} - \mathbf{K}_{\widetilde{\mathbf{x}}}\widetilde{\mathbf{x}} \tag{16}$$

$$\text{with } \mathbf{K}_{\mathbf{x}} = \sum_{e\in\mathcal{E}} \mathbf{P}_e^{\mathrm{T}} \mathbf{R}_e^{\mathrm{T}} \widetilde{\mathbf{K}}_e \mathbf{R}_e \mathbf{P}_e \tag{17}$$

$$\mathbf{K}_{\widetilde{\mathbf{x}}} = \sum_{e\in\mathcal{E}} \mathbf{P}_e^{\mathrm{T}} \mathbf{R}_e^{\mathrm{T}} \widetilde{\mathbf{K}}_e \mathbf{P}_e. \tag{18}$$

The transformation is the same as with the global co-rotation, but on the element level. Matrix $\widetilde{\mathbf{K}}_e$ represents the linearized stiffness of element e w.r.t. its reference frame $\mathbf{P}_e\widetilde{\mathbf{x}}$. Matrix $\mathbf{R}_e = \mathrm{diag}(\mathbf{R}'_e, \ldots, \mathbf{R}'_e)$ compensates for element rotation via $d+1$ rotation matrices $\mathbf{R}'_e \in \mathbb{R}^{d\times d}$ along its diagonal - one for each node of the simplex.

In either case, Rayleigh's damping model (Eq. 6) needs to be updated during model motion by

$$\mathbf{D} = \alpha\mathbf{M} + \beta\mathbf{K}_{\mathbf{x}}. \tag{19}$$

2.3. Kabsch's Algorithm

For estimating the rotation we draw on Kabsch's algorithm [Kab76, Kab78], which works in both co-rotation cases and in arbitrary dimension. Other approaches exist, but are either limited to a particular dimension, like quaternions in Coutsias et al. [CSD04], or solely apply to element co-rotation, like in Etzmuß et al. [EKS03] or Müller and Gross [MG04].

Treating the whole model and each element, respectively as a set of nodes, we find the minimum mean squared deviation rotation via covariance of current positions and reference positions. Let $\Sigma \in \mathbb{R}^{d\times d}$ be their covariance matrix. Then, its singular value decomposition

$$\Sigma = \mathbf{V}\mathbf{S}\mathbf{W}^{\mathrm{T}} \tag{20}$$

gives the nearest orthonormal transformation by $\mathbf{V}\mathbf{W}^{\mathrm{T}}$. Additionally requiring its determinant to be $+1$, i.e., a proper rotation, we conditionally flip the sign of the left or right singular vector that corresponds to the minimum singular value of the decomposition. Assuming descending ordering of singular values in $\mathbf{S}$ the optimal rotation reads

$$\mathbf{V}\,\mathrm{diag}(\underbrace{1,\ldots,1}_{d-1},\mathrm{sign}(\det(\mathbf{V}\mathbf{W}^{\mathrm{T}})))\,\mathbf{W}^{\mathrm{T}}. \tag{21}$$

2.4. Conservative Forces

In FEM-based object detection, the concept of forces is used to link the elastic model with external information such as the image data. Since the intention is always to pull the model into the right spot, the definition of forces can be quite diverse, i.e., forces that act on the boundary of the model, forces that act on its interior or a weighted combinations of the two. This way, information ranging from nearby edges, expected intensity and texture to position and orientation of the object can be integrated by supplying appropriate forces.

We restrict ourselves to conservative forces, which originate in potential fields p that describes the likelihood of the object being at some place in the domain. This class of forces renders an easily interpretable optimization problem, because, in this case, the applied load (force)

$$\mathbf{l} = -\nabla \mathrm{p} \tag{22}$$

pulls the model towards a minima of p. For this force definition, FEM-based object detection is a damped gradient descend with inertia that minimizes a combination of internal and external energies:

$$\Pi = \Pi_{int} + \Pi_{ext} \rightarrow \min! \tag{23}$$

$$\text{with } \Pi_{int} = \tfrac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u} \tag{24}$$

$$\Pi_{ext} = \sum_{e\in\mathcal{E}} \int_{\Omega_e} \mathrm{p}\, d\Omega_e. \tag{25}$$

The impact of internal energy Π_{int} is parameterized via stiffness $\mathbf{K}$, more precisely by it's elasticity parameter E. External energy Π_{ext} stems from integration of potential over the model. Its influence is entirely controlled by the scaling of the potential field. The parameterization of mass (inertia) $\mathbf{M}$ does not influence Π, it rather allows to overcome spurios local minima of the objective function. The same holds true for damping $\mathbf{D}$, which guarantees convergence to a local minimum of Π as long as p is bounded from below. If forces cannot be traced back to gradients of some potential, external energies become dependent on the path of model motion, which does not allow such an intuitive interpretation.

3. Hierarchical Finite Element Models

In part-based object detection via HFEMs we deal with multiple submodels s and thus with multiple motion equations of form 26, which are coupled among submodels by shared nodes. Providing an efficient optimization framework is the challenge of integrating the coupled submodel motion into a simple closed-form solution.

$$\mathbf{M}^s\ddot{\mathbf{u}}^s + \mathbf{D}^s\dot{\mathbf{u}}^s + \mathbf{K}^s\mathbf{u}^s = \mathbf{f}^s \tag{26}$$

We show that the motion of a HFEM can be traced back to that of a simple FEM. This enables us to apply the FEM techniques we already discussed. This further allows us to retain valuable properties of the FEM concept:

- a well-defined minimization problem
- an efficient optimization framework
- guaranteed convergence/oscillation prevention
- smooth, predictable model motion

We start with the derivation of our closed-form solution for the case of linearized motion and then generalize it to large rotations. A time discretization scheme for our HFEM is presented afterwards.

3.1. Linearized Solution

The derivation of our closed-form solution can be outlined as follows. We first split the motion equation into equations for each submodel node. We then reformulate these equations in terms of all nodes, introducing zeros where necessary. Afterwards, we aggegrate all submodel equations for a single node following the balance of forces and rewrite the coupled equations into a single equation of the desired form. In the end, we condense the explanation into simple formulas.

If we split motion equation 26 into $|\mathcal{N}^s|$ equations - one for each node of submodel s - we get

$$\sum_{n\in\mathcal{N}^s} \mathbf{M}^s_{m,n}\ddot{\mathbf{u}}^s_n + \mathbf{D}^s_{m,n}\dot{\mathbf{u}}^s_n + \mathbf{K}^s_{m,n}\mathbf{u}^s_n = \mathbf{f}^s_m. \tag{27}$$

Matrices $\mathbf{M}^s_{m,n}$, $\mathbf{D}^s_{m,n}$, $\mathbf{K}^s_{m,n} \in \mathbb{R}^{d\times d}$ define mass, damping and stiffness of the node pair (m,n). Vectors $\ddot{\mathbf{u}}^s_n$, $\dot{\mathbf{u}}^s_n$, $\mathbf{u}^s_n \in \mathbb{R}^d$ reflect acceleration, velocity and displacement of node n, while $\mathbf{f}^s_m$ is the external force that acts on node m.

We now reformulate Eq. 27 in terms of the node set $\mathcal{N} = \bigcup_s \mathcal{N}^s$ of the hierarchical model. Zeroing new mass, damping and stiffness relations by $\forall n \notin \mathcal{N}^s : \mathbf{M}^s_{m,n}, \mathbf{D}^s_{m,n}, \mathbf{K}^s_{m,n} = \mathbf{0} \in \mathbb{R}^{d\times d}$, we can express Eq. 27 equivalently by

$$\sum_{n\in\mathcal{N}} \mathbf{M}^s_{m,n}\ddot{\mathbf{u}}_n + \mathbf{D}^s_{m,n}\dot{\mathbf{u}}_n + \mathbf{K}^s_{m,n}\mathbf{u}_n = \mathbf{f}^s_m \tag{28}$$

Due to the summation over the complete node set $\mathcal{N}$, vectors $\ddot{\mathbf{u}}_n$, $\dot{\mathbf{u}}_n$, $\mathbf{u}_n \in \mathbb{R}^d$ now reflect quantities of the HFEM.

The net force must be zero for each node of the HFEM. Therefore, all forces, i.e., inertial, damping, elastic (stretching and bending) and external, must balance each other in the sum. This immediately leads to

$$\sum_s \sum_{n\in\mathcal{N}} \mathbf{M}^s_{m,n}\ddot{\mathbf{u}}_n + \mathbf{D}^s_{m,n}\dot{\mathbf{u}}_n + \mathbf{K}^s_{m,n}\mathbf{u}_n = \sum_s \mathbf{f}^s_m \tag{29}$$

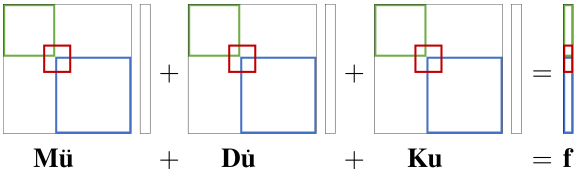


Figure 1: *Sketch of our closed-form motion equation for a hierarchy with three finite element submodels: red, green and blue; mass* $\mathbf{M}$*, damping* $\mathbf{D}$*, stiffness* $\mathbf{K}$ *and force* $\mathbf{f}$ *are aggregated where submodels share nodes*

for each $m \in \mathcal{N}$. Note that Eq. 29 is nothing more than a summation of Eq. 28 over the set of submodels.

We now introduce shorthands of form $\mathbf{M}_{m,n} = \sum_s \mathbf{M}^s_{m,n}$ for mass, damping, stiffness and force, which allows us to rewrite Eq. 29 as

$$\sum_{n\in\mathcal{N}} \mathbf{M}_{m,n}\ddot{\mathbf{u}}_n + \mathbf{D}_{m,n}\dot{\mathbf{u}}_n + \mathbf{K}_{m,n}\mathbf{u}_n = \mathbf{f}_m. \tag{30}$$

Eq. 30 follows the same form which we started with in Eq. 27, but for the complete HFEM. We can easily assemble all $|\mathcal{N}|$ equations of form 30 into a single motion equation of the desired form

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{D}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f}. \tag{31}$$

In essence, the idea of closed-form HFEM is to aggregate submodel matrices and forces as sketched in Figure 1. Therefore, we have to set up suitable projection matrices $\mathbf{P}^s \in \mathbb{R}^{|\mathcal{N}^s|d\times d|\mathcal{N}|}$, which transform from HFEM DOFs into submodel DOFs. We then end up with HFEM system matrices and forces formed via

$$\mathbf{M} = \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{M}^s\mathbf{P}^s \tag{32}$$
$$\mathbf{D} = \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{D}^s\mathbf{P}^s \tag{33}$$
$$\mathbf{K} = \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{K}^s\mathbf{P}^s \tag{34}$$
$$\mathbf{f} = \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{f}^s. \tag{35}$$

Regarding Rayleigh's damping model, the aggregation in Eq. 33 yields the same as the post-superimposition, since

$$\mathbf{D} = \sum_s \mathbf{P}^{s\mathrm{T}}\left(\alpha\mathbf{M}^s + \beta\mathbf{K}^s\right)\mathbf{P}^s \tag{36}$$
$$= \alpha\sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{M}^s\mathbf{P}^s + \beta\sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{K}^s\mathbf{P}^s \tag{37}$$
$$= \alpha\mathbf{M} + \beta\mathbf{K}. \tag{38}$$

3.2. Co-rotational Formulation

The generalization of our closed-form solution to large rotation is straightforward, because both co-rotational formulations that we already discussed apply to our HFEM immediately. Still, there may be cases where neither global nor element co-rotation best suits the problem. This is the case, e.g., if the expected rotation is mostly global except for one or more freely "hanging" submodels that can rotate independently. Global co-rotation is then insufficient to solve the problem and element co-rotation may be overkill.

To represent these cases, we introduce a formulation for submodel co-rotation. The idea is the same as with global co-rotation, but applied to each submodel individually. We propose to substitute

$$\mathbf{K}\mathbf{u} = \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{K}^s\mathbf{P}^s\mathbf{u} \tag{39}$$
$$= \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{R}^{s\mathrm{T}}\widetilde{\mathbf{K}}^s\left(\mathbf{R}^s\mathbf{P}^s\mathbf{x} - \mathbf{P}^s\widetilde{\mathbf{x}}\right) \tag{40}$$
$$= \mathbf{K}_{\mathbf{x}}\mathbf{x} - \mathbf{K}_{\widetilde{\mathbf{x}}}\widetilde{\mathbf{x}} \tag{41}$$
$$\text{with } \mathbf{K}_{\mathbf{x}} = \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{R}^{s\mathrm{T}}\widetilde{\mathbf{K}}^s\mathbf{R}^s\mathbf{P}^s \tag{42}$$
$$\mathbf{K}_{\widetilde{\mathbf{x}}} = \sum_s \mathbf{P}^{s\mathrm{T}}\mathbf{R}^{s\mathrm{T}}\widetilde{\mathbf{K}}^s\mathbf{P}^s, \tag{43}$$

where matrix $\widetilde{\mathbf{K}}^s$ represents the linearized stiffness of submodel s w.r.t. its reference frame $\mathbf{P}^s\widetilde{\mathbf{x}}$. Matrix $\mathbf{R}^s = \mathrm{diag}(\mathbf{R}^{s\prime}, \ldots, \mathbf{R}^{s\prime})$ contains $|\mathcal{N}^s|$ copies of rotation matrix $\mathbf{R}^{s\prime} \in \mathbb{R}^{d\times d}$ along its diagonal, which compensates node-wise for the entire submodel rotation. Akin to global and element co-rotation, Kabsch's algorithm is used to estimate rotations $\mathbf{R}^{s\prime} \in \mathbb{R}^{d\times d}$.

A combination of co-rotational techniques is possible as well. For instance, if the expected shape variation requires element co-rotation for some submodels while submodel co-rotation is sufficient for others, we can easily derive an adapted co-rotation formulation. We then replace $\mathbf{R}^{s\mathrm{T}}\widetilde{\mathbf{K}}^s\mathbf{R}^s$ in Eq. 42 and $\mathbf{R}^{s\mathrm{T}}\widetilde{\mathbf{K}}^s$ in Eq. 43 by element co-rotation matrices of Eq. 17 and Eq. 18, respectively. Thus, our approach allows for a trade-off of accuracy and performance according to prior knowledge.

3.3. Time discretization

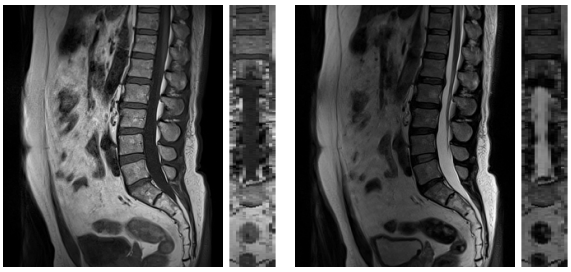
We solve the closed-form HFEM motion equation by Newmark's implicit second order scheme, which reads

$$\left(\mathbf{M} + \gamma_1\Delta t\mathbf{D} + \gamma_2\Delta t^2\mathbf{K}\right)\ddot{\mathbf{u}}_{t+\Delta t} = \mathbf{f}_{t+\Delta t} - \mathbf{K}\mathbf{u}_t - \left(\mathbf{D} + \Delta t\mathbf{K}\right)\dot{\mathbf{u}}_t - \left(\left(1-\gamma_1\right)\Delta t\mathbf{D} + \left(\tfrac{1}{2} - \gamma_2\right)\Delta t^2\mathbf{K}\right)\ddot{\mathbf{u}}_t \tag{44}$$
$$\dot{\mathbf{u}}_{t+\Delta t} = \dot{\mathbf{u}}_t + \left(1-\gamma_1\right)\Delta t\ddot{\mathbf{u}}_t + \gamma_1\Delta t\ddot{\mathbf{u}}_{t+\Delta t} \tag{45}$$
$$\mathbf{u}_{t+\Delta t} = \mathbf{u}_t + \Delta t\dot{\mathbf{u}}_t + \left(\tfrac{1}{2} - \gamma_2\right)\Delta t^2\ddot{\mathbf{u}}_t + \gamma_2\Delta t^2\ddot{\mathbf{u}}_{t+\Delta t} \tag{46}$$

For the average constant acceleration parameterization with $\gamma_1 = \frac{1}{2}$ and $\gamma_2 = \frac{1}{4}$, the scheme is unconditionally stable and most accurate. Substituting either of the co-rotation formulations into Eq. 44 gives

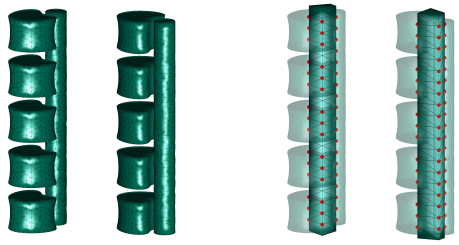
$$\left(\mathbf{M} + \gamma_1\Delta t\mathbf{D} + \gamma_2\Delta t^2\mathbf{K}_{\mathbf{x}}\right)\ddot{\mathbf{u}}_{t+\Delta t} = \mathbf{f}_{t+\Delta t} - \mathbf{K}_{\mathbf{x}}\mathbf{x} + \mathbf{K}_{\widetilde{\mathbf{x}}}\widetilde{\mathbf{x}} - \left(\mathbf{D} + \Delta t\mathbf{K}_{\widetilde{\mathbf{x}}}\right)\dot{\mathbf{u}}_t - \left(\left(1-\gamma_1\right)\Delta t\mathbf{D} + \left(\tfrac{1}{2} - \gamma_2\right)\Delta t^2\mathbf{K}_{\mathbf{x}}\right)\ddot{\mathbf{u}}_t. \tag{47}$$

We additionally substitute damping in Eq. 47 by Rayleigh's damping model $\mathbf{D} = \alpha\mathbf{M} + \beta\mathbf{K}_{\mathbf{x}}$, although other damping models are possible as well.



(a) T_1 (b) T_2

Figure 2: *Aligned abdominal magnetic resonance images of a subject from our data set; (a) central sagittal and coronal plane of the T_1 weighted image; (b) central sagittal and coronal plane of the T_2 weighted image*



(a) Morphological layer (b) Structural layer

Figure 3: *Front and back view of our two-layer lumbar column model; (a) morphological layer with part models for spinal canal and each vertebra body; (b) structural layer with a bar-shaped model enforcing part relations via shared nodes (red) with each of the part models*

If we have little to no spurious energy minima, we may not need the influence of inertia (mass) to overcome those minima. In this case we only solve $\mathbf{D}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f}$. Although an implicit first order formulation of Newmark's scheme can be derived, we can still use the second order scheme of Eq. 44-46, simply removing the mass $\mathbf{M}$ from its left-hand side.

4. Experiments and Discussion

We demonstrate our HFEM approach to object detection at the example of lumbar column detection in MRI. We first go into detail on the image data and introduce our HFEM lumbar column model. We further describe the particular potential fields that link the model to the image data as well as the parameterization of both potential fields and column model. Afterwards, we detail the test scenario and discuss its results.

4.1. Data and Model

The MRI data was acquired in the Study of Health in Pomerania seeking to investigate population risk factors for various disorders and diseases, cf. Völzke et al. [VAS*11]. Our data comprises aligned T_1 and T_2 weighted images of 49 healthy subjects, which were acquired on a Siemens 1.5 Tesla Magnetom Avanto imager. See Figure 2 for an example. During the standardized acquisition procedure, an abdominal volume of $500 \times 500 \times 66$ mm was sliced sagittally into 15 slices at a resolution of 448×448 pixels, resulting into voxels of about $1.1 \times 1.1 \times 4.4$ mm in size (see the work of Hegenscheid et al. [HKV*09] for further details on acquisition). For convenience, we rescaled image intensities equally into the $[0,1]$-interval.

Our lumbar column model is shown in Figure 3. It consists of a cylindric spinal canal submodel and five vertebra body submodels. Physical extents of the submodels as well as of gaps between submodels were defined according to the data set expectation. We use a bar structure to capture the columnar relations elastically. Layers of the model are coupled by shared nodes, which are marked in red in Figure 3.

4.2. Potentials and Parameters

To link our lumbar column model to the image data we derive two potential fields, which describe the likelihood of the lumbar column being at some place in the domain. One field p_c acts on the canal submodel and the other field p_v acts on the vertebra submodels. The fields are weighted, point-wise combinations of the T_1 and the T_2 weighted image:

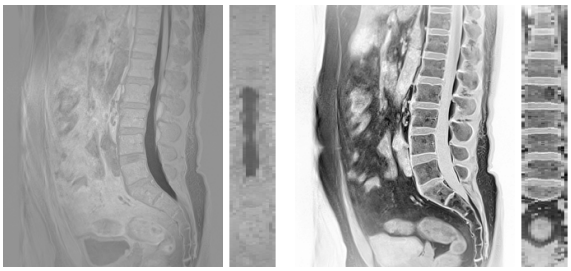
$$\mathrm{p}_c = s_c G_{\sigma_c} * ((1-w_c)\,T_1 + w_c\,(-T_2)) \quad (48)$$

$$\mathrm{p}_v = s_v G_{\sigma_v} * ((1-w_v)\,(-T_1) + w_v T_2)\,. \quad (49)$$

These potentials exploit the different intensity relations of column parts with their adjacent structures in T_1 and T_2 weighted images. Absolute information such as expected intensities may improve the discriminative power of the potential fields, which however is beyond the scope of our work.

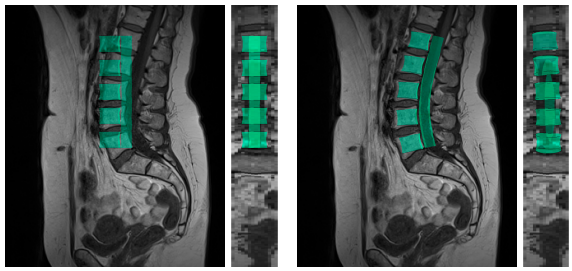
In Eq. 48 and Eq. 49, weights $w_c \in [0,1]$ and $w_v \in [0,1]$ have to be defined so that the combination results in well-defined minima along the canal and inside each vertebra body, respectively. Figure 4 shows weighted combinations for $w_c = \frac{1}{2}$ and $w_v = 0$, which we found to suit that goal best. To extend the influence region of these minima we introduced smoothing by Gaussian kernels G_{σ_c} and G_{σ_v}, respectively. Here, we aim for large influence regions without blurring the minima themselves or intermediate structures as for example the intervertebral disks. We found standard deviations of $\sigma_c = 10$ mm and $\sigma_v = 4$ mm to be appropriate.

Scaling factors s_c and s_v in Eq. 48 and Eq. 49 along with the model elasticity balance the minimization problem. For simplicity we assume isoparametric material, i.e., material parameters are constant over each submodel and all submodels have the same parameterization. For this simplification, elasticity $E = 1$ combined with scaling factors $s_c = \frac{3}{2}$ and $s_v = \frac{1}{4}$ provide good results and an almost stationary behavior for a wide parameter range around them. We fixed material density at $\rho = \frac{1}{10}$ and parameterized Rayleigh's damping by $\alpha, \beta = 1$. We used global co-rotation in our experiments, because the expected non-linear variation is rather small.



(a) $w_c = \frac{1}{2}$ (b) $w_v = 0$

Figure 4: *Weighted abdominal image combinations of a subject from our data set; (a) minima along the spinal canal; (b) local minimum inside each vertebra body*



(a) Initial Position (b) Final Position

Figure 5: *Our test scenario for a subject from our data set; (a) model position after initialization; (b) final model position after optimization*

4.3. Scenario and Results

We performed our experiment in three steps: manual model initialization, automatic model optimization, and manual result quality assessment.

During the first step, an expert manually specified a point inside the L_3 vertebra body within the central sagittal T_1 slice for each subject. We used this point in the second step to define the initial model position. More precisely, we aligned the center of the L_3 submodel with it. Model orientation was kept constant due to standardized acquisition. As shown in Figure 5, we always initially oriented the model in consent with the image coordinate system. In the third step, an expert screened the results and assessed detection quality, answering the following questions by "Yes" or "No".

1. Did the model assign L_1–L_5 correctly?
2. Did the canal model align with the spinal canal?
3. Did the vertebra models orient correctly?

We chose the combination of a large data set and a rather simple evaluation because it enables us to judge detection quality and robustness to object variability quickly.

Using a straightforward Matlab implementation on a standard 64-bit system with an Intel Core i5 4×3.30 GHz, the optimization took 1.52 s on subject average with a standard deviation of 0.39 s. Minimal and maximal computation times were 1.08 s and 2.55 s, respectively.

Table 1: *Aggregated results for our test scenario*

Answer	Question		
	1	2	3
Yes	48	46	45
No	1	3	4

Aggregated results of our evaluation are given in Table 1. Answers indicate good overall detection quality in 45 out of 49 subjects. The model completely failed only in one case, misassigning the L_1–L_5 vertebrae and misaligning with the spinal canal. Thus, answers to all Questions were "No". In two cases the model did not match the spinal canal and the vertebra orientation, hence answers to Question 2 and 3 were "No". The vertebra orientation was incorrect in one other case, resulting in a "No" for Question 3.

In the first three cases we found that the lumbar column was not entirely covered by the imaged volume. More precisely, portions of one or more vertebra bodies were not covered by the sagittal slices that were acquired. This was not the case for the rest of the data set. Possibly the energy-neutral Neumann boundary condition that we used to deal with the image border may be insufficient and hence these bad results may be avoidable using other boundary conditions. In the last case, i.e., where only the vertebra orientation was incorrect, the subject was substantially larger than the data set expectation. The model was not able to compensate for this deviation, leading to false orientation of the L_1 and L_5 vertebra. We may be able to compensate for that by relaxation of the model's resistance to variation in length.

5. Conclusion

In this work we focussed on prototypical part-based modeling for objection detection. Possible applications of our approach are face recognition, person detection and tracking as well as several medical problems. Our approach may also be of interest for fields like object tracking and segmentation.

We based on the work of Engel and Tönnies [ET09a, ET09b] who proposed a hierarchical finite element approach that allows for elastic modeling of variable parts with variable part relations. We substantially simplified their heuristical optimization framework by means of an efficiently solvable closed-form solution to their hierarchical concept. We showed how to trace the solution of a hierarchical finite element model back to that of a non-hierarchical one. This allowed us to apply standard finite element optimization techniques to hierarchical problems. Thus, we augmented hierarchical part-based object detection with an efficient local optimization framework.

We demonstrated our approach at the example of lumbar column detection in magnetic resonance imaging on a data set of 49 subjects. We used a hierarchy containing submodels for the spinal canal and each vertebra. We superimposed a bar structure on top of these submodels to realize their variable spatial relationship. Given a rough model initilization, our approach solved the detection problem reliably in 45 out of 49 cases, showing computation times of only a few seconds per subject.

In the future, we plan to extend our approach by a stochastical initialization component. The goal is to remove the need for manual initialization. In a second step we aim to provide the user with several local optima, letting her/him choose the correct solution. The intention is to design a feedback process that allows us to learn model parameters in the course of detection. Starting with a very elastic model, we may use the user's choice to optimize model parameters such that the model becomes less elastic in spots where the expected variation of part shapes or their relationship is low.

Regarding the lumbar column, we plan to use our detection results as shape prior for a suitable segmentation technique like level sets or graph cuts. Combined with our approach this may give a one click solution to exact lumbar column segmentation. We also plan to investigate the detection and segmentation of other important parts of the spine, such as the cervical column.

Acknowledgements

We thank all parties and participants of the Study of Health in Pomerania for providing us with such a comprehensive image data set. This research was partially funded by the project Visual Analytics in Public Health (TO 166/13-2), which is part of the Priority Program 1335: Scalable Visual Analytics of the German Research Foundation.

References

[BAZT04] BERGNER S., AL-ZUBI S., TÖNNIES K. D.: Deformable structural models. In *Proc. IEEE Int. Conf. Image Process.* (2004), pp. 1875–1878. 1

[CK05] CHOI M. G., KO H. S.: Modal warping: Real-time simulation of large rotational deformation and manipulation. *IEEE Trans. Vis. Comput. Graphics 11* (2005), 91–101. 3

[CSD04] COUTSIAS E. A., SEOK C., DILL K. A.: Using quaternions to calculate RMSD. *J. Comput. Chem. 25* (2004), 1849–1857. 3

[EKS03] ETZMUSS O., KECKEISEN M., STRASSER W.: A fast finite element solution for cloth modelling. In *Proc. Pac. Conf. Comput. Graphics Appl.* (2003), pp. 244–251. 3

[ET09a] ENGEL K., TÖNNIES K. D.: Hierarchical vibrations: A structural decomposition approach for image analysis. In *Proc. Int. Conf. Energy Minim. Methods Comput. Vision Pattern Recognit.* (2009), pp. 317–330. 1, 2, 7

[ET09b] ENGEL K., TÖNNIES K. D.: Stable structural deformations. In *Proc. Int. Conf. Comput. Vision Syst.* (2009), pp. 114–123. 1, 2, 7

[FGMR10] FELZENSZWALB P. F., GIRSHICK R. B., MCALLESTER D., RAMANAN D.: Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell. 32* (2010), 1627–1645. 1

[FH05] FELZENSZWALB P. F., HUTTENLOCHER D. P.: Pictorial structures for object recognition. *Int. J. Comput. Vision 61* (2005), 55–79. 1

[HKV*09] HEGENSCHEID K., KÜHN J. P., VÖLZKE H., BIFFAR R., HOSTEN N., PULS R.: Whole-body magnetic resonance imaging of healthy volunteers: Pilot study results from the population-based SHIP study. *RöFo - Fortschr. Geb. Röntgenstr. bildgeb. Verfahr. 181* (2009), 748–759. 6

[Kab76] KABSCH W.: A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A: Found. Crystallogr. 32* (1976), 922–923. 3

[Kab78] KABSCH W.: A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A: Found. Crystallogr. 34* (1978), 827–828. 3

[KS11] KAZEMI V., SULLIVAN J.: Face alignment with part-based modeling. In *Proc. Br. Mach. Vision Conf.* (2011), pp. 27.1–27.10. 1

[KTZ04] KUMAR P. M., TORR P. H. S., ZISSERMAN A.: Extending pictorial structures for object recognition. In *Proc. Br. Mach. Vision Conf.* (2004). 1

[KZT09] KUMAR P. M., ZISSERMAN A., TORR P. H. S.: Efficient discriminative learning of parts-based models. In *Proc. IEEE Int. Conf. Comput. Vision* (2009), pp. 552–559. 1

[LH04] LAN X., HUTTENLOCHER D. P.: A unified spatio-temporal articulated model for tracking. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2004), pp. 722–729. 1

[MDM*02] MÜLLER M., DORSEY J., MCMILLAN L., JAGNOW R., CUTLER B.: Stable real-time deformations. In *Proc. ACM SIGGRAPH/Eurographics Symp. Comput. Anim.* (2002), pp. 49–54. 3

[MG04] MÜLLER M., GROSS M.: Interactive virtual materials. In *Proc. Graphics Interface* (2004), pp. 239–246. 3

[RCP*09] ROBERTS M. G., COOTES T. F., PACHECO E., OH T., ADAMS J. E.: Segmentation of lumbar vertebrae using part-based graphs and active appearance models. In *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* (2009), pp. 1017–1024. 1

[RF03] RAMANAN D., FORSYTH D.: Finding and tracking people from the bottom up. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2003), pp. 467–474. 1

[SP93] SCLAROFF S. E., PENTLAND A. P.: *Modal matching for correspondence and recognition*. Tech. Rep. 201, Massachusetts Inst. Tech., Media Lab. Percept. Comput. Sect., 1993. 3

[TW88] TERZOPOULOS D., WITKIN A.: Physically based models with rigid and deformable components. *IEEE Comput. Graphics Appl. 8* (1988), 41–51. 3

[VAS*11] VÖLZKE H., ALTE D., SCHMIDT C. O., RADKE D., LORBEER R., FRIEDRICH N., AUMANN N., LAU K., PIONTEK M., BORN G., HAVEMANN C., ITTERMANN T., SCHIPF S., HARING R., ET AL.: Cohort profile: The Study of Health in Pomerania. *Int. J. Epidemiol. 40* (2011), 294–307. 6

[ZHU*11] ZHANG S., HUANG J., UZUNBAS M., SHEN T., DELIS F., HUANG X., VOLKOW N., THANOS P., METAXAS D. N.: 3d segmentation of rodent brain structures using hierarchical shape priors and deformable models. In *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* (2011), pp. 611–618. 1