A Quality–of–fit Function for Evaluating Deformable Model–based Segmentations of Anatomical Structures in Medical Data

Karin Engel and Klaus Toennies

Otto von Guericke University Magdeburg, Germany

Abstract. Localising shapes in medical images is an important, but difficult problem due to the tremendous anatomical variability and the variation in image quality. Hence, segmentation often involves variants of deformable models that can handle part of these problems by considering certain shape constraints and image features over the actual segmentation. Quality of the segmentations obtained by deformable shape models is rarely evaluated, but may be specifically useful for automating the segmentation process. This paper presents a quality–of–fit function for the evaluation of segmentations obtained by parametric deformable models. It combines a measure of non–rigid shape deformation with a measure that estimates the external energy of the model based on application–specific external forces. Experimental results are presented with data from different medical applications and show that the proposed function allows determining success or failure of the model fit.

1 Introduction

Segmentation of specific objects in medical images is a difficult task because medical data often come noisy and incomplete, while the anatomical structures of interest may vary dramatically from patient to patient and from scan to scan. One possible solution to this problem is to introduce smoothness constraints to the extracted boundaries and prior information about the shape and possible deformations of the desired object. This kind of a-priori knowledge can be represented by deformable models, such as the statistical Active Shape Model (ASM) [1, 2], or variants of the energy–minimising Active Contour Model [3–7].

Under the deformable model framework, segmentation is considered as the problem of finding the contour that represents the smooth boundary of a specific object Ω in the image I. A deformable shape model $\mathcal{T}$ represents additional information about the objects shape and a set of constraints that control the preference of the template to deform into similar shapes. Success of deformable models crucially depends on both, reliable external and internal energies (that drive the model fit and constrain the deformation, respectively), as well as on the initial estimate (i.e. information about the location, size and orientation of the desired object in an image). Currently human guidance, e.g. in terms of initialisation and generation of training data, is often needed to obtain correct segmentation results [8–10]. On the other hand, the quality of the segmentations obtained by deformable shape models is rarely evaluated, but may be employed for controlling the segmentation process. Hill and Taylor proposed an image–based objective function that compares the model–based segmentation $I(\mathcal{T})$ with the expected image data [11]. This measure may not always be appropriate, for example for the segmentation of data obtained using non–deterministic image modalities, such as sonography or SPECT. Moreover, the image–based objective function does not provide information about the shape deformation, which may be useful, e.g. for the monitoring of diseases, or comparative studies. Bergner et al. recently proposed a deformation measure for the evaluation of mass–spring–models of shape [12]. The sensitivity to the scaling of the model instances represents, however, a drawback of their measure, because the size of an object depends on imaging parameters (such as the selected field of view). To address these issues, we present a quality–of–fit (QOF) function, which for the first time not only considers the support provided by image features, but includes also the amount of non–rigid deformation necessary to match the data.

2 Method

Our method builds on the assumption that a prototypical representation of shape sufficiently captures variation of the desired object class. We therefore exemplarily employ the finite element decomposition of shape, which supports an efficient simulation of deformation [8]. The quality of such a model instance projected into the image domain and deformed according to external, image–based model forces can be evaluated in a straightforward manner. For using it in medical image interpretation, the quality–of–fit function is designed such that it indicates the probability that the segmentation results are correct.

2.1 Prototypical Parametric Deformable Models

A parametric deformable template $\mathcal{T}(\mathbf{p})$ represents the objects undeformed shape (rest shape) and a set of parameters $\mathbf{p}$ that define how it deforms under applied forces (cf. Sect. 2.2). The rest shape of a n–dimensional object can be understood as a continuous domain $\Omega \subset \mathbb{R}^n$, and its deformation is described by a boundary value partial differential equation that is solved numerically for the unknown displacement field $u(\mathbf{x})$, $\mathbf{x} \in \Omega$. The Finite Element Method yields an algebraic function that relates the deformed positions $\mathbf{x}$ of all finite element nodes in the object to the internal elastic forces,

$$\mathbf{f}_{int} = \mathbf{K}(\mathbf{x} - \mathbf{x}^0) = \mathbf{K}\mathbf{u}. \tag{1}$$

Here, $\mathbf{x}^0$ represents the rest positions of the nodes, and $\mathbf{K}(E, \nu)$ encapsulates the material (stiffness) properties as well as the type of mesh and discretisation used.

When the elastic template is subject to a dynamic load $\mathbf{f}(t)$, the displacement field varies over time, and the elastic forces $\mathbf{f}_{int}$ are replaced by the sum of body forces, inertial and damping forces and external forces. The dynamic equilibrium equation has the form

$$\ddot{\mathbf{u}}(t) = \mathbf{M}^{-1}(-\mathbf{C}\dot{\mathbf{u}}(t) - \mathbf{K}\mathbf{u}(t) + \mathbf{f}(t)), \tag{2}$$

where $\ddot{\mathbf{u}}(t) \approx \frac{\partial^2 \mathbf{u}}{\partial t^2}|_{t>0}$ and $\dot{\mathbf{u}}(t) \approx \frac{\partial \mathbf{u}}{\partial t}|_{t>0}$. To simplify analysis, $\mathbf{C}$ may approximate a velocity–dependent damping force, and $\mathbf{M}$ may represent a constant function of material density, ρ [4, 5]. For simulating the deformation of the template, the finite element equations (2) are integrated over time until an equilibrium is reached. (Note that the linear elasticity model that is commonly used in real–time animations of deformable objects is not invariant to rigid–body transformations. Artefacts can be avoided using the stiffness warping method [5, 13].)

Similar to the ASM [1], the deformed positions $\mathbf{x}(t) = \mathbf{x}^0 + \mathbf{u}(t)$ can be expressed in terms of a superposition of $m = m_2 - m_1$ displacement fields,

$$\mathbf{x}(t) = \mathbf{x}^0 + \sum_{k=m_1}^{m_2} \phi_k \mathbf{q}_k(t), \tag{3}$$

where $m_1 \geq 1, m_2 \leq nN$. The modal vectors ϕ_k are solutions to the eigenproblem $(\mathbf{K} - \omega_k^2 \mathbf{M})\phi_k = 0$ and $\mathbf{q}_k$ contains the nodal coordinates in embedded space (modal amplitudes) [5].

2.2 External Model Forces

In our case, the external model forces $\mathbf{f}(t)$ in Equation 2 shall attract the nodes to characteristic object features in the image I. Such dynamic loads are created by sensors at the finite element nodes. The sensors usually sample a scalar potential field $\mathcal{P}$, whose local minima coincide with features of interest, such that

$$\mathbf{f}(t) = -\nabla \mathcal{P}(\mathbf{x}). \tag{4}$$

For example, the contour nodes typically represent object boundaries and are therefore associated with sensors that sample gradient magnitude maps, i.e. $\mathcal{P} = -\kappa|\nabla(G_\sigma * I)|^2$, where G_σ denotes a Gaussian low pass filter with standard deviation σ, and $\kappa > 0$ [6]. In order to render the influence of different object features (such as edges and corners) independent from the specific filters used, the feature values can be normalised into the interval $[0, 1]$, e.g. by a linear transformation of the filtered image. The potential energy function may also be defined based on the distance from the current nodal positions to points in the image at which, e.g., the filter response is high. Using a map $\mathcal{D}(\mathbf{x})$ of the distance of each pixel to the closest relevant point,

$$\mathbf{f}_i(t) = \kappa \mathcal{D}(\mathbf{x}_i(t)), \qquad \kappa > 0, \tag{5}$$

where $\mathbf{x}_i(t)$ denotes the current position of sensor $i = 1, \ldots, N$. In contrast to the energy minimising formulation of deformable models, our formulation also permits the use of more general types of external forces, which may not be written as the negative gradient of a scalar potential field. For example, such forces may be defined by employing region statistics over the current segmentation [14].

2.3 The Quality–of–fit Function

Based on an initial estimate (obtained by a transformation of the model to the image coordinate system), the model is deformed by minimising its associated energy function. This criterion describes the model instances energy w.r.t. the undeformed reference shape and image data. In our case, a locally optimum image segmentation result is indicated by a

balance of the internal and external model forces (cf. Equation 2). The associated quality–of–fit (QOF) value should be high, if the image data explained by the model instance, $I(\mathcal{T}(\mathbf{p}^t))$, corresponds with the expected image data without significant deformation from the reference shape $\mathcal{T}(\mathbf{p}^0)$. A maximum QOF–value shall indicate cases, where model and data match perfectly without non–rigid deformation. From our experience, three types of error can result in low segmentation quality. First, the model may not be adequate (w.r.t. its shape constraints, elasticity model, or response to the crucial feature detectors), such that it produces false positive solutions (target error). Second, the model may be badly initialised (projection error). And third, the evaluation measure may be inappropriate (evaluation error). In order to reflect these types of error in an adequate manner, our objective function incorporates both internal and external energies. More specifically, the QOF–function combines a measure of deformation Q^d of a model instance with its correlation with the image data Q^s for estimating overall energy,

$$\mathcal{Q} = \zeta Q^d + (1 - \zeta) Q^s, \tag{6}$$

where $\zeta \in [0, 1]$ is a regularisation parameter. In order to ease interpretation we aim at formulating Q^s and Q^d such that their values are normalised to the interval $[0, 1]$, where values close to 1 indicate high quality.

Since parametric models perform a discrete data sampling at the node positions, the mean value of the sensor input (e.g., the average gradient magnitude at the deformed nodal positions) may be used to estimate the external energy Q^s of a model instance [11, 12]. This simple approach may, however, not always be appropriate (for example, if the sensor inputs do not stem from a scalar Gaussian potential field, but represent texture information [2, 14, 15]). We therefore directly employ the force formulation of deformable models, and let

$$Q^s = \mathcal{F}\big(\mu(|\mathbf{f}_i(t)|^2)\big) \tag{7}$$

indicate the amount of correspondence of the deformed model with the data. In Equation 7, the function μ computes the mean value, and $\mathcal{F}(q) = \exp(-\alpha q^2), \alpha \leq 1$, such that the resulting values are within $[0, 1]$.

The model deformation criterion Q^d shall measure the degree of discrepancy between $\mathcal{T}(\mathbf{p}^0)$ and $\mathcal{T}(\mathbf{p}^t)$. In contrast to [12], our measure evaluates intrinsic properties in terms of the amount of non–rigid deformation of the shape model instance in its un–rotated reference frame. More specifically, the strain energy that is associated with the finite element nodes is adapted from [5], i.e.

$$Q^d = \mathcal{F}\big(\mu\big(\mathbf{q}_i(t)^2 \omega_i^{-2}\big)\big), \tag{8}$$

estimates the energy needed to match model and data. Since the low–order modes represent global variations, including the rigid body modes, while the high–frequency modes are the ones most sensitive to noise, we only consider the modal amplitudes corresponding to the m intermediate vibration modes, which explain a proportion β, e.g. $\beta = 0.25$, of the total shape variation (Sect. 2.1).

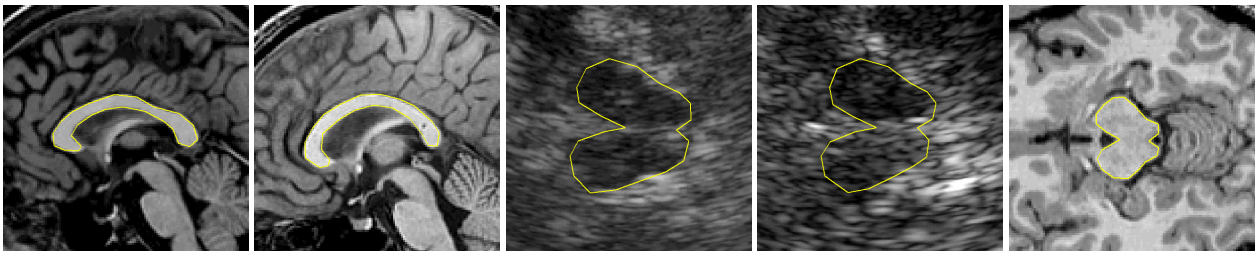
3 Experimental Evaluation

We tested the ability of the proposed quality–of–fit function to measure the quality of segmentation results in three representative medical applications, segmentation of the corpus callosum in sagittal slices from cerebral MRI data (30 data sets), segmentation of the mesencephalic brain stem in transcranial sonographies (TCS) (10 data sets), and segmentation of the midbrain in transversal slices from cerebral MRI data (30 data sets). We therefore constructed 2d–shape models that represent the corpus callosum and the mesencephalic brain stem in the different data sets[1].

For the segmentation of the MRI data sets, we computed Gaussian potential forces from gradient magnitude maps according to Section 2.2, and let $\kappa = 100$. Due to the specific noise characteristics in the TCS data, the boundaries of the mesencephalic brain stem may or may not be characterised by high intensity gradients. We therefore used for the TCS segmentation texture–based external model forces, as defined in [14]. All our experiments have been done with the same set of empirically chosen values for the material parameters ($E = 0.9, \nu = 0.25, \rho = 1$) that yielded a plausible behaviour of the elastic templates.

The model instances were initialised within the image domain by the 2d–affine transformation of the prototype from the model coordinate frame to the image coordinate frame. The transformations were characterised by the set of parameters position (x, y), orientation ψ and isotropic scaling s of the model instances in the image. (Note that the rigid–body deformations are represented by the low–order modes, which are in our case neglected in Equation 8.)

[1] Each of the application–specific shape models was constructed from a manual segmentation of an example data set.



(a) $\mathcal{Q} = 0.83$ (b) $\mathcal{Q} = 0.77$ (c) $\mathcal{Q} = 0.88$ (d) $\mathcal{Q} = 0.82$ (e) $\mathcal{Q} = 0.79$

Figure 1. Typical segmentation results and corresponding QOF–values $\mathcal{Q}$ for the example applications studied (note that the images were cropped for displaying the regions of interest).

A visual inspection of the segmentations by two neuroscientists suggested that the deformed model instances always provided correct (valid) segmentation results if the shape models were initialised at an adequate scale and orientation in the direct vicinity of the objects. The approximation of damping characteristics using (mass–proportional) Rayleigh damping (cf. Sect. 2.1, [5]) was sufficient, i.e. the model fit converged after 392 ± 78 iterations of the numerical simulation. No fix stopping criterion was required. Due to an adequate choice of external model forces, force linearisation [5, 13] and valid shape constraints incorporated within the shape prototypes, segmentation accuracy was very stable for variation in the values for the affine transformation parameters. The shape models adapted to the desired object within the images in case of imprecise initial positioning (up to $20mm$ displacement from the ideal position), variations at the order of 10% in scale and up to $40°$ variation in orientation. From our experience, the method is also robust to changes in the values for the material parameters. For an evaluation of the segmentations obtained in this experiment, we computed the quality–of–fit according to Equation 6, using $\alpha = 0.1, \beta = 0.25$ and $\zeta = 0.5$. In all cases, where the model instances were initialsed very close to the objects, the QOF–value of the fitted model instances was above 0.7 (Figure 1).

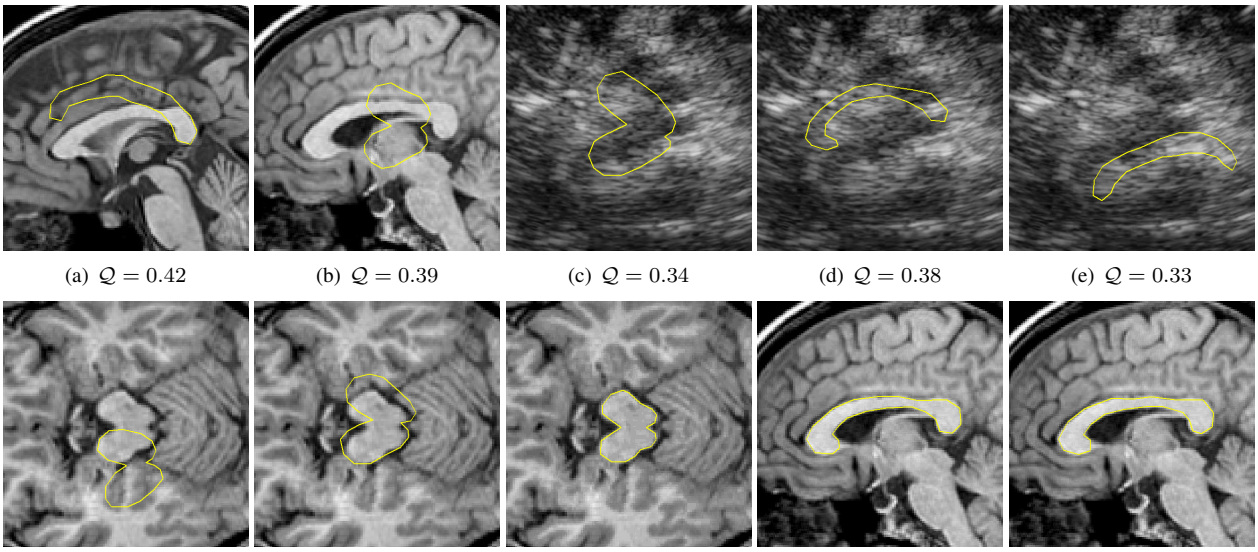
In order to test for the ability of the QOF–function to decide upon success and failure of the deformable model–based segmentation, we further simulated target, projection and evaluation errors, as discussed in Section 2.3. In a first experiment, we selected inadequate shape models, i.e. we applied for example the corpus callosum prototype to the segmentation of the TCS data, or initialised a shape model instance of the correct class at different positions in the data set that did not contain the desired object. Again, we calculated the quality–of–fit and found significantly lower values in all three applications ($p < 0.01$, one–sided t–test). A threshold of $\tau_Q = 0.5$ always allowed the separation of correct segmentations from invalid results (Figure 2).

In a second experiment, we used a purely image–based objective function (simulated by letting $\zeta = 0$ in Equation 6). As a result, correct solutions were indicated only in 54% of all cases. We also calculated the deformation quality according to [12], and used this measure in Equation 6. Our results have shown that this measure provided lower QOF–values for correct segmentations. The difference in the QOF–values compared to our implementation using Equation 8 was statistically significant ($p < 0.05$). However, in both cases no preferred value $\zeta \in [0.4, 0.6]$ for the influence of the deformation energy became evident from our results, and in more than 60% of all cases valid solutions were indicated by high QOF–values (72% using Equation 8). This underlines the importance of including the amount of deformation necessary to fit the data in a measure for the evaluation of the segmentation quality.

We further compared the QOF–values with a user rating of example segmentations. In some cases the experts favoured segmentations with relatively low QOF–values, and vice versa (Figures 2(i), 2(j)). In most cases, however, values of $\mathcal{Q} > 0.7$ could be interpreted as a high probability that the segmentation result corresponds to a correct solution.

4 Conclusion

We presented a quality–of–fit function for the evaluation of model–based segmentations based on the external and internal energies associated with a parametric shape model. In contrast to existing approaches, it combines a measure of the non–rigid shape deformation and a measure of the discrepancy between the deformed model and the expected image data. Experimental results were presented for different medical applications and showed that the proposed function allows determining success or failure of the model fit. In all cases where small deformations were required for bringing the model into correspondence with the image data, the segmentations were indicated by high QOF–values close to 1, while poorly fitted model instances yielded low values close to 0. A threshold of 0.5 allowed for deciding upon the success of the segmentation. This effect was independent from the external model–force formulation.



(a) $\mathcal{Q} = 0.42$ (b) $\mathcal{Q} = 0.39$ (c) $\mathcal{Q} = 0.34$ (d) $\mathcal{Q} = 0.38$ (e) $\mathcal{Q} = 0.33$

(f) $\mathcal{Q} = 0.4$ (g) $\mathcal{Q} = 0.41$ (h) $\mathcal{Q} = 0.78$ (i) $\mathcal{Q} = 0.81$ (j) $\mathcal{Q} = 0.86$

Figure 2. Examples of misleading segmentations due to target and projection error are shown in (a)–(g). Figures (f)–(h) demonstrate that incorrect results were in most cases indicated by QOF–values $\mathcal{Q} < 0.5$, while values above 0.7 could be interpreted as a high probability for a correct solution. The segmentation shown in Figure (i) was favoured by the expert, but assigned a smaller QOF–value compared with the (likewise correct) solution in Fig. (j).

Since the proposed objective function builds on a decomposition–based shape description (cf. Equation 2), it can also be used in combination with a statistical shape model, and employed in an ASM–based search [1, 2]. We believe that the proposed QOF–function may support object detection and localisation, and render a deformable shape search for the globally optimum parametrisation of the model (e.g., using regular sampling [16] or genetic algorithms [11]) more efficient. Future research will therefore focus on a more detailed evaluation of the QOF–function in cooperation with multiple experts for assessing its utility in view of its applicability for automatic medical image analysis.

Acknowledgements

We like to thank L. Niehaus and A. Brechmann for the fruitful discussion and for providing the data used in this paper.

References

1. Cootes T., et al.: The use of Active shape models for locating objects in medical images. *Imag Vis Comp* **12**(16) (1994) 355–366
2. Cootes T., et al.: Active Appearance Models. *Proc. ECCV* (1998) **2** 484–498
3. Kass M., et al: Snakes: Active contour models. *Int J Comp Vis* **1**(4) (1988) 321–331
4. Cohen L.: On active contour models and balloons. *CVGIP Imag Underst* (1991) **53**(2) 211-218
5. Sclaroff S. and Pentland A.: Modal matching for correspondence and recognition. *IEEE Trans Patt Anal Mach Intell* (1995) **17**(6) 545-561
6. Xu C. and Prince J.: Snakes, shapes, and gradient vector flow. *IEEE Trans Imag Proc* (1998) **7**(3) 359–369
7. McInerney T., et al.: Deformable organisms for automatic medical image analysis. *Med Imag Anal* **6**(3) (2002) 251–266
8. McInerney T. and Terzopoulos D.: Deformable Models in Medical Image Analysis. *Med Imag Anal* **1**(2) (1996) 91–108
9. He L., et al.: A comparative study of deformable contour methods on medical image segmentation. *Imag Vis Comp* **26** (2008) 141–163
10. Suri J. and Farag A.: *Deformable Models - Theory and Biomaterial Applications.* Springer, New York (2007)
11. Hill A. and Taylor C.: Model-based image interpretation using genetic algorithms. *Imag Vis Comp* (1992) **10**(5) 295-300
12. Bergner S., et al.: Deformable structural models. Proc. *IEEE ICIP* (2004) 1875–1878
13. Müller M., et al.: Stable Real–Time Deformations. *Proc. ACM SIGGRAPH Symp Comp Anim* (2002) 49–54
14. Engel K. and Toennies K.: Segmentation of the midbrain in transcranial sonographies using a two-component deformable model. *Proc. MIUA* (2008) 3–7
15. Ivins J. and Porrill J.: Statistical Snakes: Active Region Models. *Proc. BMCV* **2** (1994) 377–386
16. Felzenszwalb P.: Representation and detection of deformable shapes. *IEEE Trans Patt Anal Mach Intell* **27**(2) (2005) 208-220